How single cells sense Smad3 signal

Thesis by
Christopher Lee Frick

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
Defended 3 May 2019

Christopher Lee Frick
ORCID: 0000-0001-6823-5920

ACKNOWLEDGEMENTS

The list of people to thank is long, but I will hopefully keep this acknowledgement short. I would first like to thank, Lea, who consistently inspires me to both think bigger and be more concise. I would also like to thank the members of the lab who offered provided frequent helpful discussions: Kibeom Kim, Harry Nunns, and Noah Olsman. I would like to thank the rest of the lab during my time here (Mike Abrams, David Gold, Ty Basinger, Andrew Liu, and Mengsha Gong) for their friendship and support. I would like to draw attention to the work done by Laura Ratliff for this project as well. During her high school internship, Laura successfully cloned CRISPR plasmids that were used in the generation of endogenous Smad3 clones.

A big thank you is deserved as well by the Caltech Flow Cytometery and Cell Sorting Facility. Rochelle Diamond, Diana Perez, Jamie Tijerina, and Patrick Cannon—thank you for the hours of cell sorting. A big thanks as well to the many members of many other Caltech labs who took time out of their busy research to help train me: Zakary Singer, Eric Lubeck, and Long Cai for training in single molecule FISH; Yaron Antebi for his time training me in developing cell tracking software; and James Linton, for being a constant resource of scientific wisdom.

To my collaborators, Felix Dietrich and Yannis Kevrekidis, thank you for the incredible hard work you have poured into this project. You have both been a joy to work with and never cease to teach me new things. I would also like to thank my thesis committee for their support and direction, which I believe have greatly improved the quality and presentation of the work herein described.

Finally, I would like to thank my friends and family for their continued support, encouragement and prayers. The last leg of this thesis has been particularly demanding and would not have been possible without your help. Mom and Dad, thanks for your love and support. To Courtney, my wife, thank you for cheering for me, for making countless sacrifices for me during the last seven years, and for changing all of the diapers I missed while working long hours in lab. You are a constant source of joy and courage. And to Eli and Charlotte, thank you for joy and love. Dad loves you so much.

ABSTRACT

Animal cells possess the remarkable ability to send, receive, and respond to molecular signals. Accurate processing of these signals is essential for the development and maintenance of complex cell fates and organization. The regulation of cell behavior in response to signal is mediated by signal transduction pathways, which are highly conserved protein-protein interaction networks. Recent work has shown that the activation of biomolecular networks is highly sensitive to natural cellular variation in protein levels, making it unclear how these pathways accurately and reliably transmit signals in single cells. In this thesis, I address this question in the Transforming Growth Factor-β (Tgf-β) pathway, a major intercellular signaling pathway in animal cells. First, we asked whether extracellular signal is accurately transduced into pathway activation in single cells. Examining pathway dynamics in live reporter cells, we found evidence for fold-change detection. Although the level of nuclear Smad3 varied across cells, the fold change in the level of nuclear Smad3 was a more precise outcome of ligand stimulation. Indeed, by measuring Smad3 dynamics and gene expression in the same cells, we confirm that the fold-change in Smad3 carries signal in the pathway. These findings suggest that cells encode Tgf-β signal in a precise Smad3 fold-change as a strategy for coping with cellular noise. Second, we brought two significant advancements, which enabled us to ask how tightly signaling dynamics dictates target gene expression. By imaging endogenous dynamics of both signaling and gene expression in clonal cells, and correlating the full dynamics with a nonlinear manifold learning approach, we show that knowing the full dynamics of Smad3 is necessary but not sufficient to predict the full dynamics of target gene expression. Indeed, we find evidence for the role of mTOR, MEK5, and cell cycle as cell-specific variables that influence how a cell responds to Smad3. This demonstrates the extent to which, even across clonal cells, response to signal considerably varies, as each cell computes decisions based on its own internal state.

PUBLISHED CONTENT AND CONTRIBUTIONS

Frick, Christopher L, Clare Yarka, Harry Nunns, and Lea Goentoro (2017). "Sensing relative signal in the Tgf-β/Smad pathway". In: *Proceedings of the National Academy of Sciences*, p. 201611428. DOI: 10.1073/pnas.1611428114.
C.L.F. participated in the design and execution of the experiments, data analysis, figure development, and writing of the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 The language of cell-cell communication

In 1924, Spemann and Mangold made a remarkable discovery—cells communicate with one another (Spemann and Mangold, 1924). Studying amphibian development, they found that the development of an embryo into an adult organism depends on a small region of tissue known as an organizer. If this tissue is transplanted to the other side of an embryo, it induces dorsoanteriorization on that side, resulting in a two-headed tadpole. The presence of a specific set of cells determines the developmental fate of surrounding cells. It was thus shown that information can be transferred from one cell to another, and that this process of communication facilitates the breathtaking process of animal development. The nature of the signal being transferred and how this signal's instructions were read and responded to, however, were not yet known.

Nearly half a century later, laborious efforts of geneticists and biochemists had uncovered that communication between cells is mediated by networks of proteins. Information (signal) in these networks is propagated from one molecule to another. Extracellular signals (commonly referred to as ligands) bind to receptors on receiving cells. These receptors activate proteins within the cell that participate in a network of interactions, known as a signal transduction network, which generally culminate in activation of a transcription factor to effect changes in gene expression (Figure 1.1). Individual single cells, therefore, possess the machinery to receive and respond to instructions.

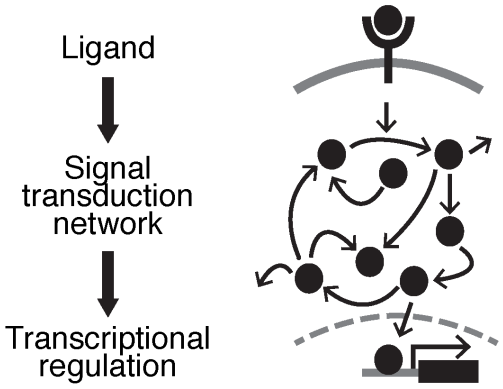

Figure 1.1: Schematic showing the general structure of a signaling pathway. Ligand binds to receptors at the cell surface, activating the intracellular signal transduction network, which culminates in changes in gene expression mediated by the pathway transcription factor.

At the early stages, the logic of cellular communication appeared straightforward. Similar to words in a human language, different ligands give different instructions. The discrete biochemical differences between these secreted molecules lead to interaction with different receptors and thus activation of different signal transduction pathways. A further layer of complexity was added when it was discovered that the instruction a ligand conveys can depend on the identity of the receiving cell—the same signal will give different responses in different cell types.

1.2 How to ensure precision in signal transduction?

Another layer of complexity appeared when it was discovered that signal transduction pathways derive meaning not just from the identity of the ligand, but also its quantity. The response of one given cell type to a ligand can depend on the concentration of that ligand the cells are exposed to. For example, it has been shown that the Tgf-β pathway can distinguish at least five different concentrations of Activin ligand in Xenopus embryo cells—the five concentrations lead to different cellular outcomes–specification of epidermis, posterolateral mesoderm, muscle, notochord, and neural inducing organizer tissue (Green, New, and Smith, 1992). Analysis showed that upon exposure to increasing ligand concentration, the activation of the Activin-dependent transcription factor, Smad2, increased proportionally, suggesting a simple model in which cell responses are determined by signal crossing multiple defined thresholds.

However, it was not long after it was found that signal is interpreted quantitatively

that the question of precision arose. The urgency of answering this question became apparent when it was discovered that protein concentrations, even in mammalian cells, are highly variable and fluctuating (Sigal et al., 2006). The approaches to study cell signaling at the time – utilizing using bulk analyses, i.e., lysing populations of cells, which averages out differences between single cells – became insufficient. Studies at the level of single cells would be necessary to determine how these complex signal transduction networks can both have highly variable components and function accurately in each single cell.

1.3 Dynamics matter

Theoretical work at the end of the 20th century suggested a potential solution. Despite the inevitable variations in concentration of network nodes, perturbation analysis of molecular networks showed that they can indeed exert accurate control over their responses (Barkai and Leibler, 1997; Shinar and Feinberg, 2010). In most networks, the outcome depends strongly on the initial conditions and parameter values—perturbing the network in even a small way can significantly alter its functioning. However, some networks can be wired in such a way that the network response is insensitive to variations in protein concentration. This observation raised the possibility that signaling pathways might exhibit robustness.

Indeed, as biochemical analysis provided clearer pictures of signal transduction network interactions, it became apparent that these pathways contain many features of engineering systems: positive and negative feedbacks, time delays, switches etc (Azeloglu and Iyengar, 2015). Signal transduction networks, therefore, could in theory be capable of performing complex computations and doing so robustly.

It was quite a surprise then, when the signal transduction activation was quantified in single cells and found to be highly variable from cell to cell. In cells exposed to the same concentration of ligand, signal transduction was different from cell to cell. In some instances, the differences were dynamic—the activation of transcription factor was asynchronously pulsatile (Tay et al., 2010). In others, the dynamics appeared uniform, but the activation of the terminal transcription regulator was imprecise (Cheong et al., 2011; Voliotis et al., 2014). Despite the highly conserved network of interactions, and the potential for robustness, these networks do not appear to accurately tune the activation of the transcription factor. How then do these pathways achieve precise control over the processes they regulate?

The continued study of signaling in single cells, however, soon began to provide

new insights that clarified how signal flows through pathways. Observations of the NF-κB signaling pathway in single cells first demonstrated that signaling activation could be oscillatory and further that target gene response could depend on the nature of the oscillations (D. Nelson et al., 2004). Work in this pathway and others led to models in which signal from ligand is "encoded" in complex dynamics of the pathway transcription factor, which is then "decoded" into gene activation or repression (Figure 1.2).

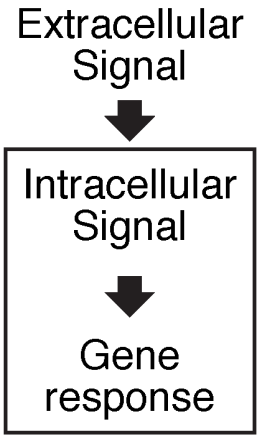

Figure 1.2: The signal from ligand is transduced into an intracellular signal, which is the dynamics of pathway activation. The signal carried by pathway dynamics must then be decoded into gene response

Subsequently, it was shown that the full dynamics of pathway activation encode significantly more information about extracellular ligand concentration than any time point measurements (Selimkhanov et al., 2014). Thus pathways could exhibit more precise control over gene expression if target genes response is a product of the full dynamics.

1.4 Robustness is encoded in dynamic features of signaling

Indeed, shortly thereafter it was discovered that some signal transduction networks are wired to accurately tune the dynamics of signaling. Studies of the Erk pathway and Wnt pathway were first to suggest that signaling networks are wired to transmit signal in accurate dynamics rather than accurate steady state levels. Specifically it was shown that these pathways encode signal in relative changes: while the concentration of the pathway transcription factor was highly variable in cells ,

the fold-change was nearly constant (Cohen-Saidon et al., 2009; Goentoro and Kirschner, 2009). This same strategy was found later in another major developmental signaling pathway, the TNF/NF-κB pathway (Lee et al., 2014). Therefore, signaling pathways have indeed evolved to produce signaling response that is robust to cell-to-cell variability, with the response being the ligand-induced fold change in the concentration of transcription factor.

The question that then must be asked of each signaling pathway is, How is signal transmitted? The finding that signaling pathways have evolved to encode robustness suggests an approach: to understand the central functioning of a pathway, one need only to look for the robust feature. From this search, insights into the design of the pathway naturally arise: how do the pathway interactions encode this response? What are the advantages and constraints of transmitting signal through this dynamical feature? Having identified the robust feature in the pathway it is also now possible to ask, How precise is the signaling process? Further, knowing the form of signal, we can ask, To what extent does signal dynamics dictate cellular response? This thesis pursues these questions in a major metazoan signaling pathway, the Tgf-β pathway.

1.5 Tgf-β pathway function and architecture

The Transforming growth factor beta (Tgf-β) is named for the function it was originally discovered to perform: transforming normal cells into actively growing malignant cells. It rapidly became apparent, however, that the functioning of Tgf-β was diverse and by no means limited to only malignant transformation: within 10 years of its discovery in 1983, it was shown that Tgf-β regulated a huge array of normal physiological functions of normal, non-cancerous tissues such as wound healing (Sporn et al., 1983), immune cell function (Wahl et al., 1987), and even embryogenesis and organogenesis (Flanders, Thompson, et al., 1989; Flanders, Ludecke, et al., 1991; Heine et al., 1991). As early as 1990, there were lists of 22 different cell-types and the effects of Tgf-β exposure (Massague, 1990).

The transducers of Tgf-β pathway are the Smad proteins. The Smads are a family of transcription factors, which become activated upon exposure to ligands in the Tgf-β superfamily. This ligand superfamily is comprised by 33 distinct secreted peptides, that fall broadly into two classes: Tgf-β-Smad activating ligands and BMP-Smad activating ligands. Two of the eight Smad proteins in mammals are specific to the Tgf-β ligands: Smad2 and Smad3, which are known as the receptor regulated

Smads. Smad4, which is known as the common Smad, is a common mediator to all Tgf-β superfamily members. When Tgf-β binds to Tgf-β-specific receptors at the cell surface, Smad2 and Smad3 are phosphorylated, form dimeric and trimeric complexes with Smad4, translocate to the nucleus, and become transcriptionally active. Smads bind to chromatin in conjunction with cofactors, and upregulate and downregulate hundreds of target genes. The specific target genes that activate depend on cell-type specific cofactors expressed (Mullen et al., 2011), mutations in pathway components (David et al., 2016), or active signaling crosstalks (Uttamsingh et al., 2008).

While the pathway appears linear in its logic (Tgf-β–>Smads–>Gene expression), the Smads undergo constitutive cytoplasmic-nuclear cycling. After phosphorylation, complex formation and nuclear localization, the Smad complexes can fall apart, and be dephosphorylated and exported back to the nucleus. Smads are thus constantly shuttling in and out of the nucleus, and constantly being phosphorylated and dephosphorylated. Further, Smad proteins are rapidly turned over, with a half-life of 1 to 2 hrs (Lo and Massagué, 1999; Xing et al., 2015; Hough, Radu, and Doré, 2012). The addition of Tgfβ results in to nuclear accumulation of Smad protein, as it increases phosphorylation and decreases nuclear export. This rapid cycling process has been hypothesized to enable rapid, continuous monitoring of extracellular ligand concentration (Schmierer et al., 2008)).

1.6 The need for a quantitative understanding of Tgf-β/Smad signaling in single cells

Despite the importance of Tgf-β/Smad signaling, the current working model for the Tgf-β pathway is derived almost exclusively from bulk analysis of populations of cells. Applying this model to single cells has been shown to be insufficient. This is highlighted most succinctly by recent measurements in fixed single cells revealing that the number of ligand-stimulated Smad complexes can vary by a factor of 40 from cell to cell (Zieba et al., 2012). How can we reconcile the precision required in determining cell fate during development with such imprecision observed in single-cell level?

This thesis aims to address the lack of understanding single-cell dynamics in the Tgf-β/Smad pathway. In Chapter 2, I explored the precision by which extracellular signal is transduced into intracellular signal. I find that the Tgf-β pathway employs fold-change detection as a strategy against noise. In Chapter 3, I examined how

intracellular signal is transduced into gene response. I find that, even within a population of clonal cells, intracellular signal dynamics only weakly predicts gene response. Instead, intrinsic variables within clonal individual cells strongly influence response to Smad dynamics. Indeed, I further uncover candidates for these intrinsic variables that govern a cell's interpretation of environmental signal. The totality of the work in this thesis has uncovered how signal is quantitatively transmitted through Tgf-β/Smad signaling pathway in single cells, revealing that cells send relative signals as a strategy against noise and that predicting a cell's response to signal requires knowing the cell itself.

Chapter 2

SENSING RELATIVE SIGNAL IN THE TGF-β /SMAD PATHWAY

Frick, Christopher L, Clare Yarka, Harry Nunns, and Lea Goentoro (2017). "Sensing relative signal in the Tgf-β/Smad pathway". In: *Proceedings of the National Academy of Sciences*, p. 201611428. DOI: 10.1073/pnas.1611428114.

Abstract How signaling pathways function reliably despite cellular variation remains a question in many systems. In the transforming growth factor-β (Tgf-β) pathway, exposure to ligand stimulates nuclear localization of Smad proteins, which then regulate target gene expression. Examining Smad3 dynamics in live reporter cells, we found evidence for fold-change detection. Although the level of nuclear Smad3 varied across cells, the fold change in the level of nuclear Smad3 was a more precise outcome of ligand stimulation. The precision of the fold-change response was observed throughout the signaling duration and across Tgf-β doses, and significantly increased the information transduction capacity of the pathway. Using single-molecule FISH, we further observed that expression of Smad3 target genes (*ctgf*, *snai1*, and *wnt9a*) correlated more strongly with the fold change, rather than the level, of nuclear Smad3. These findings suggest that some target genes sense Smad3 level relative to background, as a strategy for coping with cellular noise.

2.1 Introduction

Variability in the abundance of signaling components, across cells and contexts, is a well-documented feature of multiple signaling pathways (Cheong et al., 2011; Cohen-Saidon et al., 2009; Lee et al., 2014). Several ideas have been proposed for how cells can overcome variability: cells may integrate responses from multiple signaling pathways (Cheong et al., 2011), compensate via cross-talk with other pathways (Uda et al., 2013), average responses across neighboring cells (Cheong et al., 2011), use negative feedbacks (Voliotis et al., 2014), or measure signal dynamics (Selimkhanov et al., 2014). In our own work, we found another strategy, where cells overcome variability by interpreting signaling relative to background (Goentoro and Kirschner, 2009). Specifically, working in the canonical Wnt pathway, we presented evidence that signal is transduced through the ligand-induced fold change

in β-catenin level, rather than the absolute level.

Detecting fold change in signal level allows a cell to assign meaning to signal relative to its own background, enabling faithful transduction despite cellular variability. In addition to the Wnt pathway, fold-change detection has been proposed in the Erk pathway (Cohen-Saidon et al., 2009) and, subsequently, in the NF-κB pathway (Lee et al., 2014), calcium signaling (Thurley, Tovey, et al., 2014), and cytokine signaling (Thurley, Gerecht, et al., 2015). The evidence for fold-change detection in more and more systems suggests a conserved strategy across signaling pathways in animal cells. Motivated by these findings, we explored in this study for the presence of fold-change detection in a major channel of communication in cells, the transforming growth factor-β (Tgf-β) pathway.

The Tgf-β pathway functions across diverse contexts and tissues, and regulates fundamental processes, including proliferation, differentiation, morphogenesis, stem-cell maintenance, and regeneration (Massagué, 2012). These diverse functions are mediated by a highly conserved set of proteins. The Tgf-β pathway senses signal from a large family of secreted ligands, whose members include Tgf-β, Bmp, and Activin. Signal transduction is primarily mediated by the Smad proteins (Figure 2.1A): five ligand-specific receptor Smads (R-Smads; Smad1, Smad2, Smad3, Smad5, and Smad8), one common Smad (Smad4), and two regulatory Smads (Smad6 and Smad7) that act as feedback. The Smad proteins transduce signal in a dynamic process: They continually shuttle between the cytoplasm and nucleus, and ligand stimulation tunes this process. Specifically, ligand binds a complex of type I and type II serine/threonine kinase receptors, which phosphorylate the R-Smads. Phosphorylated R-Smads form a complex with the common Smad4. In their heteromeric form, the Smad proteins are retained more strongly in the nucleus through reduced export rate, as well as, as proposed recently (Schmierer et al., 2008), accelerated import rate. Thus, ligand activation leads to a net accumulation of the Smad complex in the nucleus, where it regulates target genes.

The Tgf-β pathway is a particularly interesting system for testing for fold-change detection because it is known that the expression levels of its components vary considerably from cell to cell. A recent study using proximity ligation assay in fixed cells revealed that the levels of Smad3/4 and Smad2/4 complexes vary by more than 40-fold across cells (Zieba et al., 2012). Consistent with this finding, our immunofluorescence analysis in a clonal cell population revealed significant overlap between the level of nuclear Smad3 in unstimulated and stimulated cells (Figure 2.7).

The variability in the level of Smad proteins from cell to cell raises the question as to how cells can reliably sense information about their external environment through the Tgf-β pathway.

Motivated by these observations, we tested whether signal in the Tgf-β pathway is sensed in an absolute manner or relative to background. Finding fold-change response in the Tgf-β pathway would expand our understanding of how information flows in the pathway, and how pathway activity should be interpreted appropriately across contexts and diseases. The alternative finding is equally interesting: If we find that cells monitor the absolute level of Smad proteins despite their variability, this finding will suggest that the mechanism that produces robust cellular outcomes is downstream from Smads.

2.2 Results

To investigate what aspects of Smad dynamics regulate gene response, we used live-cell imaging of the Tgf-β pathway. Responding to Tgf-β ligands, specifically, are the receptor-regulated Smad2 and Smad3 (R-Smads). Although structurally similar, Smad2 and Smad3 affect distinct genes (Brown, Pietenpol, and Moses, 2007), and it is also known that Smad3 can bind directly to DNA, whereas the predominant isoform of Smad2 does not (Gaarenstroom and Hill, 2014). Here, we focused on Smad3 and generated a reporter C2C12 cell line stably expressing an mNeonGreen-Smad3 construct (NG-Smad3; the sequence is shown in Figure 2.8A). Smad3 protein tagged on the N terminus retains phosphorylation at the C-terminal SXS motif, complex formation with Smad4, nuclear translocation, DNA binding, and transcriptional activity (Nicolás et al., 2004). We determined via Western blotting that NG-Smad3 is expressed at a moderate level, at twofold more than the endogenous Smad3 (Figure 2.8 B and C). We confirmed that NG-Smad3 is phosphorylated (Figure 2.8D) and translocates to the nucleus upon ligand stimulation (Figure 2.8E). We further confirmed that the signaling response is quantitatively identical across three clonal lines (Figure 2.9B). One clone was chosen for the measurements described here.

Stimulating the clonal reporter cells with Tgf-β ligand resulted in nuclear accumulation of NG-Smad3 (Figure 2.1B and Movie 2.15), as expected from published studies (Nicolás et al., 2004; Baker and Harland, 1996; Macías-Silva et al., 1996; Warmflash et al., 2012). In this experiment, cells were stimulated with purified recombinant Tgf-β1 (2.4 ng/mL). Images were acquired every 4 min, starting at 1 h

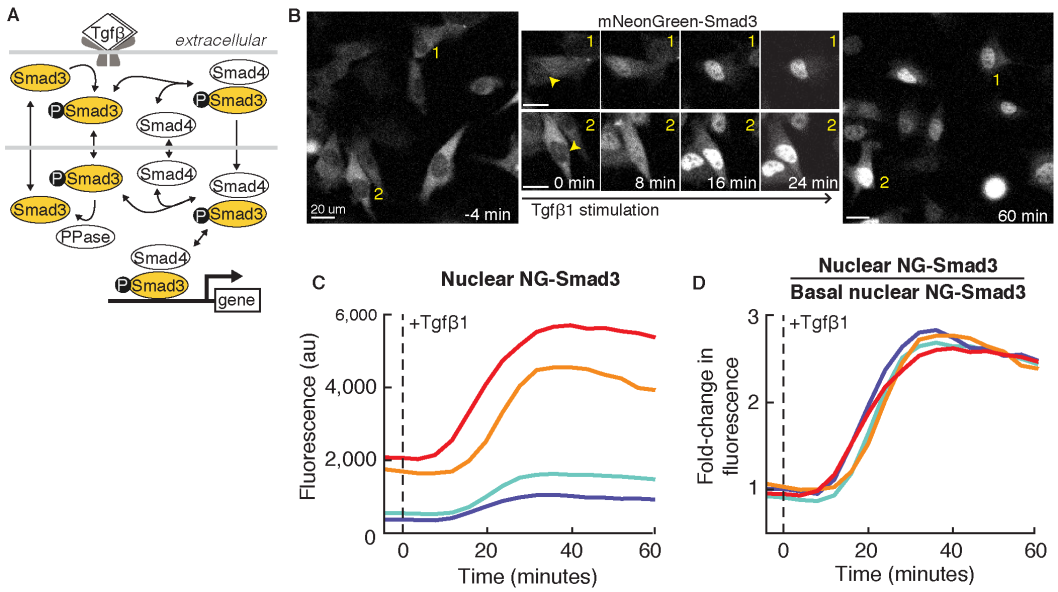

Figure 2.1: Ligand-induced nuclear accumulation of NG-Smad3. (*A*) Illustration of Smad3 activation and nucleocytoplasmic shuttling in the Tgf-β pathway. Ligand stimulation leads to phosphorylation of Smad3. Phosphorylated Smad3 complexes with Smad4 are shown. The Smad complex translocates to the nucleus and regulates target genes. The Smad complex may also dissociate, allowing Smad3 dephosphorylation and export back to the cytoplasm. (*B*) NG-Smad3 in C2C12 clonal reporter cells responding to ligand stimulation. Purified Tgf-β1 (2.4 ng/mL) was added to the cells at the start of the experiment (denoted as t = 0 min). (*Left*) Cells are shown before stimulation. (*Center*) Two individual cells are tracked over time. (*Right*) Same cells 60 min after stimulation. (Scale bars: 20 μm.) (*C*) Quantitation of the level of nuclear NG-Smad3 during Tgf-β1 stimulation. Each line corresponds to an individual cell. The dashed line indicates when Tgf-β1 was added. au, arbitrary units. (*D*) Fold change in nuclear NG-Smad3 from the same cells measured in *C*. Basal level is measured as the average of the fluorescence level 24 min before ligand stimulation.

before to up to 4 h after ligand stimulation. We segmented the nucleus using fluorescence signal from constitutively expressed nuclear mCerulean3, and then quantified the median fluorescence of NG-Smad3 in the nucleus for each cell (Materials and Methods). Nuclear accumulation of NG-Smad3 began immediately upon Tgf-β1 addition and peaked in most cells after 30 min, consistent with previously reported time scales of R-Smad (Nicolás et al., 2004; Warmflash et al., 2012). In nearly all cells, NG-Smad3 remained predominantly nuclear during the 4 h of imaging, with a slight decrease over time. We confirmed that cells only exposed to buffer showed no response (Figure 2.10A).

Quantifying the NG-Smad3 response in single cells, we indeed observed that the

level of nuclear NG-Smad3 varied across cells, even after Tgf-β1 stimulation (Figure 2.1C). In fact, cells with a high initial level of nuclear NG-Smad3 responded more strongly to Tgf-β1 stimulation, arriving at a higher final level (e.g., the orange and red traces in Figure 2.1C). In contrast, cells with a lower initial level of nuclear NG-Smad3, rather than compensating for the lower start, responded less to Tgf-β1 stimulation (e.g., the cyan and purple traces in Figure 2.1C). Thus, the cells do not appear to adjust the strength of their response to produce a robust level of nuclear NG-Smad3.

In comparison, these same cells exhibited more precise fold-change responses. The fold-change responses can be seen in Figure 2.1D, where we have now plotted the level of nuclear NG-Smad3 relative to the basal, prestimulus level. Indeed, the variability in Figure 2.1C arises because exposure to ligand stimulates an increase in nuclear Smad3 proportional to the initial level (e.g., 200–600 and 2,000–6,000 are both threefold changes). The linear proportionality between the basal and stimulated levels of nuclear NG-Smad3 holds for nearly two orders of magnitude (Figure 2.10C).

To confirm the higher precision in fold-change response, we quantified 299 cells responding to Tgf-β1 stimulation (Figure 2.2). Across these cells, the level of nuclear NG-Smad3 varied from cell to cell (Figure 2.2A; the distribution is shown in Figure 2.2C). By contrast, the fold change in the level of nuclear NG-Smad3 is substantially more precise (Figure 2.2B), resulting in a response distribution that is 3.7-fold more narrow than the absolute level distribution (computed using quartile coefficient of dispersion; Figure 2.2D). These results are reproducible across experiments (Figure 2.9A) and across three clonal cell lines (Figure 2.9B).

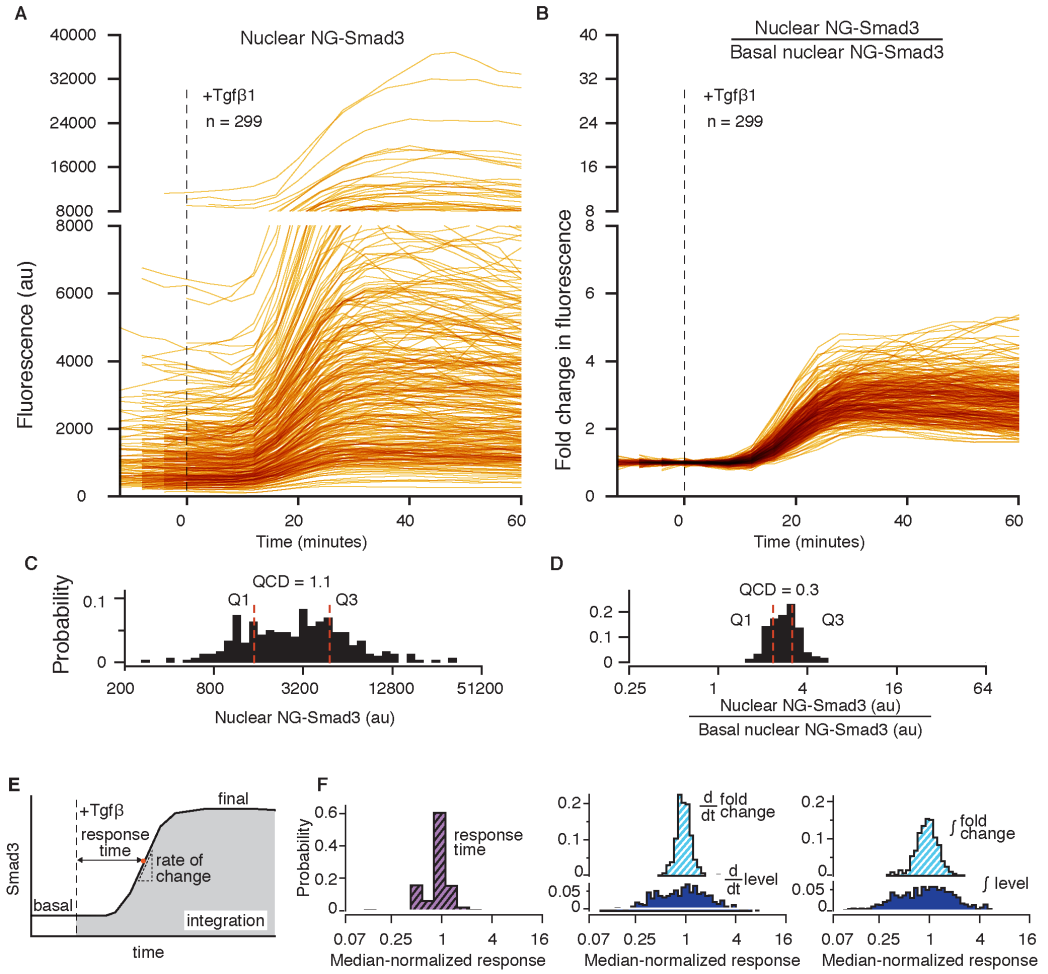

Figure 2.2: Fold change in nuclear NG-Smad3 is a more precise response to ligand stimulation. (*A* and *B*) NG-Smad3 responding to 2.4 ng/mL Tgf-β1 stimulation. The dashed line indicates when Tgf-β1 was added. Each line is a trace from a single cell, plotted as the absolute fluorescence level (*A*) or relative to its basal level (*B*). Basal level was computed as the average of nuclear NG-Smad3 fluorescence in the cell over 24 min before ligand stimulation. These data came from multiple experiments. The fluorescence distribution from each experiment was adjusted so that the median fluorescence is equal across experiments. No systematic differences were observed across experiments (Figure 2.9). Distribution of the level (*C*) and fold change in the level (*D*) of nuclear NG-Smad3 at 32 min after ligand stimulation are shown. Quartile coefficient of dispersion (QCD) is defined here as: (Q3 - Q1)/Q2, where Q1, Q2, and Q3 are the 25th, 50th, and 75th percentiles, respectively. (*E*) Illustration of the different response features examined in *F*. Response time was computed as the time from ligand addition until the inflection point in the response curve. Rate of change was computed as the maximum derivative of the response curve. Integrated response was computed over 52 min after ligand stimulation. (*F*) Distributions of the response time (purple), maximum rate of change, and integrated response computed from the absolute level (light blue) or fold change in nuclear NG-Smad3 (dark blue). The distributions are median-normalized to facilitate comparison. The distribution for response time is the same for fold change and absolute level, because it is defined as the time to maximum rate of change (the inflection point). Thus, only one distribution is shown.

Two features of the fold-change response are notable. First, the higher precision of the fold-change response persists throughout the duration of signaling, and not only at steady state (Figure 2.11A). Further, the response time of the fold-change response is well preserved across cells (purple-hatched distribution in Figure 2.2F). As a result, multiple features of the fold-change response are also more precise than the corresponding features computed using the absolute response, specifically, any monotonic functions, such as the integrated amount or the rate of change (Figure 2.2 E and F). Cells therefore may derive multiple robust computations from the Smad3 response sensed relative to background (e.g., integration, rate detection, timer).

The higher precision of the fold-change response suggests that cells could better sense external ligand by monitoring Smad3 response relative to background. To assess this possibility, we collected dose–response data, which we then analyzed using information theory. First, to test if the precision of the fold-change response is maintained at different doses of ligand, we stimulated cells with Tgf-β1 concentrations between 10 pg/mL and 2.4 ng/mL (which spans the dynamic range in our system). We observed that the fold change of NG-Smad3 increased as the dose of ligand stimulation increased, and remained a more precise response across ligand doses (Figure 2.3). Correspondingly, monotonic functions of the fold-change response (e.g., integration, derivative) also maintained precision across ligand doses and display dose dependence.

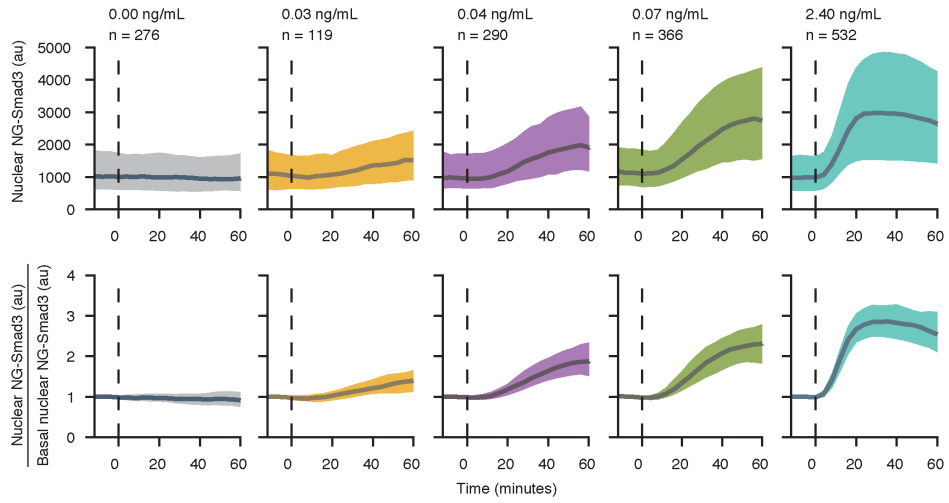

Figure 2.3: Higher precision of fold-change response holds across doses of Tgf-β. Plotted is the median (bold line) bounded by the 25th percentile and 75th percentile of the data (shaded area) from traces of the level of nuclear NG-Smad3 (*Upper*) and the fold change of nuclear NG-Smad3 (*Lower*). The dashed line indicates the time of Tgf-β addition.

Next, we analyzed the dose–response data using information theory. Pioneered by Claude Shannon in telecommunication (Shannon, 1948), information theory provides a mathematical framework for assessing information transmission across a communication channel, whether it be an electronic device (e.g., telegraph) or a signal transduction pathway (Cheong et al., 2011; Uda et al., 2013; Voliotis et al., 2014; Selimkhanov et al., 2014; Hansen and O'Shea, 2015). Specifically, the metric mutual information describes the extent to which measuring a particular response reduces uncertainty about the input (detailed in Materials and Methods). Because of noise in the communication channel, a given input will not necessarily produce a given response, but rather maps to a distribution of possible responses. The greater the noise, the greater the overlap is in the response distributions, and the lower the information is that the response gives about the input (Figure 2.4A).

To assess how the fold-change response facilitates information transduction in the Tgf-β pathway, we computed the maximum mutual information in the system. This quantity, also known as the channel capacity, describes the maximum amount of transducible information for a given input–response pair, and can be computed from measured dose–response distributions without making assumptions about the statistical properties of the input, the specifics of the transduction process, or the noise properties. Using the single-cell dose–response data in Figure 2.3, we first determined the maximum mutual information between the level of nuclear NG-Smad3 and Tgf-β input (Materials and Methods). The level of nuclear NG-Smad3 produced overlapping distributions across Tgf-β doses (Figure 2.4B), and could transduce, at most, ~0.2 bits of information (Figure 2.4D). The fold change of NG-Smad3, in contrast, produced considerably less overlap across Tgf-β doses (Figure 2.4C), and could transduce 1.2 bits of information (Figure 2.4D). Importantly, sensing fold change provides more information than absolute level throughout the entire signaling dynamics, even after only 8 min after ligand stimulation (Figure 2.11B). We extended the analysis to multiple computations from the Smad3 response (e.g., integration, rate of change), and found that relative computations consistently transduce higher information than what their absolute counterparts could transduce (Figure 2.4D). Therefore, as expected from the higher precision across cells, our results suggest that information about the Tgf-β input is more accurately transmitted through the fold change in nuclear NG-Smad3.

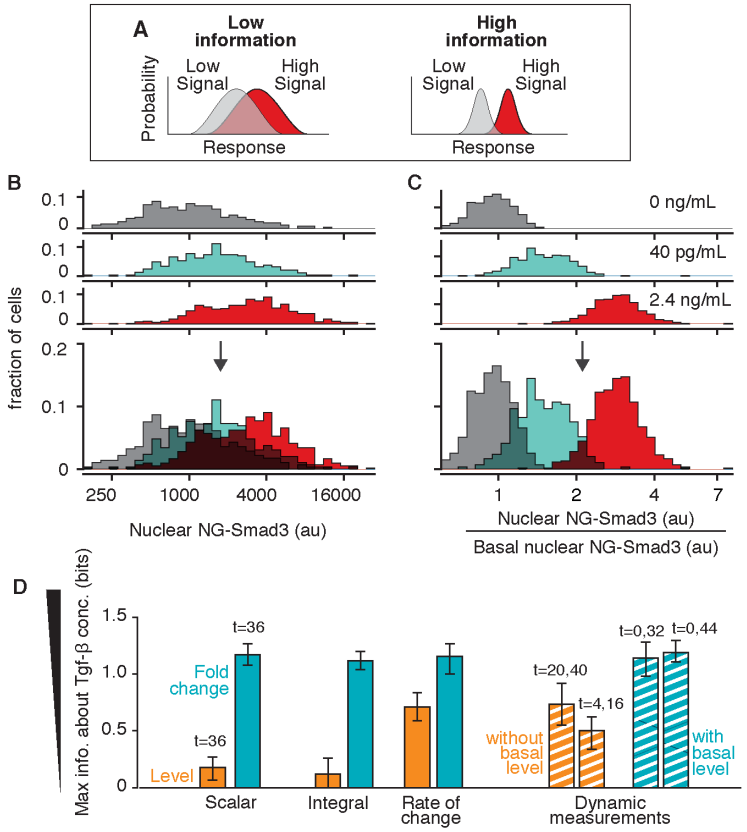

Figure 2.4: Fold-change response has higher information transduction capacity. (*A*) Noisy, overlapping response distributions provide low information about the strength of ligand input. (*B* and *C*) To compute the maximum mutual information, we stimulated the cells with different doses of Tgf-β1 (Figure 2.3). The response distributions for three doses are shown here, of the absolute fluorescence level (*B*) or the fold change (*C*). (*Bottom*) Overlay of the distributions. For low, medium, and high doses, the number of cells examined was 277, 290, and 532, respectively. (*D*) We computed the maximum mutual information between ligand input and different features of the nuclear NG-Smad3 response. Features computed using the absolute response are shown in orange, and features computed using the fold-change response are shown in blue. Level and fold change of nuclear NG-Smad3 were evaluated at steady state, at 36 min after Tgf-β addition (comparison at different time points is shown in Figure 2.11B). Rate of change in the NG-Smad3 response was computed as the maximum of the derivative of the response curve. To compute the integral of the NG-Smad3 response, the response was integrated over the first hour of ligand stimulation. For dynamic measurements, the level of nuclear NG-Smad3 was measured at multiple time points, as indicated, and mutual information was computed with a 2D distribution (Figure 2.14). Error bars are 90% confidence intervals computed using bootstrap resampling. The total number of cells examined for each calculation was 1,650.

The robustness of and the higher information carried by the fold-change response

raise the question of whether target genes use this feature and respond to the fold change in nuclear Smad3. To test this possibility, we combined live-cell imaging with single-molecule RNA FISH (Figure 2.5A). By keeping track of the position of the cells within the imaging field, we could measure the NG-Smad3 response, and subsequently obtain mRNA counts from the same cell. We can, therefore, correlate, within single cells, both the signaling dynamics and target gene expression (Lee et al., 2014; Singer et al., 2014). We first filmed cells responding to Tgf-β1 stimulation, and then fixed cells and stained for mRNA. To count the mRNA molecules, we took optical z-sections of the entire cells and performed automated detection of mRNA foci using custom MATLAB scripts (Materials and Methods). We examined known direct targets of Smad3, *snail*, and *ctgf*. Smad3 has been shown to bind directly to the promoters of these genes upon Tgf-β stimulation (Brandl et al., 2010; Zhang et al., 2011). Both genes are involved in various processes, including epithelial-mesenchymal transition, cell adhesion, fibrosis, and extracellular matrix remodeling (Fan, Pech, and Karnovsky, 2000; Medici, Hay, and Olsen, 2008). For each gene, we characterized the expression profile over 6 h and report here the transcript counts at peak expression, which occurs at 1 h after ligand stimulation.

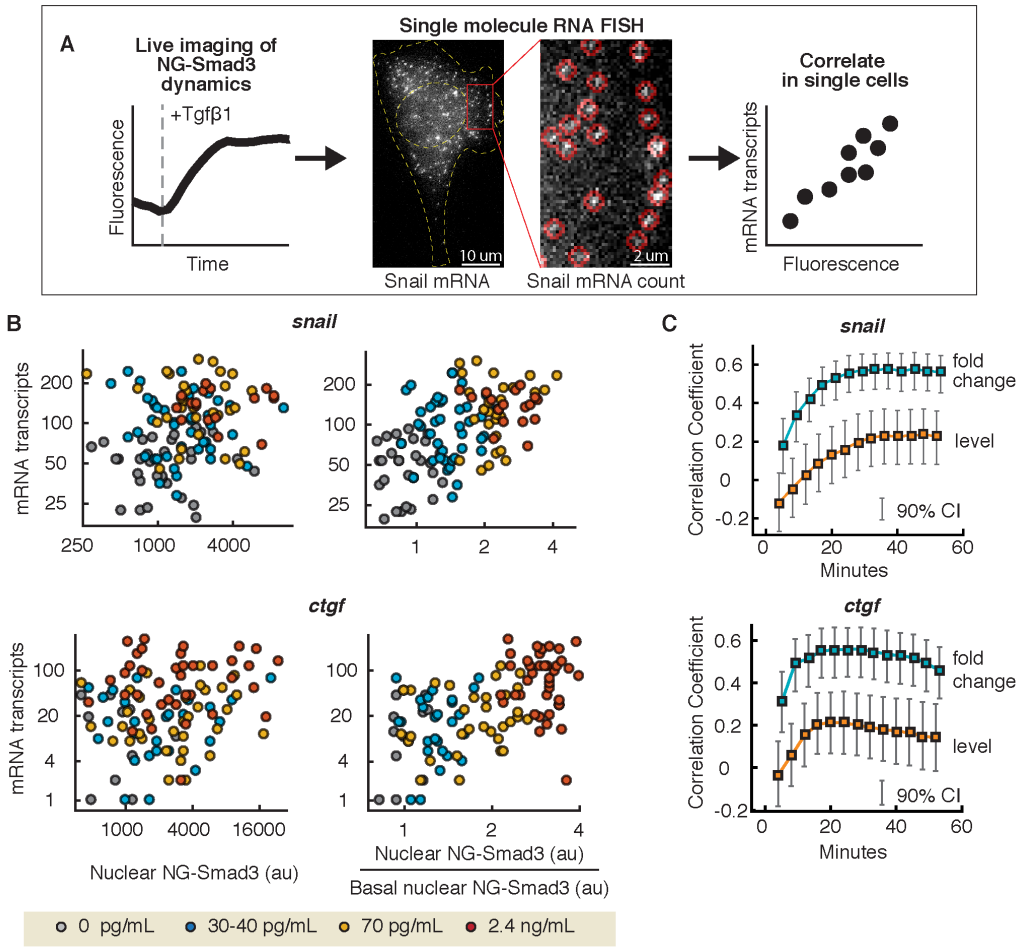

Figure 2.5: Expression of target genes correlates more strongly with the fold change in nuclear NG-Smad3. (*A*) To correlate NG-Smad3 dynamics and transcription response within a single cell, we combined live-cell imaging with smFISH. Cells were stimulated with Tgf-β1 and imaged. The same cells were then fixed, stained against specific mRNA, and imaged again. Foci corresponding to individual mRNA molecules were quantified using custom MATLAB scripts. The mRNA transcript counts were then plotted against features of NG-Smad3 response from the same cells. (*B*) Number of mRNA transcripts plotted against the level (*Left*) or fold change (*Right*) of nuclear NG-Smad3 measured in the same cell. The mRNA transcripts were counted after 1 h of Tgf-β stimulation, and plotted here with response features measured at 44 min (*snail*) and 28 min (*ctgf*) after ligand stimulation. (*C*) Plotted is the correlation between mRNA transcripts (at 1 h after ligand stimulation) and NG-Smad3 response measured throughout the entire signaling dynamics. The correlation coefficient is Spearman's rank correlation coefficient. Error bars are 90% confidence intervals (CI), computed using bootstrap resampling. At each time point after ligand addition, the correlation with fold change (blue) is significantly higher than the correlation with level (orange) ($P < 0.01$, Steiger's Z test; a complete statistical analysis is shown in Table 2.1).

We observed variability in the mRNA expression of the target genes, which may be due to variability in cell size, chromatin state, cell-cycle phase, other extrinsic variables, or stochastic noise. Although many factors can contribute to gene regulation, we focus here on discerning the effects of Smad3 dynamics. While the variability may mask some of the correlation, clear trends were observable Figure (2.5). The expression of both target genes appeared to be linearly proportional to the magnitude of fold change in Smad3 level and showed no strong dependence on the absolute level of Smad3. Figure 2.5 shows the mRNA counts plotted against the absolute level (Figure 2.5B, *Left*) or the fold change in the level of nuclear NG-Smad3 (Figure 2.5B, *Right*). Further, the higher correlation of mRNA expression and the fold-change response was apparent throughout the entire duration of NG-Smad3 dynamics (Figure 2.5C), and was statistically significant ($P < 0.01$, Steiger's Z test; Table 2.1). The same result was observed with another direct target gene, wnt9 (Figure 2.12). These results suggest that some target genes of the Tgf-β pathway respond to the fold change in Smad3, rather than the absolute level.

2.3 Discussion

There has been growing evidence that the dynamics of Smad proteins are important for their functioning (Schmierer et al., 2008; Warmflash et al., 2012). In this study, we explored the dynamics of R-Smads to investigate how the Tgf-β pathway solves the problem of cellular variability. First, using single-cell live imaging, we found that fold change in Smad3 level, rather than absolute level, is the outcome of ligand stimulation that is more robust to cell-to-cell variation. Then, analyzing the response distributions across doses, we found that measuring fold change in Smad3 indeed confers higher information transduction capacity to the Tgf-β pathway. The robustness and higher information transduction capacity suggest that fold change in Smad3 is a meaningful signal sensed by the cells. Measuring gene response and Smad3 dynamics in single cells, we found that some direct targets of Tgf-β indeed correlate more strongly with fold change in Smad3, rather than absolute level. Altogether, these findings suggest that, at least in some contexts, cells sense the relative level of Smad3 as a way to transmit information more accurately despite cellular variability (Figure 2.6).

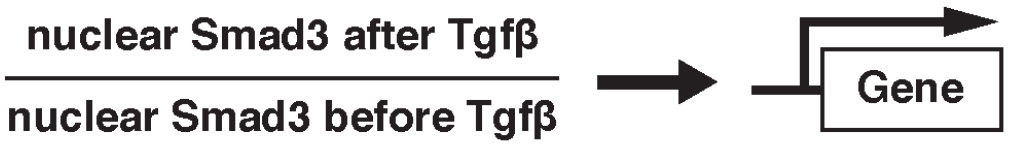

Figure 2.6: Our finding suggests that, at least in some contexts, Smad signal in the Tgf-β pathway is sensed in a relative manner.

Sensing the relative change in Smad3 may be useful for allowing the Tgf-β pathway to function reliably in diverse processes (e.g., migration, differentiation, cell death) and diverse tissues (e.g., adipose, muscle, epithelia) (Massagué, 2012), where concentrations of the components of the pathway are known to vary significantly across cells (Zieba et al., 2012), tissues (Flanders, Kim, and Roberts, 2001), and developmental stages (Flanders, Kim, and Roberts, 2001). The finding that signal may be sensed relative to background also means that a high level of nuclear Smad complex does not necessarily indicate a high level of signaling, and this finding may have implications for understanding the context-dependent outcomes of the Tgf-β pathway. For instance, the Tgf-β pathway is known to act as a tumor suppressor in early tumorigenesis and as a tumor promoter in the later stages (Akhurst and Hata, 2012). This and other context-dependent outcomes are thought to arise from cell type-specific interactions between Smads and master transcription factors, the epigenetic status of target gene promoters/enhancers, and cross-talk with other pathways (Massagué, 2012; Akhurst and Padgett, 2015). In the framework of fold-change detection, a part of the context-dependent outcomes may arise from the signal being interpreted relative to different quantitative backgrounds.

The higher precision of the fold-change response is predicted by mathematical modeling of the Tgf-β pathway. A mathematical model capturing the nucleocytoplasmic dynamics of Smads upon Tgf-β stimulation was developed by Schmierer et al. (Schmierer et al., 2008). The model displays a high degree of predictive power, fitting four independent measurements in cells and predicting results from two independent datasets that were not used to construct the model. We find that simulations of the Tgf-β model predict that fold change in nuclear Smad3 will be more robust to parameter variation than the absolute level of nuclear Smad3 (Figure 2.13). The robustness also holds and is even more pronounced for the fold change in the Smad complex. The parameter sensitivity analysis also indicates that the robustness in

the fold-change response starts breaking down with large-parameter variations, indicating that cells must operate in a specific parameter regime to achieve this robust feature in Smad3 response (Figure 2.13C and D). Similar parameter tuning is also required in the fold-change response in the canonical Wnt pathway (Goentoro and Kirschner, 2009). Overall, these results suggest that the robust fold-change response arises from the conserved interactions of the pathway.

The higher precision of the fold-change response of Smad3 significantly increases the information transduction capacity of the Tgf-β pathway. It was recently proposed that measuring the absolute signal dynamics over multiple time points may increase channel capacity in Erk, NF-κB, and the calcium pathway (Selimkhanov et al., 2014). We tested this idea in the Tgf-β pathway and found that measuring the signaling dynamics of Smad3 at multiple time points can give comparable mutual information to the fold-change response, provided that one of the time points was the basal state (Figure 2.4D and Figure 2.14), therefore strengthening our findings here. Our findings suggest that, despite variability in the Smad level, there is indeed reliable signal processing within the Tgf-β pathway.

Looking downstream in the pathway, the correlation between target gene expression and fold change in Smad3 necessitates a mechanism for computing fold changes. A recurrent motif in transcriptional networks, the type-1 incoherent feedforward loop, was shown to have the ability to provide fold-change computation (Goentoro, Shoval, et al., 2009). Smads, known activators, also effect repression through recruitment of repressors, such as ATF3 or E2F4 (Massagué, Seoane, and Wotton, 2005), or inducing specific microRNAs that repress their own target genes (Butz et al., 2012). In the context of our findings, these seemingly opposite actions of Smads may mediate fold-change detection.

Our present work places the Tgf-β pathway among the increasing number of signaling pathways where fold-change response has now been identified or proposed (Cohen-Saidon et al., 2009; Lee et al., 2014; Goentoro and Kirschner, 2009; Thurley, Tovey, et al., 2014; Thurley, Gerecht, et al., 2015). Our finding reinforces an emerging theme across signaling pathways in animal cells, that signaling dynamics are sensed in a relative manner. Beyond signaling in cells, sensing signal in a relative manner brings to mind the Weber's law in sensory systems (Laming, 1986; Shepard, Kilpatric, and Cunningham, 1975; Weber, Ross, and Murray, 1996), and highlights a convergence between biological sensory systems at the single-cell and organismal levels.

References

Akhurst, Rosemary J and Akiko Hata (2012). "Targeting the TGFβ signalling pathway in disease". In: *Nature reviews Drug discovery* 11.10, p. 790.

Akhurst, Rosemary J and Richard W Padgett (2015). "Matters of context guide future research in TGFβ superfamily signaling". In: *Sci. Signal.* 8.399, re10–re10.

Baker, Julie C and Richard M Harland (1996). "A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway." In: *Genes & Development* 10.15, pp. 1880–1889.

Brandl, Martina et al. (2010). "IKKα controls canonical TGFβ–SMAD signaling to regulate genes expressing SNAIL and SLUG during EMT in Panc1 cells". In: *J Cell Sci* 123.24, pp. 4231–4239.

Brown, Kimberly A, Jennifer A Pietenpol, and Harold L Moses (2007). "A tale of two proteins: Differential roles and regulation of Smad2 and Smad3 in TGF-β signaling". In: *Journal of cellular biochemistry* 101.1, pp. 9–33.

Butz, Henriett et al. (2012). "Crosstalk between TGF-β signaling and the microRNA machinery". In: *Trends in pharmacological sciences* 33.7, pp. 382–393.

Cheong, Raymond et al. (2011). "Information transduction capacity of noisy biochemical signaling networks". In: *science* 334.6054, pp. 354–358.

Cohen-Saidon, Cellina et al. (2009). "Dynamics and variability of ERK2 response to EGF in individual living cells". In: *Molecular cell* 36.5, pp. 885–893.

Fan, Wen-Hua, Michael Pech, and Morris J Karnovsky (2000). "Connective tissue growth factor (CTGF) stimulates vascular smooth muscle cell growth and migration in vitro". In: *European journal of cell biology* 79.12, pp. 915–923.

Flanders, Kathleen C, Edward S Kim, and Anita B Roberts (2001). "Immunohistochemical expression of smads 1–6 in the 15-day gestation mouse embryo: signaling by BMPs and TGF-βs". In: *Developmental dynamics: an official publication of the American Association of Anatomists* 220.2, pp. 141–154.

Gaarenstroom, Tessa and Caroline S Hill (2014). "TGF-β signaling to chromatin: how Smads regulate transcription during self-renewal and differentiation". In: *Seminars in cell & developmental biology*. Vol. 32. Elsevier, pp. 107–118.

Goentoro, Lea and Marc W Kirschner (2009). "Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling". In: *Molecular cell* 36.5, pp. 872–884.

Goentoro, Lea, Oren Shoval, et al. (2009). "The incoherent feedforward loop can provide fold-change detection in gene regulation". In: *Molecular cell* 36.5, pp. 894–899.

Hansen, Anders S and Erin K O'Shea (2015). "Limits on information transduction through amplitude and frequency regulation of transcription factor activity". In: *Elife* 4, e06559.

Laming, Donald (1986). *Sensory analysis*. Academic Press.

Lee, Robin EC et al. (2014). "Fold change of nuclear NF-κB determines TNF-induced transcription in single cells". In: *Molecular cell* 53.6, pp. 867–879.

Macías-Silva, Marina et al. (1996). "MADR2 is a substrate of the TGFβ receptor and its phosphorylation is required for nuclear accumulation and signaling". In: *Cell* 87.7, pp. 1215–1224.

Massagué, Joan (2012). "TGFβ signalling in context". In: *Nature reviews Molecular cell biology* 13.10, p. 616.

Massagué, Joan, Joan Seoane, and David Wotton (2005). "Smad transcription factors". In: *Genes & development* 19.23, pp. 2783–2810.

Medici, Damian, Elizabeth D Hay, and Bjorn R Olsen (2008). "Snail and Slug promote epithelial-mesenchymal transition through β-catenin–T-cell factor-4-dependent expression of transforming growth factor-β3". In: *Molecular biology of the cell* 19.11, pp. 4875–4887.

Nicolás, Francisco J et al. (2004). "Analysis of Smad nucleocytoplasmic shuttling in living cells". In: *Journal of cell science* 117.18, pp. 4113–4125.

Schmierer, Bernhard et al. (2008). "Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system". In: *Proceedings of the National Academy of Sciences* 105.18, pp. 6608–6613.

Selimkhanov, Jangir et al. (2014). "Accurate information transmission through dynamic biochemical signaling networks". In: *Science* 346.6215, pp. 1370–1373.

Shannon, Claude Elwood (1948). "A mathematical theory of communication". In: *Bell system technical journal* 27.3, pp. 379–423.

Shepard, Roger N, Dan W Kilpatric, and James P Cunningham (1975). "The internal representation of numbers". In: *Cognitive psychology* 7.1, pp. 82–138.

Singer, Zakary S et al. (2014). "Dynamic heterogeneity and DNA methylation in embryonic stem cells". In: *Molecular cell* 55.2, pp. 319–331.

Thurley, Kevin, Daniel Gerecht, et al. (2015). "Three-dimensional gradients of cytokine signaling between T cells". In: *PLoS computational biology* 11.4, e1004206.

Thurley, Kevin, Stephen C Tovey, et al. (2014). "Reliable encoding of stimulus intensities within random sequences of intracellular Ca2+ spikes". In: *Sci. Signal.* 7.331, ra59–ra59.

Uda, Shinsuke et al. (2013). "Robustness and compensation of information transmission of signaling pathways". In: *Science* 341.6145, pp. 558–561.

Voliotis, Margaritis et al. (2014). "Information transfer by leaky, heterogeneous, protein kinase signaling systems". In: *Proceedings of the National Academy of Sciences* 111.3, E326–E333.

Warmflash, Aryeh et al. (2012). "Dynamics of TGF-β signaling reveal adaptive and pulsatile behaviors reflected in the nuclear localization of transcription factor Smad4". In: *Proceedings of the National Academy of Sciences* 109.28, E1947–E1956.

Weber, EH, Helen E Ross, and David J Murray (1996). "EH Weber on the tactile senses". In:

Zhang, Yingze et al. (2011). "High throughput determination of TGFβ1/SMAD3 targets in A549 lung epithelial cells". In: *PLoS One* 6.5, e20319.

Zieba, Agata et al. (2012). "Intercellular variation in signaling through the TGF-β pathway and its relation to cell density and cell cycle phase". In: *Molecular & Cellular Proteomics* 11.7, pp. M111–013482.

2.4 Supplementary Figures

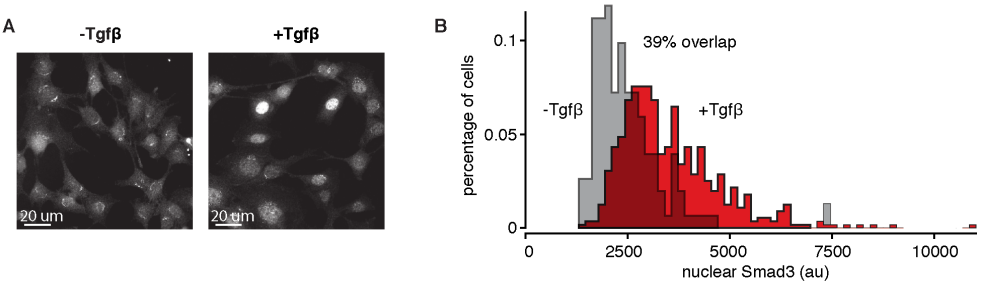

Figure 2.7: Immunofluorescent staining reveals overlap of the Smad3 level in unstimulated and stimulated cells. (*A*) Immunofluorescent staining of Smad3 in C2C12 cells: unstimulated (*Left*) and stimulated with 2.4 ng/mL Tgf-β1 (*Right*). (*B*) Quantitation of the fluorescence signal from unstimulated cells (gray, n = 152) and cells stimulated for 1 h with 2.4 ng/mL Tgf-β1 stimulation (red, n = 557). The overlap did not decrease at subsequent time points. For immunofluorescence, cells were grown overnight on 96-well glass-bottomed plates (Griener Bio-One; 655892). After Tgf-β stimulation, cells were fixed with 4% (wt/vol) paraformaldehyde for 20 min. Following fixation, cells were rinsed twice with 1×PBS, blocked, permeabilized with 5% (vol/vol) goat serum and 0.1% Triton X-100 in PBS, and then incubated with primary antibody diluted in blocking buffer overnight at 4 °C. Cells were then washed three times with 1×PBS + 0.05% Tween-20 for 5–10 min each time, and incubated in secondary antibody diluted in 1×PBS + 0.05% Tween-20 for 1 h in the dark at room temperature. The cells were washed again three times with 1×PBS + 0.05% Tween-20 for 5–10 min each time, with DAPI counterstain added in the third wash, followed by one more wash with 1×PBS for 5 min. Imaging was done in 1× PBS. Monoclonal rabbit anti-Smad3 (Cell Signaling; C67H9) was used at 1:100, and goat–anti-Rabbit DyLight 650 (Thermo Fisher Scientific; SA510034) was used at 1:2,000. au, arbitrary units.

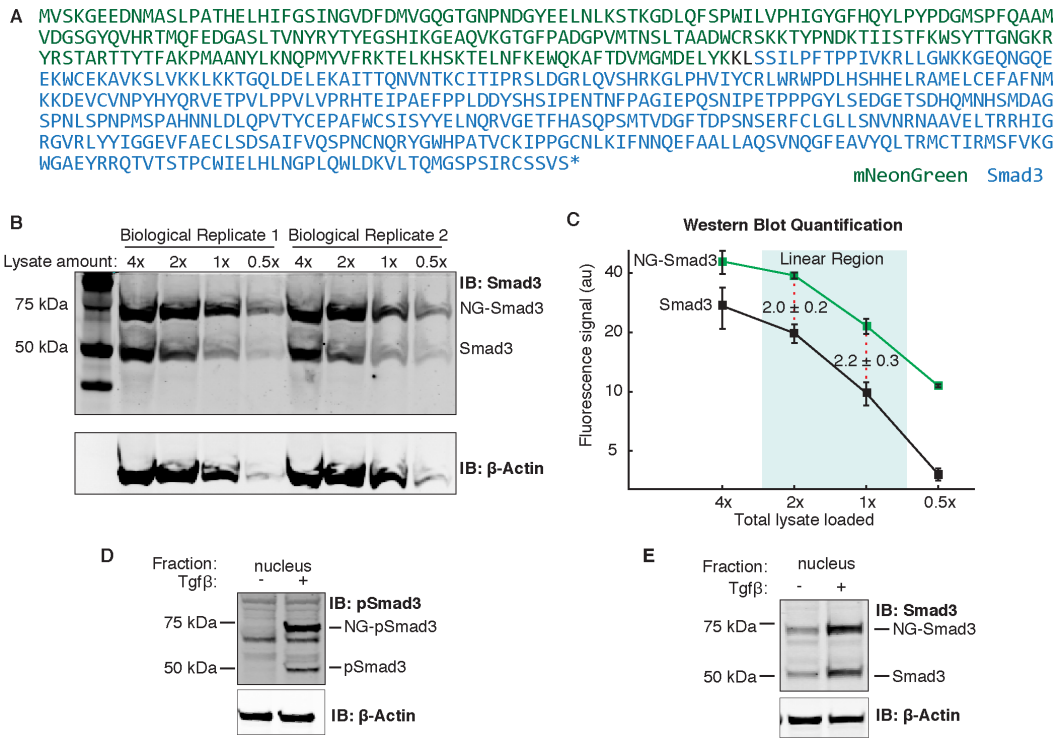

Figure 2.8: Characterization of the NG-Smad3 construct. (*A*) Amino acid sequence of the NG-Smad3 used in the study. The asterisk denotes a stop codon. (*B* and *C*) Quantifying the level of NG-Smad3 relative to endogenous Smad3 using quantitative Western blotting. (*B*) Cell lysate was loaded at different concentrations to find the range where antibody staining is linear. IB, immunoblot. (*C*) Linear region (blue-shaded area) is where the change in lysate amount is proportional to the change in fluorescent signal. The relative expression of NG-Smad3 to Smad3 was determined by taking the average of two biological replicates. (*D*) Western blot against phosphorylated Smad3 (pSmad3) in the nuclear fraction. (*E*) Western blot against Smad3 in the nuclear fractions. In both *D* and *E*, cells were collected 1 h after Tgf-β stimulation (2.4 ng/mL). For Western blotting, cells were treated with Tgf-β1 or a carrier, harvested by trypsinization, and then either pelleted and frozen or fractionated into nuclear and cytoplasmic fractions using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific; 78833). Cell lysates were loaded on Bolt 4–12% gradient SDS/PAGE (Invitrogen) and transferred onto nitrocellulose membranes by wet transfer using a standard wet transfer buffer [25 mM Tris, 192 mM glycine, 20% (wt/vol) methanol] for 1 h at 200 mA at 4 °C. Membranes were dried, blocked using Odyssey blocking buffer (LI-COR; 927-50000) for 1 h at room temperature, and incubated with primary antibodies at 4 °C overnight and secondary antibodies for 1 h at room temperature. Imaging and quantification were performed using a LI-COR Odyssey infrared scanner. Primary antibodies were as follows: rabbit–anti-Smad3 (Cell Signaling; C67H9) at a 1:1,000 dilution, rabbit–anti-pSmad3 (Cell Signaling; C25A9) diluted at 1:1000, and mouse-anti–β-actin (Cell Signaling; 8H10D10) diluted at 1:20,000. All primary antibodies were diluted in blocking buffer. Secondary antibodies, goat–anti-rabbit IRDye 680LT (LI-COR; 925-68021), and goat–anti-mouse IgG (H+L) DyLight 800 Conjugate (Cell Signaling; 5257) were diluted at 1:5,000. All secondary antibodies were diluted in blocking buffer + 0.1% Tween-20 + 0.01% SDS.

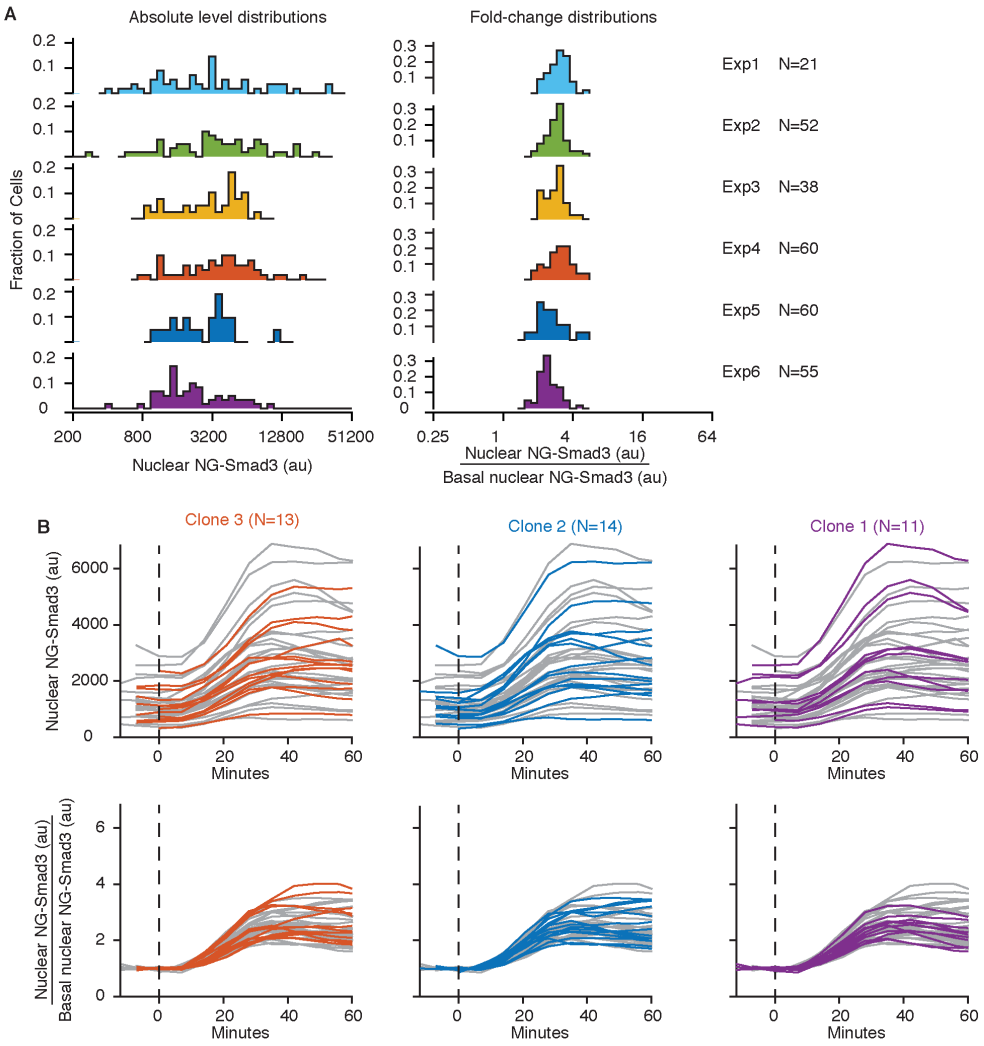

Figure 2.9: Higher precision of fold-change response is reproducible across experiments and clonal lines. (*A*) Data from six of 20 independent experiments performed in the study. (*Left*) Distribution of the level of nuclear NG-Smad3 at 32 min after ligand stimulation (2.4 ng/mL Tgf-β1). (*Right*) Fold change in nuclear Smad3. (*B*) Data from three clonal cell lines. In each plot, the colored lines are traces from the indicated clone and the gray lines are traces from all clones together. The data are plotted as the absolute fluorescence level (*Top*) or relative to its basal level (*Bottom*). The dashed line indicates when Tgf-β1 was added (2.4 ng/mL). Data from clone 3 are presented in the main text.

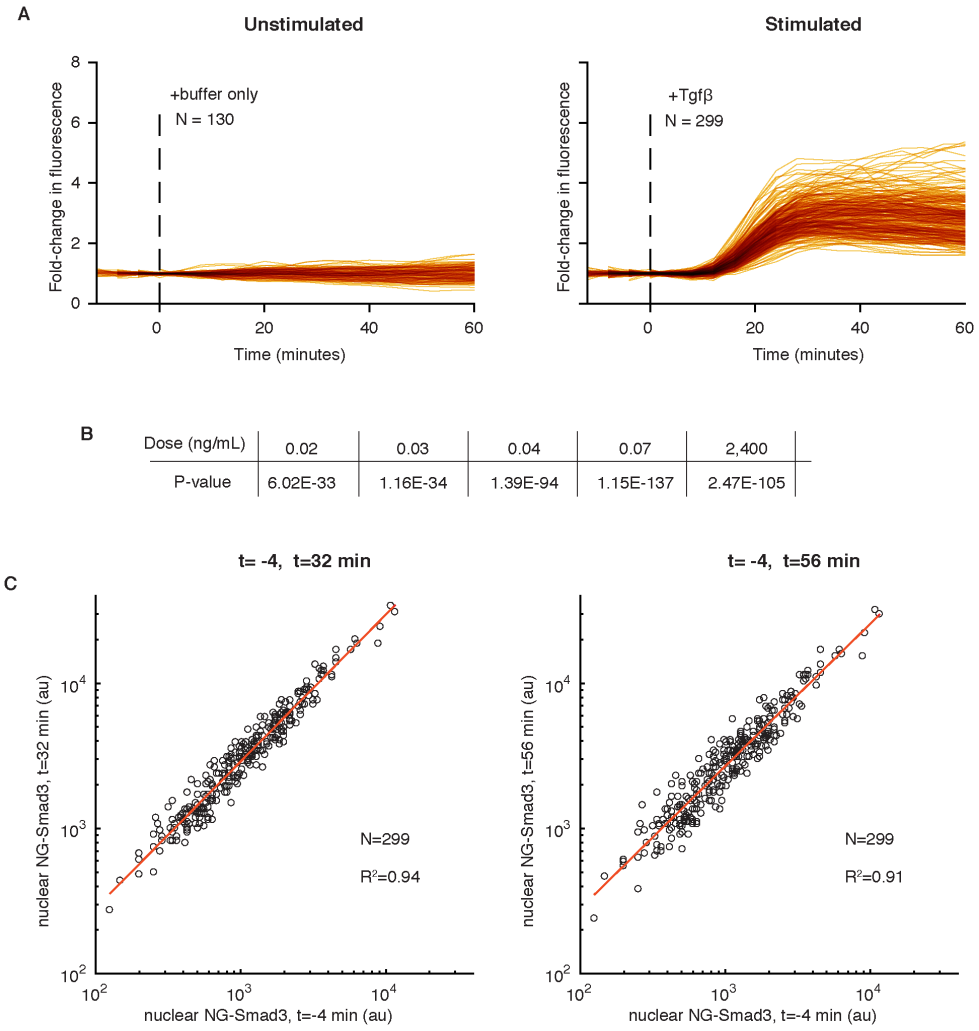

Figure 2.10: Smad3 responses to stimulation are statistically different from unstimulated cells. (*A*, *Left*) We confirmed that cells only exposed to buffer showed no response. (*A*, *Right*) Cells exposed to 2.4 ng/mL Tgf-β show an increase in nuclear Smad3. (*B*) Exposure to all doses of Tgf-β tested produces Smad3 responses that are statistically different from unstimulated cells. Distributions of unstimulated cells at t = 60 min were compared with distributions of cells stimulated with the indicated doses of Tgf-β using the Student's t test. P values are shown in the table. (*C*) Nuclear Smad3 level poststimulation correlates linearly with the basal nuclear Smad3 level. Each data point comes from a single cell. The entire plot comes from measurements of 299 cells. Shown on the y axis is the level of nuclear NG-Smad3 32 min (*Left*) and 56 min (*Right*) after ligand stimulation. The same linearity is observed at other time points. R^2 is the square of Pearson's correlation coefficient.

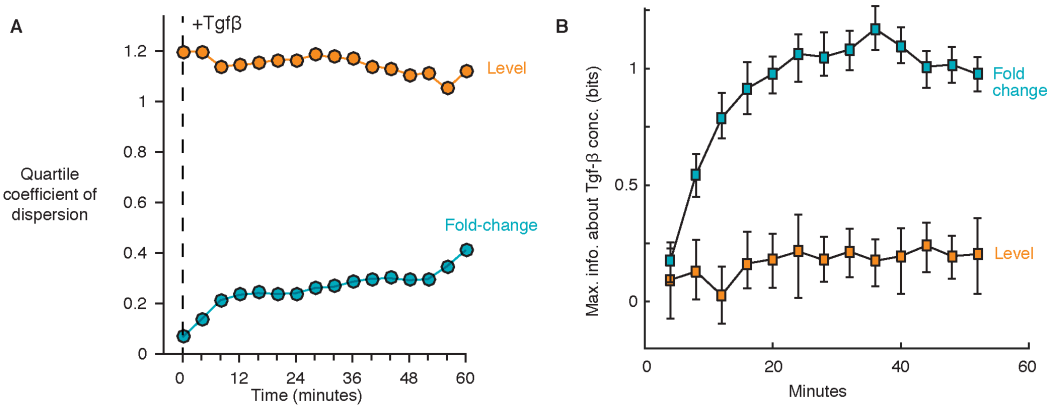

Figure 2.11: Fold-change response is more precise and has higher information capacity throughout the duration of signaling. (*A*) Plotted is the quartile coefficient of dispersion (QCD) for the level (orange) and the fold change (blue) in the level of nuclear NG-Smad3. The QCD is defined here as follows: (Q3 - 1)/Q2, where Q1, Q2, and Q3 are the 25th, 50th, and 75th percentiles, respectively. The QCD values reported here are computed using the cell traces in Figure 2.2 A and B. (*B*) Maximum mutual information between Tgf-β input and nuclear Smad3 level (orange) or fold change (blue) was determined at each time interval after Tgf-β addition (t = 0). Error bars are 90% confidence intervals computed using bootstrap resampling. The total number of cells examined for each calculation was 1,650.

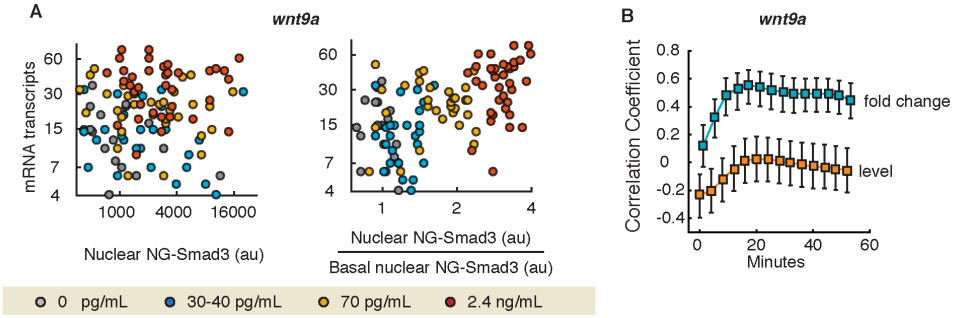

Figure 2.12: Expression of *wnt9a* correlates more strongly with fold change in nuclear NG-Smad3. Plotted is the correlation between NG-Smad3 responses and *wnt9a* mRNA transcripts counted in the same cell. (*A*) Level of nuclear NG-Smad3 (*Left*) and the fold change in nuclear level of Smad3 (*Right*) plotted versus the number of *wnt9a* mRNA counts per cell. The mRNA transcripts were counted at 1 h after ligand stimulation as described in the main text. (*B*) Plotted is the correlation between *wnt9a* mRNA transcripts (at 1 h after ligand stimulation) and NG-Smad3 response measured throughout the entire signaling dynamics. The correlation coefficient is Spearman's rank correlation coefficient. Error bars are 90% confidence intervals, computed using bootstrap resampling. At each time point after ligand addition, the correlation with fold change (blue) is significantly higher than the correlation with level (orange) ($P < 0.01$, Steiger's Z test; Table 2.1).

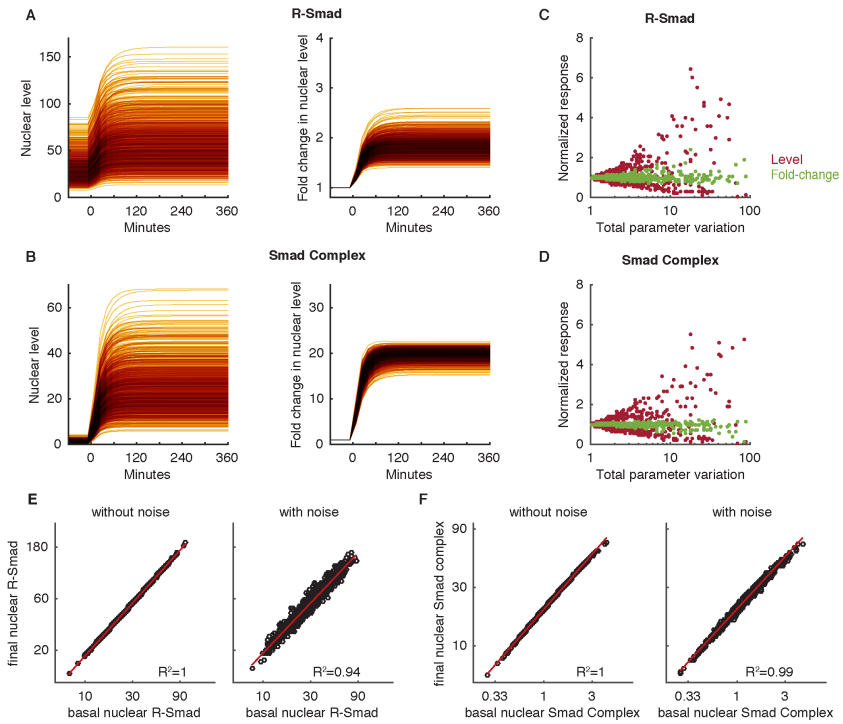

Figure 2.13: Mathematical model of Tgf-β pathway predicts fold-change robustness to parameter variation. We tested whether a mathematical model predicts robustness of Smad fold change to endogenous cellular variability. The mathematical model used is a system of ODEs describing R-Smad nucleocytoplasmic shuttling in response to Tgf-β (Schmierer et al., 2008). To test for robustness to cell-to-cell variation, we performed 1,000 simulations with random parameter variation. (*A*) Response dynamics of nuclear R-Smad level (*Left*) and fold change (*Right*) from 1,000 individual simulations with random parameter variation. (*B*) Response dynamics of nuclear Smad complex level (*Left*) and fold change (*Right*) from 1,000 individual simulations with random parameter variation. Plots of the normalized steady-state response (absolute level, red; fold change, green) vs. the total parameter variation for R-Smad (*C*) and the Smad complex (*D*) are shown. The responses were normalized to the median of all 1,000 simulations as a way of facilitating comparison. Each dot corresponds to an individual simulation. Plots of basal nuclear vs. final nuclear level of Smad3 (*E*) or the Smad complex (*F*) are shown. Each data point represents one simulation of the Tgf-β model either without (*Left*) or with (*Right*) random parameter variation. In each plot are data from 1,000 simulations. R^2 is the square of Pearson's correlation coefficient. Simulations of the system of ordinary differential equations (ODEs) were performed in MATLAB using the numerical solver, ode15i. To perform parameter variation, parameter values were chosen randomly in each iteration of the simulation from a log-normal distribution whose mode is the parameter value of the published model. The log-normal distributions were generated by multiplying the parameter by a randomly generated log-normal distribution with mu equal to 0 and sigma equal to 0.1, except for the expression of R-Smad, where sigma was 0.4 (mu and sigma are the log-mean and log-standard deviation of the log-normal distribution, respectively). We began all simulations with a basal level of Tgf-β (i.e., 0.02 nM), followed by stimulation with 0.5 nM. Total parameter variation, k, is defined as: $\log(k) = \sum_{n=1}^{L} |\log(k_n/k_n^0)|$, where k_n^0 are the published biochemical parameters of the model and k_n are the biochemical parameters of the altered system.

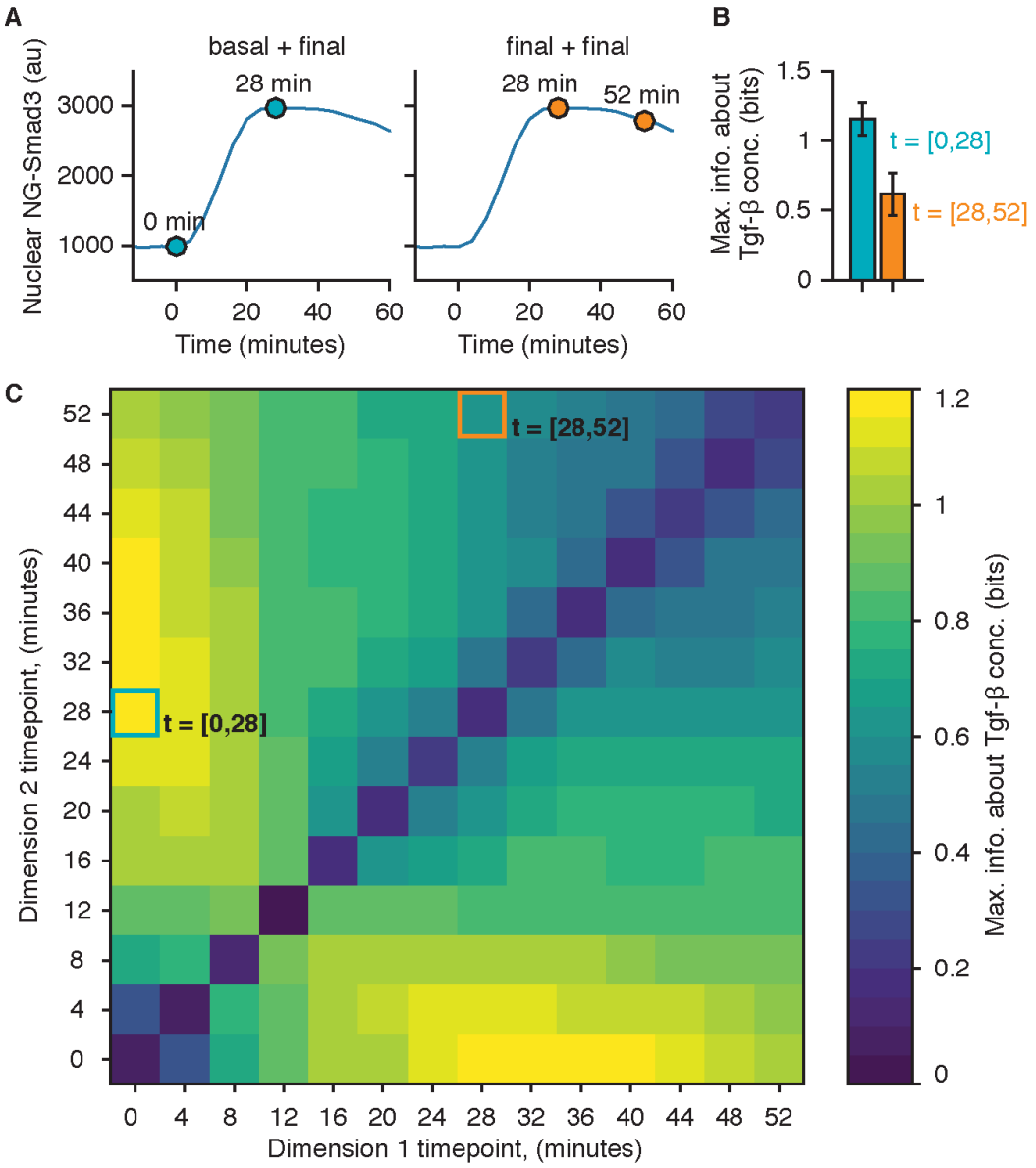

Figure 2.14: Mutual information computed with dynamic measurements. (*A*) Illustration of dynamic measurement. In dynamic measurement, the signaling response of a cell is defined as a vector containing multiple time points (e.g., t = [0,28 min] [*Left*] and t = [28,52 min] [*Right*]) instead of a single time point (as described in ref. Selimkhanov et al., 2014). (*B*) Maximum mutual information between NG-Smad3 response and Tgf-β input using the specified dynamic measurement time point combinations. The data used for the information calculation are the same used in Figure 2.4 (n = 1,650). Maximum mutual information was computed as described in Materials and Methods, with R in Equation 2.1 as the multivariate. Each entry in R contains two time point measurements. (*C*) Shown is a matrix of the maximum mutual information between NG-Smad3 response and Tgf-β input for all time-point combinations. The x axis is the first time point, and the y axis is the second time point. The boxes indicate the location of the time point combinations from *A* and *B*.

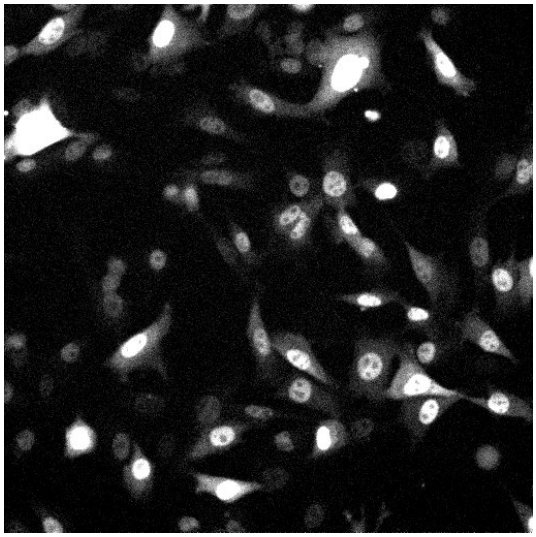

Figure 2.15: MOVIE: NG-Smad3 cells responding to Tgf-β1. Tgf-β1 is added between t = -4 min and t = 0 to a final concentration of 2.4 ng/mL. Images were acquired at 4-min intervals.

2.5 Materials and Methods

Expression Construct.

The human Smad3 cDNA was a gift from Joan Massague (Addgene; plasmid 27010). Human Smad3 and mouse Smad3 contain 100% sequence identity. The mNeonGreen (NG) gene was obtained from Allele Biotechnology (ABP-FP-MNEONSA). The mCerulean3-C1 cDNA was a gift from Klaus Hahn (Addgene; plasmid 22030). The NG-Smad3 construct was placed downstream of a CMV promoter, and the mCerulean3 gene was fused with a 3x nuclear localization sequence (NLS) and placed downstream of an SV40 promoter.

Cell Culture.

C2C12 cells (American Type Culture Collection, CRL-1772) were cultured at 37 °C and 5% (vol/vol) CO2 in DMEM (ThermoFisher Scientific; 11995) supplemented with 10% (vol/vol) FBS (Invitrogen; A13622DJ), 100 U/mL penicillin, 100 μg/mL streptomycin, 0.25 μg/mL amphotericin, and 2 mM L-glutamine (Invitrogen). To generate the NG-Smad3 C2C12 cell line, cells were transfected with the NG-Smad3 plasmid using FuGene 6 reagent (Promega; E2693). Stable expression was selected for using puromycin at a concentration of 2 μg/mL. Cells were sorted using FACS and then plated in a 96-well plate to select single clones. NG-Smad3 cells were maintained in media containing 2 μg/mL puromycin.

Live-Cell Imaging.

Cells were grown on 24-well glass-bottomed plates (Griener Bio-One; 662892) overnight before imaging in 2 mL of FluoroBrite DMEM (Gibco-Life Technologies; A18967) containing 10% (vol/vol) FBS, 1 mM glutamax (Gibco-Life Technologies; 35050), 1 mM sodium pyruvate, 100 U/mL penicillin, 100 μg/mL streptomycin, and 0.25 μg/mL amphotericin. Cells were imaged using a Zeiss Axio Observer.Z1 inverted fluorescence microscope under incubation [37 °C and 5% (vol/vol) CO2, with humidification] on a motorized stage. In each experiment, 21–38 positions were imaged and focus was maintained using Zeiss Definite Focus. Images were acquired at 4-min intervals with a 20x, 0.8-N.A. Plan Apo objective and Evolve 512 EM-CCD camera (Photometrics). Cells were imaged for at least 1 h before stimulation with Tgf-β1 (PeproTech; 100-21). One hundred microliters of Tgf-β–containing media was added to cells to achieve desired final concentrations of Tgf-β in cell growth media for experiments. Buffer-only media were added in 0 ng/mL experiments to control for any effects of adding liquid (e.g., shear stimulation).

Image Analysis, Cell Tracking, and Fluorescence Quantification.

Time-lapse movies were quantified after flat-field correction, bleaching correction, and background subtraction. We followed the standard protocol described by Waters (Waters, 2009). In flat-field correction, to capture the shape of fluorescence illumination, we imaged a well containing media only. We imaged five different positions within the well, and computed the median of the images. Flat-field correction was performed by dividing each experimental image by this media-only image. This procedure was repeated for each fluorescence channel. Bleaching correction was performed for each fluorescence channel by correcting for the global change in fluorescence throughout the duration of imaging. For background correction, images were segmented such that the entireties of cells were broadly outlined, and fluorescence signal from the background was then averaged and subtracted from the image. This procedure was repeated for all images at each time frame.

Fluorescence Quantification.

We report the median fluorescence intensity of NG-Smad3 fluorescence in the nuclei. The nuclei of cells were first segmented based on the fluorescence of the constitutively expressed mCerulean3-3NLS (3xNLS). Next, segmented nuclei were tracked across all time frames. Finally, the fluorescence data from the segmented nuclei were extracted. We only tracked and quantified fluorescence from cells that maintained consistent morphology for at least 1 h of imaging. Cells that divided, balled up, left the imaging field of view, or displayed some abnormality (e.g., double-nucleated, abnormally large) were excluded. For correcting experimental fluctuations during the imaging period, we used the constitutive mCerulean3 fluorescence as an internal control, dividing each time trace for NG-Smad3 nuclear fluorescence by the normalized mCerulean fluorescence time trace in the same cell. Subsequently, individual time traces were smoothed using a running three-frame average. mCerulean normalization and time trace averaging turn out to be minor corrections (likely because our tracked cells maintained consistent morphology and our imaging setup was stable during imaging duration), and we obtained the same conclusions both with and without these corrections applied (raw data are available upon request). We performed all segmentation, tracking, and fluorescence quantitation steps using the Lineage Tracker ImageJ (NIH) plug-in (Downey et al., 2011) and custom MATLAB (MathWorks) scripts (available upon request).

Single-Molecule FISH.

Following time-lapse imaging, cells were fixed using 4% (wt/vol) paraformaldehyde for 20 min, permeabilized in 70% (vol/vol) ethanol for at least 1 d at -20 °C, and then hybridized overnight with HPLC-purified single-molecule FISH (smFISH) probes at 30 °C using a protocol adapted from Raj et al. (Raj et al., 2008). Images of stained cells were acquired using a 40x, 1.4-N.A. Plan Apo Oil Objective with Immersol 518F (Zeiss; 444960) and an Orca Flash 4.0 V sCMOS camera. To ensure the entirety of each cell was imaged at each position, a z-stack of 20 or more images was collected at 0.6-μm intervals. Differential interference contrast (DIC) images from the middle Z-slice of image stacks were used for manual segmentation of cells. The mRNA foci were detected using custom MATLAB scripts. Briefly, fluorescence images were convolved with a 5 x 5 Laplacian-of-Gaussian kernel, and then thresholded such that mRNA foci were only identified in cells (scripts available upon request). Probe sets targeting *snai1*, *wnt9a*, and *ctgf* mRNA were designed using Stellaris Probe Designer Version 4.1 and ordered from Biosearch. Each probe is a 20-mer with a mdC(TEG-Amino) 3' modification, which was used to couple the probe to Alexa Fluor 594 NHS Ester (Molecular Probes; A20004) or Alexa Fluor 647 NHS Ester (Molecular Probes; A20006). Following the coupling reaction, fluorescently labeled probes were purified using HPLC. The smFISH probe sequences are provided in Table 2.2.

Mutual Information Estimation.

To estimate the mutual information between Tgf-β input and NG-Smad3 response, we followed the steps described in the methods of Voliotis et al. (Voliotis et al., 2014). Mutual information is expressed as follows:

$$I(R;S) = h(R) - h(R|S) = h(R) - E[h(R|S = s)], \tag{2.1}$$

where R, the pathway output, is continuous and possibly multivariate, and S, the ligand input, is the distribution of ligand concentrations. R is a vector containing experimentally determined responses (e.g., a vector of fold-change responses at t = 32 min or a vector of nuclear NG-Smad3 levels at t= 32 min), and S is a vector containing the probabilities of ligand doses. The unconditional entropy, $h(R)$, and the conditional entropy, $h(R|S = s)$, are estimated using the nearest-neighbor (knn) method [i.e., equation 20 of Kraskov et al. (Kraskov, Stögbauer, and Grassberger, 2004)], which is reproduced below:

$$\hat{H}(X) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N}\sum_{i=1}^{N} \log \varepsilon(i). \tag{2.2}$$

The knn method performs better than "binning with bias correction" (used in ref. (Cheong et al., 2011)), at a smaller sample size (n < 200), giving more accurate estimations of mutual information with smaller mean squared error and bias (Voliotis et al., 2014). We use $k = 3$ nearest neighbors for all estimations performed in this work. For each calculation of mutual information, we performed 100 iterations of random sampling (without replacement) of the dataset. We confirmed that the distributions of the mutual information estimator are similar for 100 and 1,000 iterations. To determine maximum mutual information, we tested 100 different signal input distributions, S, ranging from uniform, to unimodal, to bimodal, to trimodal, and determined the signal at which mutual information was maximum. We report in Figure 2.4D and Figure 2.14 the maximum mutual information. To compute the mutual information using dynamic measurement (Selimkhanov et al., 2014), R is multivariate, with each entry in R (corresponding to a single cell) containing two time point measurements.

2.6 Supplementary Tables

	snai1 (n = 125)					*ctgf* (n = 114)					*wnt9a* (n = 119)				
Time, min	r_{13}	r_{23}	r_{12}	Z	P value	r_{13}	r_{23}	r_{12}	Z	P value	r_{13}	r_{23}	r_{12}	Z	P value
4	0.18	-0.12	0.21	2.616	0.0075	0.31	-0.03	-0.02	2.634	0.0072	0.32	-0.20	0.05	4.171	1.08E-05
8	0.34	-0.04	0.18	3.410	4.14E-04	0.49	0.06	-0.08	3.406	4.51E-04	0.48	-0.12	-0.04	4.814	2.81E-07
12	0.43	0.03	0.27	3.837	6.00E-05	0.51	0.15	0.03	3.073	0.0016	0.53	-0.04	0.03	4.797	3.20E-07
16	0.50	0.09	0.35	4.309	5.08E-06	0.55	0.21	0.13	3.145	0.0013	0.56	0.01	0.11	4.862	2.04E-07
20	0.53	0.13	0.39	4.417	2.73E-06	0.56	0.21	0.19	3.265	7.81E-04	0.54	0.03	0.17	4.690	5.71E-07
24	0.56	0.16	0.39	4.454	2.30E-06	0.56	0.21	0.22	3.343	5.66E-04	0.52	0.02	0.19	4.555	1.26E-06
28	0.57	0.20	0.37	4.166	1.21E-05	0.56	0.20	0.23	3.472	3.29E-04	0.51	0.01	0.20	4.569	1.12E-06
32	0.57	0.22	0.34	3.893	4.97E-05	0.55	0.19	0.24	3.467	3.32E-04	0.49	0.00	0.20	4.555	1.19E-06
36	0.58	0.23	0.31	3.772	9.02E-05	0.54	0.18	0.24	3.404	4.30E-04	0.49	-0.02	0.21	4.685	5.26E-07
40	0.57	0.23	0.29	3.647	1.59E-04	0.53	0.17	0.22	3.366	5.05E-04	0.49	-0.03	0.19	4.726	4.10E-07
44	0.58	0.23	0.28	3.678	1.40E-04	0.52	0.17	0.19	3.254	8.00E-04	0.49	-0.03	0.16	4.669	6.07E-07
48	0.57	0.24	0.27	3.522	2.78E-04	0.49	0.15	0.16	3.073	0.0016	0.48	-0.04	0.13	4.554	1.25E-06
52	0.56	0.23	0.28	3.512	2.88E-04	0.46	0.14	0.14	2.756	0.0049	0.45	-0.05	0.12	4.320	4.88E-06

Table 2.1: Correlations, z-scores, and p-values from Steiger's Z test for Figures 2.5 and 2.12. Here we are testing whether the correlation between fold-change and mRNA count is significantly different from the correlation between level and mRNA count. Steiger's Z test is used to determine whether two non-independent correlations are significantly different (Steiger, 1980). Correlations are non-independent when they share a term. In this case the fold-change and level are both correlated with the same term, mRNA count data. The correlations in the table are the following: r_{13} is the correlation between fold-change and mRNA count, r_{23} is the correlation between level and mRNA count, and r_{12} is the correlation between fold-change and level, which is not the question of interest, but is necessary to restrict the level of deviation between the two correlations.

snai1		*ctgf*		*wnt9a*	
probe#	Probe (5'>3')	probe#	Probe (5'>3')	probe#	Probe (5'>3')
1	gcaaggtcggtagtcaactc	1	tcggagctggagtggatctg	1	gtcagagggaggatagtcag
2	cagagcgacctaggtagtcg	2	gtaggaggatgcacagcagg	2	aactggtactggcactccag
3	ctgcagctcgctatagttgg	3	agagagcgaggagcaccaag	3	tgcagttccagcgctcaaag
4	gaaggtgaactccacacacg	4	agccgaaatcgcagaagagg	4	gaaagcagtctccttgaagc
5	aggagagagtcccagatgag	5	cagtgcacactccgatcttg	5	cggcagaagagatggcgtag
6	tcatcggacagcgaggtcag	6	aagacacagggtgcaccatc	6	ctgctgtacttgaggttgtc
7	aggtggacgagaaggacgac	7	tttggaaggactcaccgctg	7	ccaggaactccttgacaaac
8	aagccagggaaggcgatgaa	8	caaatgtgtcttccagtcgg	8	ctcgcaaatccttgctagag
9	ttgcagttgaagatcttccg	9	tcgcatcatagttgggtctg	9	acgaggttgttgtggaagtc
10	tgaggtactccttgttacaa	10	cacaggtcttagaacaggcg	10	agcctttatcaccttcacac
11	tcggatgtgcatcttcagag	11	tggtaactcgggtggagatg	11	atggcatttgcaagtggttt
12	cagacacaaggcagcgtgtg	12	cagtctgcagaaggtattgt	12	agcgaggtctcatatttgtg
13	tagagaaggcctttccacag	13	cttcttaatgttttcctcca	13	cagtggcttcattggtagtg
14	cagtgggagcaggagaatgg	14	gtgtccggatgcacttttttg	14	tgtacaagctctggtgttcg
15	gttggagcggtcagcaaaag	15	aaacttgacaggcttggcga	15	cagaagctgggagagtcgtc
16	cactggtatctcttcacatc	16	acactggtgcagccagaaag	16	cggccacaacaaatactctc
17	caaggacatgcgggagaagg	17	agaacttagccctgtatgtc	17	acacgactctgtgtgttgtg
18	cagactcttggtgcttgtgg	18	atttgaactccactggcaga	18	tcctctctctgtgtacactg
19	gaagatgccagcgaggatgg	19	tttttcatgatctcgccatc	19	tgctgtgtgcaaagtccaca
20	gaccaaggctggaaggagtc	20	aggcacaggtcttgatgaac	20	tgcaggtgtagaattcttct
21	aggaaccaggtccagacatg	21	actcaaagatgtcattgtcc	21	ctacactccctgcagaaagg
22	ggcaaaggccaccaagagag	22	gtacatcttcctgtagtaca	22	ccctttgcactatgtgcaaa
23	gtacaaaggcactccatcag	23	agtctaatgagttcgtgtcc	23	gctcactccatgcagattaa
24	acctcatgaatactgagggg	24	cctcccacggtagttaaaaa	24	cacagcaggataggctggac
25	cgaagcagctgtgtccagag	25	ttctcactttggtgggatag	25	ctgtgttgtttgtaaccctg
26	tgggttggctttagttctat	26	acttgtatggccatgacata	26	acctcttcatccttgtagac
27	ttgagggaggtagggaagtg	27	cagtgtctgaggttgacaga	27	gcccaaggcataagcaaaag
28	tgtacctcaaagaaggtggc	28	gtcaagtgtaaactgtctcg	28	aagcattactgcaacgctct
29	ggggaactattgcatagtct	29	gcactgtgcgctaatgaaca	29	tcagggacagaggcaactga
30	ctgaggcatggttacagctg	30	cactgttccaggagactcac	30	cctgacaagggctatactga
31	ctggacatgtgtccagtaac	31	ggcacactgctgcttttaaa	31	acagagggccagtcaacatg
32	aggtgtcaccaggacaaatg	32	aaagaacagctggactcagc	32	atgcctcagaggaagccaag
33	gacagttaaactgctgtgaa	33	atggtcagagctgaaactgc	33	cgattatgaaggtgcacctg
34	cattattcatggtcccttct	34	cctgacaagtgtcactggaa	34	ggctctctgaagaagacagt
35	ttggcccctaacaagtgatg	35	gtctaacagacaaggctctg	35	actgatactccacggaagtg
36	acacattggccaggctgaag	36	acttacttgccacaagctgt	36	gaagagagggaaggcaaggc
37	tgttgggccccaaaatagtt	37	aaatctggcttgttacaggc	37	tagagcaagcaaaggagggg
38	atgtaaacatctttctcccg	38	ggctatcagtttaaaatccc	38	tggattccagaagccttgtg
		39	acctatggtgtttggagttt	39	ctggggaaggtagaattggg
		40	tcacagataagcttcgtgtc	40	ctatatccagcaacctgatc
		41	gcagtatctccttttgttttg	41	cttggggacacagagcaaga
		42	gagtcactcaggtcacaatt	42	gatgacagcaagtgtgtctg
		43	tgcacagcatttgttctgac	43	cctgtagagacatcagcaag
		44	ctgacttccaatacatagct	44	agctcagtgtagtcaccttt
		45	tgaccacatttcctactaga	45	caactggatgctgaccatgc
		46	catttgttcaccaacaggga	46	ataccgcagagggaacacag
		47	taggaatcggaccttaccct	47	tttctgggagccaaagagtc
		48	cttttggtcacactctcaac	48	tgggttcccacaaacctaag

Table 2.2: Sequences for smFISH probe sets.

References

Cheong, Raymond et al. (2011). "Information transduction capacity of noisy biochemical signaling networks". In: *science* 334.6054, pp. 354–358.

Downey, Mike J et al. (2011). "Extracting fluorescent reporter time courses of cell lineages from high-throughput microscopy at low temporal resolution". In: *PloS one* 6.12, e27886.

Kraskov, Alexander, Harald Stögbauer, and Peter Grassberger (2004). "Estimating mutual information". In: *Physical review E* 69.6, p. 066138.

Raj, Arjun et al. (2008). "Imaging individual mRNA molecules using multiple singly labeled probes". In: *Nature methods* 5.10, p. 877.

Schmierer, Bernhard et al. (2008). "Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system". In: *Proceedings of the National Academy of Sciences* 105.18, pp. 6608–6613.

Selimkhanov, Jangir et al. (2014). "Accurate information transmission through dynamic biochemical signaling networks". In: *Science* 346.6215, pp. 1370–1373.

Steiger, James H (1980). "Tests for comparing elements of a correlation matrix." In: *Psychological bulletin* 87.2, p. 245.

Voliotis, Margaritis et al. (2014). "Information transfer by leaky, heterogeneous, protein kinase signaling systems". In: *Proceedings of the National Academy of Sciences* 111.3, E326–E333.

Waters, Jennifer C (2009). *Accuracy and precision in quantitative fluorescence microscopy*.

Chapter 3

A CELL'S RESPONSE TO SMAD3 SIGNAL DEPENDS ON ITS INTERNAL STATE

3.1 Introduction

The functioning of a multicellular organism relies on signaling between cells. Complex processes including embryonic development and tissue homeostasis involve a multitude of decisions made by cells in response to signals secreted by other cells. It is therefore important to understand how a cell responds to signal.

In a typical signaling process, cellular response depends on two transduction processes (Figure 3.1A). First, extracellular signal (usually called ligand) is transduced into activation of intracellular signal. Subsequently, intracellular signal is transduced into cellular response, typically regulation of specific genes. In this study, we focus on the transduction of intracellular signal (input) into gene response (output) (Figure 3.1B).

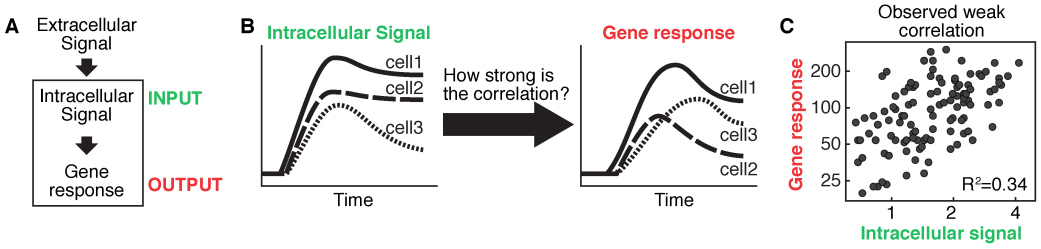

Figure 3.1: To what extent does signal dictate a cell's response? (A) In a typical signaling process, extracellular signal is transduced into intracellular signal, which is then transduced into gene response. (B) This study analyzed in single cells the input-output correlation between the dynamics of intracellular signal and gene response. (C) Where it has been measured, the correlation between signal input and gene output is unimpressive. To illustrate, this dataset is reproduced from Frick et al., 2017, where signal input is fold change of Smad3 (x-axis), gene output is snail mRNA count (y-axis), and R is the Spearman's rank correlation coefficient.

Despite the typical flow of signal transduction, it is well established that gene response varies across tissues and processes. For instance, Tgf-β/Smad3 signaling activates fibrogenic genes (e.g., collagen) in fibrosis (F. Xu et al., 2016), represses

insulin in pancreatic islet β-cells (H.-M. Lin et al., 2009), and activates different cell markers in different immune cells (Malhotra and Kang, 2013). Different gene response across tissues and processes has been attributed to context-specific differences, such as epigenetic state, morphology, microenvironment, and tissue-specific co-factors (Mullen et al., 2011; David, Huang, et al., 2016; Uttamsingh et al., 2008; David and Massagué, 2018; Derynck and Budi, 2019).

Less understood, however, is whether, and to what extent, gene response varies across cells within a clonal population. In clonal populations, many studies have focused on quantifying the variability by which ligand activation is transduced into intracellular signal (Strasen et al., 2018; J. Yao, Pilko, and Wollman, 2016; Sero et al., 2015). Variability in the downstream gene response has also been observed, but it has been typically attributed to the variability in the intracellular signal dynamics. In spite of which, where it has been measured, the correlation between intracellular signal and gene response, bafflingly, is weak (e.g., Figure 3.1C; Gillies et al., 2017; Wong et al., 2018; Rand et al., 2012; Tidin et al., 2019; Paek et al., 2016; Reyes et al., 2018; Sung et al., 2014; Frick et al., 2017; Purvis et al., 2012; R. E. Lee et al., 2014).

Motivated by this puzzle, we set out to systematically quantify the extent to which intracellular signal dynamics is transduced into gene response dynamics (Figure 3.1B). We identified three technical factors in present studies that may contribute to the measured variability in gene response. First, single-cell imaging typically relies on tracking fluorescently labeled signal expressed from exogenous DNA transfected to the cells. This means we are not tracking the full population of signal, as the endogenous signal remains unlabeled. To address this problem, we used genome editing to label the endogenous signal. Second, gene expression is often measured at fixed time points (e.g., using single-molecule FISH), such that temporal heterogeneities, initial expression level, and full dynamics cannot be accounted for. To address this problem, we used genome editing to label an endogenous target gene locus. Finally, even in studies that avoid these pitfalls, correlation analysis has so far been done on fixed-point features of dynamics (e.g., steady-state level or fold change of signal versus mRNA counts of target gene, in Frick et al., 2017; R. E. Lee et al., 2014). To analyze the correlation between the full input-output dynamics, we adopted a technique from data science: nonlinear manifold learning, which was developed to address the challenge of high-dimensional datasets.

We pursued this effort in the Tgf-β pathway (Figure 3.2 A). In the Tgf-β path-

way, ligand activation is transduced into phosphorylation of the receptor-regulated Smads, Smad2 and Smad3. Phosphorylated Smad2 and Smad3 form complexes with the common-Smad, Smad4. The Smad heteromeric complexes shuttle into the nucleus, bind to chromatin in conjunction with cell type-specific factors, resulting in regulation of specific gene expression. We quantified in single living cells the input-output dynamics in the Tgf-β pathway, and asked: Within a clonal population of cells, how strongly does Smad dynamics predict target gene response dynamics?

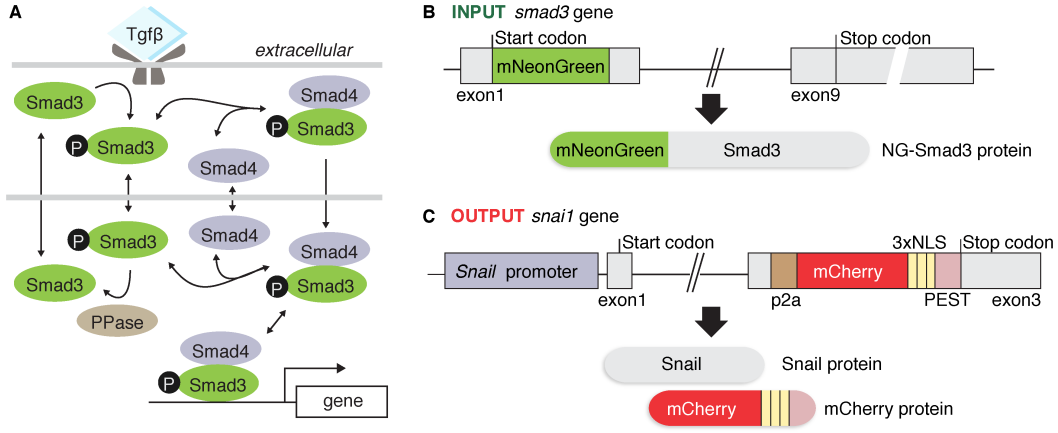

Figure 3.2: Quantifying input-output dynamics in the Tgf-β pathway in single cells. (A) Binding of Tgf-β to membrane receptors leads to translocation of Smad complex into the nucleus, where it regulates specific genes. (B-C) A double CRISPR system to track input-output dynamics in single cells. (B) To track signal input, we knocked mNeonGreen into the N-terminus of *smad3*. (C) To track gene output, we knocked p2a-mCherry-3xNLS-PEST into the C-terminal end of *snai1*. The p2a signal directs co-translational cleavage leading to production of mCherry-3xNLS-PEST in one-to-one ratio with Snail.

3.2 Results

A double CRISPR system to quantify endogenous input-output dynamics in single cells

We studied Tgf-β signaling in mouse myoblast C2C12, a cell system with intact Tgf-β pathway, where Smad signaling and its direct target genes have been characterized (D. Liu, Black, and Derynck, 2001; Warmflash et al., 2012). To track signal input, we focused on Smad3, the dominant effector of Tgfβ in C2C12 cells (D. Liu, Black, and Derynck, 2001; Yagi et al., 1999). To label Smad3, we used CRISPR/Cas9 to knock mNeonGreen (NG) into the N-terminus of smad3 (Figure 3.2B). We chose mNeonGreen because it is monomeric, matures rapidly (<10 min), and 2.7 times brighter than EGFP, which is important as endogenous protein levels

are typically much lower than exogenously expressed proteins (Shaner et al., 2013; Balleza, J. M. Kim, and Cluzel, 2018). We labeled Smad3 at the N terminus, as it preserves Smad3 functions, i.e., phosphorylation, nuclear import, binding to DNA and transcriptional activity (X. Liu et al., 1997; Nicolás et al., 2004; Frick et al., 2017). N-terminal labeling, however, does not allow the use of antibiotics. We therefore carefully optimized various aspects of the transfection and recombination (detailed in Methods), and used cell sorting to enrich for fluorescent cells. After several rounds of sorting of more than ten million cells, we obtained 1 heterozygous and 4 homozygous clones (Figure 3.8A and B). We validated that the endogenous NG-Smad3 behaves as expected. Upon Tgf-β1 stimulation, NG-Smad3 is phosphorylated and translocates to the nucleus (Figure 3.8C and D). All 5 clones show highly similar dynamics of Smad3 translocation, with time scales consistent with previous reports (Figure 3.8F; Hill, 2009; X. Lin et al., 2006; Yingling et al., 1996; Lönn et al., 2010; Y. Li et al., 2018; Strasen et al., 2018; Lo and Massagué, 1999). The clones also recapitulate further subtle regulation, e.g., as in Smad3, NG-Smad3 is excluded from the nucleolus (see Figure 3.3A and 3.8D; Chen et al., 2005).

Next, to track the gene output, we focused on Snail, a master regulator of epithelial-mesenchymal transition (EMT). Snail is a direct target of Smad3: Smad3 binds directly to the snai1 promoter (Hoot et al., 2008; Brandl et al., 2010; Thuault et al., 2008) and knocking out Smad3 abolishes Snail expression (Zavadil et al., 2004; Sato et al., 2003; Ju et al., 2006). To track Snail activation, we began with one homozygous NG-Smad3 clone, and performed another round of CRISPR/Cas9 to knock mCherry into snai1 (Figure 3.2C). We used bicistronic strategy, using the p2a sequence to direct cleavage of mCherry from Snail during translation (Osborn et al., 2005). This strategy leads to mCherry production in one-to-one ratio with Snail. A nuclear localization sequences (NLS) was used to concentrate mCherry signal and facilitate nuclear segmentation. A degradation signal (PEST, Rechsteiner and Rogers, 1996) was used, since destabilized fluorescent reporters track dynamics more rapidly (X. Li et al., 1998). After sorting of more than ten million of cells and analyzing more than 500 clones, we obtained 6 snail:mCherry clones (Figure 3.9A and B). We never found homozygous snail:mCherry clones. Fortunately, the heterozygous reporter clones show sufficiently high signal-to-noise ratio that enables tracking Snail activation. We validated the snail:mCherry clones: the 6 clones showed similar response dynamics (Figure 3.9C); and SB-431542, which inhibits Smad2/3 phosphorylation (Inman et al., 2002), suppressed mCherry expression (Figure 3.9D). We present here in-depth analysis in one clone, and have verified the

key findings in another clone.

Endogenous input-output dynamics are weakly correlated

Using the double CRISPR cell system, we analyzed the extent to which signal dynamics dictate gene response dynamics. We used time-lapse fluorescence microscopy to film NG-Smad3 and snail:mCherry dynamics in single living cells (Figure 3.3A, Movie 3.22 and Movie 3.23). Upon ligand stimulation, NG-Smad3 translocates to the nucleus. The nuclear Smad3 level peaks within 30 minutes, and slowly adapts to basal level over 6 hours (Figure 3.3A-B). Some cells interestingly show a second nuclear peak that may or may not be higher than the first peak. In response to NG-Smad3 dynamics, mCherry begins to increase at 1 hour after ligand addition (the delay is consistent with the time scale of protein production and mCherry fluorescence maturation), reaches a maximum in most cells at about 4 hours, and remains elevated for >10 hours (Figure 3.3A, 3.3C).

First, we verified our previous observations. Using exogenous NG-Smad reporter, we previously determined that Snail transcript abundance correlates more strongly with fold change in Smad3 level, rather than absolute level (Frick et al., 2017). Fold change in Smad3 is defined as the ratio between the peak level of nuclear Smad3 relative to the basal, pre-stimulated level (see Figure 3.10A). In cells with more than one peak in Smad3 dynamics, we define fold change using the first peak. Analyzing endogenous NG-Smad3 in this study, we confirmed the fold-change dependence: testing various input and output functions (Figure 3.10), we observed the highest correlation between the amount of snail:mCherry produced and the fold change in Smad3 (Figure 3.3D).

However, even though we are already examining endogenous proteins, the correlation remains rather unimpressive (R = 0.60) – comparable to what was previously observed with exogenous NG-Smad3 (Frick et al., 2017), as well as in studies across signaling systems (e.g., in NF-κB, R. E. Lee et al., 2014, in Erk, Gillies et al., 2017). The weak correlation is even more pronounced when we examined the actual traces within individual cells (Figure 3.3E). There is no readily apparent pattern of correlation between NG-Smad3 and snail:mCherry dynamics. Cells with nearly identical NG-Smad3 dynamics could give significantly different snail:mCherry responses (see cells 5 and 6 in Figure 3.3E), whereas cells with different NG-Smad3 dynamics could show similar snail:mCherry response (see cells 7 and 8 in Figure 3.3E). To verify that the apparent weak correlation is not an artifact of one particular clone, we

confirmed that a similarly weak correlation between NG-Smad3 and snail:mCherry was observed in another clonal line (Figure 3.11).

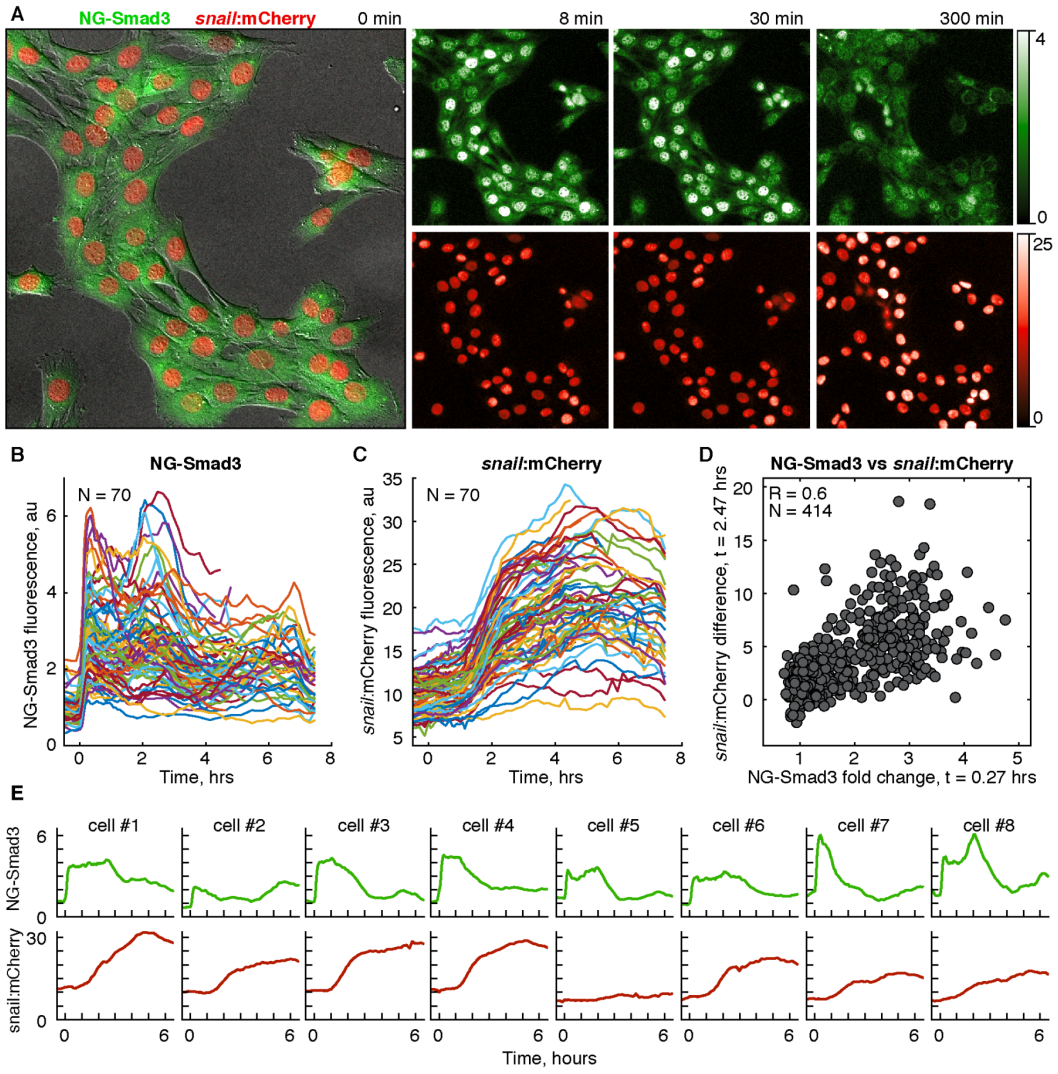

Figure 3.3: Correlation between endogenous input and output dynamics in the Tgf-β pathway is weak. (A) Dynamics of endogenous NG-Smad3 and snail:mCherry upon treatment with Tgf-β1 (2.4 ng/mL). The left panel shows an overlayed image of the single cells–NG-Smad3 fluorescence in green, snail:mCherry fluorescence in red and the DIC transmitted light image in white. The right hand panels show the response of single cells over time. (B) Single cell trajectories of endogenous NG-Smad3 fluorescence in response to Tgf-β1, added at t=0. Fluorescence is quantified as the median nuclear fluorescence. (C) Single cell trajectories of endogenous snail:mCherry fluorescence in response to Tgf-β1, added at t=0. Fluorescence is quantified as the total nuclear fluorescence to capture total production of snail. (D) snail:mCherry expression correlates most strongly with fold change in NG-Smad3. (E) Quantitation of endogenous NG-Smad3 and snail:mCherry dynamics within single cells.

Manifold learning to facilitate analysis of correlation between full input-output dynamics

The weak correlation suggests that, even though Smad3 is the input to snail, Smad3 dynamics does not carry all of the information that explains Snail dynamics. However, in our correlation analysis so far (Figure 3.3D and S3), we have only examined features at fixed time points (e.g., in Figure 3.3D). Recent work, however, argues that entire dynamics, rather than features at specific time points, carries considerably more information (Selimkhanov et al., 2014). While we know how to analyze correlation between variables (e.g., with regression), how would we analyze correlation between dynamics?

One approach for working with high-dimensional dataset is to use manifold learning (J. A. Lee and Verleysen, 2007). Manifold learning is a non-linear dimensionality reduction technique, motivated by the observation that real-world data are often much simpler than the dimensionality suggests. Dimensionality reduction techniques to search for the parsimonious representation of complex data are therefore powerful tools, as they reveal the meaningful relationships in the data and enable the use of statistical tools established in the low-dimensional world, such as regression-based correlation analysis.

Manifold learning algorithms seek for a low-dimensional manifold embedded within the high-dimensional space of the data, that captures the essential features of the data (Figure 3.4A). Manifold learning is different from linear dimensionality reduction techniques (such as Prinipal Component Analysis), in that it allows the manifold to be of any shapes, not just linear ones. Since many natural phenomena are nonlinear, nonlinear techniques often work significantly better than the corresponding linear ones in preserving local structures (Tenenbaum, De Silva, and Langford, 2000).

Of the several nonlinear dimensionality techniques that have been developed so far, we chose the Diffusion Maps, introduced by Coifman and Lafon (Coifman and Lafon, 2006; detailed in Methods). The Diffusion Maps algorithm leverages the relationship between diffusion process and random walk. It embeds the manifold in a lower-dimensional space, where the Euclidean distance between points approximates the diffusion distance in the original space, computed using Markov chain. Compared to other methods, Diffusion Maps has proven to be more robust to noise in the data and computationally inexpensive.

We applied Diffusion Maps algorithm to the measurements of the NG-Smad3 and snai1:mCherry dynamics. For each cell, NG-Smad3 measurements constitute a

48-dimensional dataset (48 time points, sampled from 30 minutes before to 4 hours after ligand addition). We input into the analysis time-lapse trajectories of hundreds of cells (425 cells in experiment 1, and 375 cells in experiment 2). We made no assumptions about which features of the dynamics are most important, and simply used the raw data in the analysis.

The analysis reveals that the NG-Smad3 dynamics can be described by a much lower-dimensional manifold. As Figure 3.4B shows, despite the broad variation across cells, the variability in the NG-Smad3 dynamics lies mostly in 3 dimensions (Figure 3.4C). The first dimension clearly corresponds to the level of NG-Smad3 expression, and the other two dimensions describe the shape of NG-Smad3 dynamics (Figure 3.12). Applying manifold learning to the snai1:mCherry dynamics, we again find that the broad variability in the snai1:mCherry dynamics can be captured by 2 dimensions (Figure 3.4D and Figure 3.12 for trajectories colored by manifold dimensions). We verified that the Diffusion Maps analysis robustly obtained similar manifolds across two independent experiments (Figures 3.14 and 3.15).

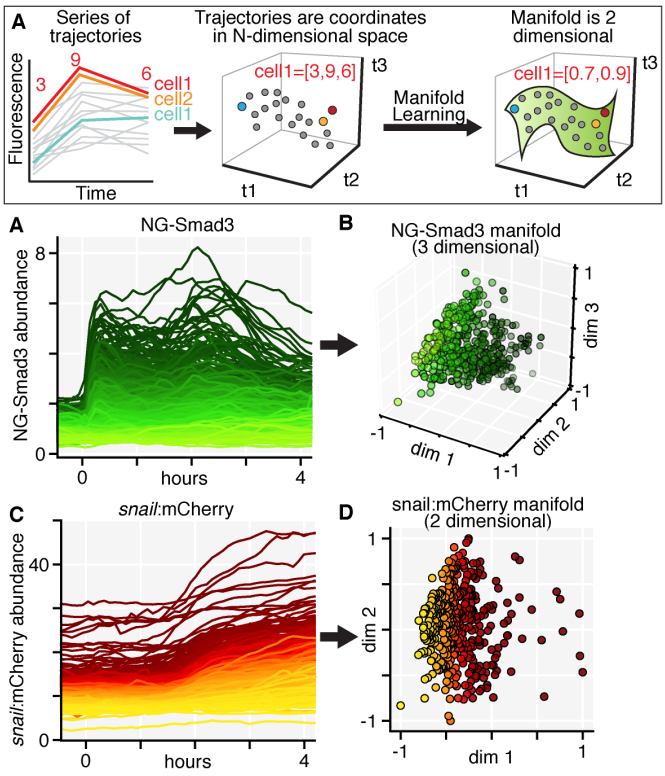

Figure 3.4: Manifold learning to uncover dimensionality of the input-output dynamics. (A) An illustration demonstrating a case where a 3-dimensional dataset can be described using a 2-dimensional manifold (B) Cell-to-cell variability in the NG-Smad3 dynamics lies in a 3-dimensional manifold. (C) Cell-to-cell variability in the snai1:mCherry dynamics lies in a 2-dimensional manifold.

Even when the entire dynamics were analyzed, endogenous input-output dynamics remains weakly correlated

Having uncovered the low-dimensional manifolds that describe the NG-Smad3 dynamics and snai1:mCherry dynamics, we can now assess the correlation between the input and output manifolds. To predict the input-output function from data, we employed supervised learning using the non-parametric Gaussian process regression (Rasmussen, 1999).

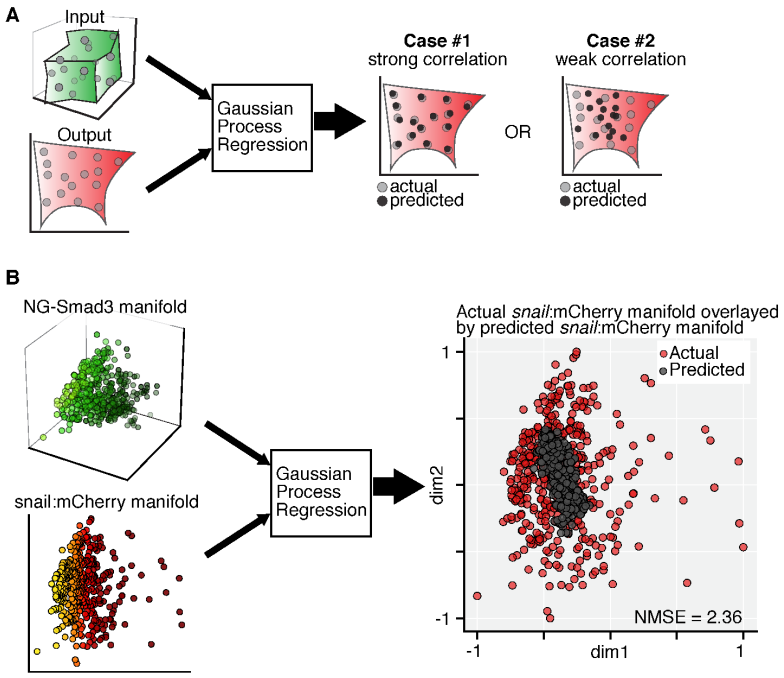

Figure 3.5: The NG-Smad3 manifold can not accurately predict the *snail*:mCherry manifold (A) Illustration of Gaussian Process Regression outcomes for scenario in which input and output are strongly correlated (case#1) and input and output are weakly correlated (case#2). (B). We used Gaussian process regression to analyze the correlation between the NG-Smad3 and snai1:mCherry manifolds. The left plots are the two manifolds, and the right plot is the predicted vs actual snail:mCherry manifold. NMSE is the average normalized mean square error computed using 5-fold cross validation.

Performing Gaussian process regression on the NG-Smad and snail:mCherry manifolds, we find a weak correlation (Figure 3.5). To compute the goodness of the predicted function, we used 5-fold cross validation, and found the normalized mean squared error to be 2.36. To verify that the combined method of manifold learning and Gaussian process regression can recover a strong correlation where there is one, we used the model of the Tgf-β/Smad pathway (Figure 3.16). We simulated cell-to-cell variation by introducing random variation in the biochemical parameters and protein concentrations. We show that if Snail is fully dictated by Smad3, manifold learning followed Guassian process regression would robustly recover the correlation (Figure 3.16G). Correspondingly if gene expression depends on Smad3 as well as another independent variable, then we cannot recover the correlation (Figure 3.16H).

Search for the hidden cellular variables identifies cell cycle, MEK5 and mTOR
Our analysis suggests that, even though Smad3 is a direct input to Snail, knowing Smad3 dynamics does not fully predict the Snail response. Thus, Snail response is function of Smad3 and at least one other independent variable. Tgfβ has been shown to signal through Smad-independent pathways (reviewed in Zhang, 2017. We tested this and found that including the Tgfβ dose in the analysis does not significantly improve the prediction beyond adding white noise (Figure 3.17). The weak correlation therefore suggests the presence of cell-intrinsic factors that influence how a cell responds to Smad3.

What might be these factors? Stochastic noise due to some biochemical reactions occurring at low-copy numbers or the bursty nature of RNA production has been shown to contribute to the variability of gene output from cell to cell (Suter et al., 2011). However, a recent work also showed that a large degree of transcript variability across single cells is deterministic, and can be predicted with multivariate models of cellular phenotypes (Battich, Stoeger, and Pelkmans, 2015). Motivated by this, we asked if the variability in how clonal cells respond to Smad3 may be due to hidden cellular variables. Returning to our Tgf-β simulation rsults, we confirm the feasibility of applying our method to address this question (Figure 3.19).

Thus we searched for candidate variables that affect how Snail responds to Smad3. We examined multiple classes of factors: (1) Physical features of cells that are readily measurable using microscopy, e.g., nucleus size, cell speed, confluency; (2) Signaling molecules known to regulate Tgf-β pathway, e.g., MEK, ERK, BMP, Wnt; (3) Key cellular processes, including growth, metabolism, epigenetic regulators, and cell cycle. Figure 3.18 provides the complete list of factors tested.

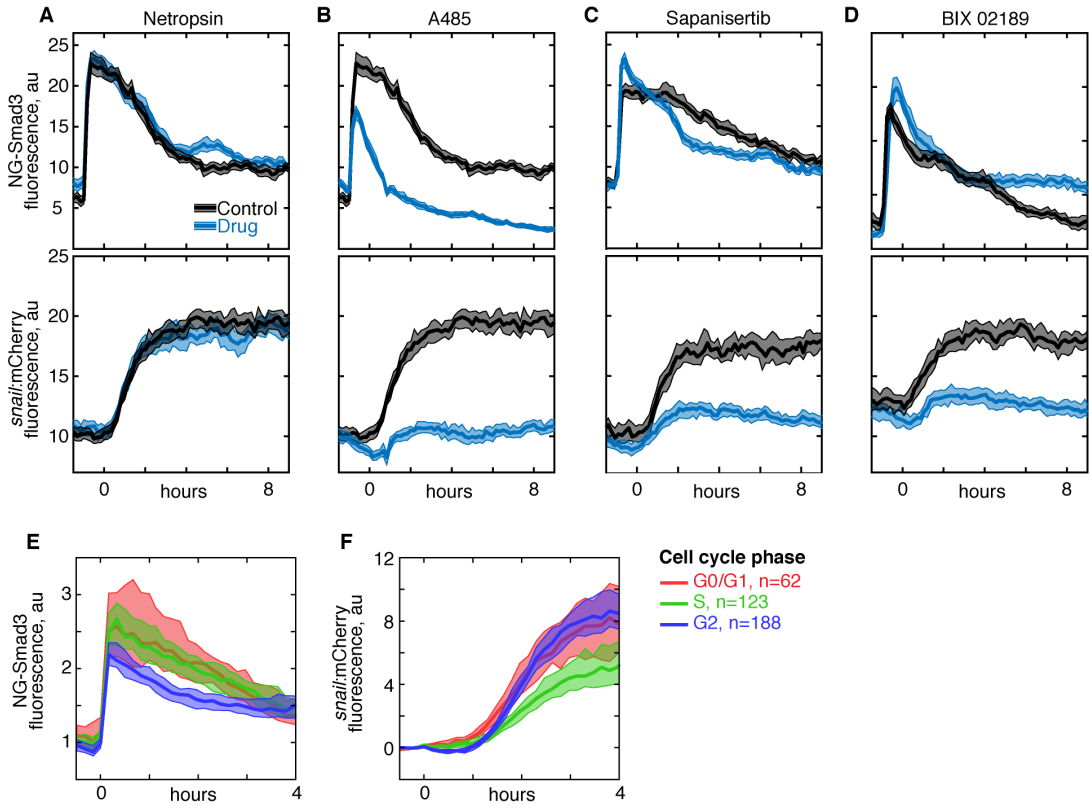

Figure 3.6: Identification of cellular variables that influence how a cell activates Snail in response to Smad3. (A-D) Dynamics of NG-Smad3, *top*, and snail:mCherry, *bottom* in response to Tgf-β1 (2.4 ng/mL, added at t=0) with or without addition of drug. NG-Smad3 fluorescence is median nuclear fluorescence intensity. snail:mCherry fluorescence is total nuclear fluorescence intensity. (A) Treating cells with Netropsin did not affect NG-Smad3 or snai1:mCherry dynamics. (B) Treating cells with A485 altered both NG-Smad3 and snai1:mCherry dynamics. (C-D) Treating cells with Sapanisertib (C) or BIX 02189 (D) altered only snai1:mCherry dynamics. (E-F) Cell cycle stage affects NG-Smad3 and snail:mCherry dynamics discordantly (E) NG-Smad3 dynamics for cells in different phase of cell cycle. (F) snail:mCherry dynamics for cells in different phase of cell cycle. Lines and areas are median and 95% confidence interval of median, respectively.

We found no positive hit with the physical factors examined. For instance, adding different measurements of confluency or migration speed shows no effect beyond adding white noise. Many of the molecular factors examined did not alter NG-Smad3 or snai1:mCherry dynamics (e.g., treating cells with Netropsin in Figure 3.6A). We also found some molecular factors that affected both NG-Smad3 and snai1:mCherry dynamics (e.g., treating cells with A485 in Figure 3.6B). Thus, these factors might simply affect snai1:mCherry through its effects on NG-Smad3.

However, we found 3 factors that affected snail:mCherry dynamics without affecting NG-Smad3 dynamics.

Our first hit is MEK5, a component of the MAPK signaling pathway. Treating cells with BIX02189, a selective inhibitor of MEK5, did not affect NG-Smad3 dynamics, but dramatically inhibited snail:mCherry response (Figure 3.6D). As a control that the effect is specific to MEK5, treatment with U0126 which inhibits MEK1/2, caused an amplified NG-Smad3 response and concomitant amplified snail:mCherry response (Figure 3.20). Our second hit is mTOR, a master regulator of growth. There are two mTOR complexes, mTORC1 and mTORC2. Treatment with sapanisertib, an inhibitor of both mTORC1 and mTORC2, had minimal effect on NG-Smad3 dynamics, but significantly decreased snail:mCherry response (Figure 3.6B). To verify this finding, treatment with rapamycin, which selectively targets mTORC1, showed a similar effect (Figure 3.21). The sufficiency of inhibiting only mTORC1 suggests mTORC2 inhibition is not required. Our third hit is the cell cycle state. Grouping single-cell trajectories by cell cycle, we find that snail:mCherry responds differently to NG-Smad3 depending on the cell-cycle stage. The dynamics of Smad3 were nearly identical across cell cycle stages, but showed the lowest response in G2. In contrast, snail:mCherry response was highest in both G0/G1 and G2, and significantly lower in cells in S phase (Figure 3.6D). The effects of S phase and G2 on Smad3 and Snail dynamics are inverted.

Thus, we found that MEK activity, mTOR activity, and cell cycle state affect snai1:mCherry dynamics independently of NG-Smad3. Does including information on MEK activity, mTOR activity, and cell cycle state along with NG-Smad3 dynamics increase how well we can predict snai1:mCherry? To test this, we coupled live microscopy with sequential antibody staining (Figure 3.7A-B). We used the recently published iterative indirect immunofluorescence imaging protocol by Gut, Herrmann, and Pelkmans (2018). Following live cell imaging of NG-Smad3 dynamics and snai1:mCherry, we measured in the same cells mTORC1/2 targets (p4E-BP1, eIF4E, pS6, and pAKT), and cell cycle state (estimated using multivariate measure of DNA content and nucleus area—see ref. Gut, Tadmor, et al., 2015). We also stained Smad4 and pSmad3. (Due to the lack of high quality antibodies against MEK5 and Erk5, we chose multiple targets of mTORC1/2.) Figure 3.7B shows antibody staining sequentially performed on the same cells. Combining the fluorescence measurements from the antibody staining, Hoecsht stain, and NG-Smad3 manifold, we find statistically significant improvement in the prediction of

snai1:mCherry manifold (Figure 3.7E-F). The modest improvement is impressive in that the measurement fluorescence intensity from antibody staining is not necessarily linear to protein concentration; thus any information about the activity improves prediction.

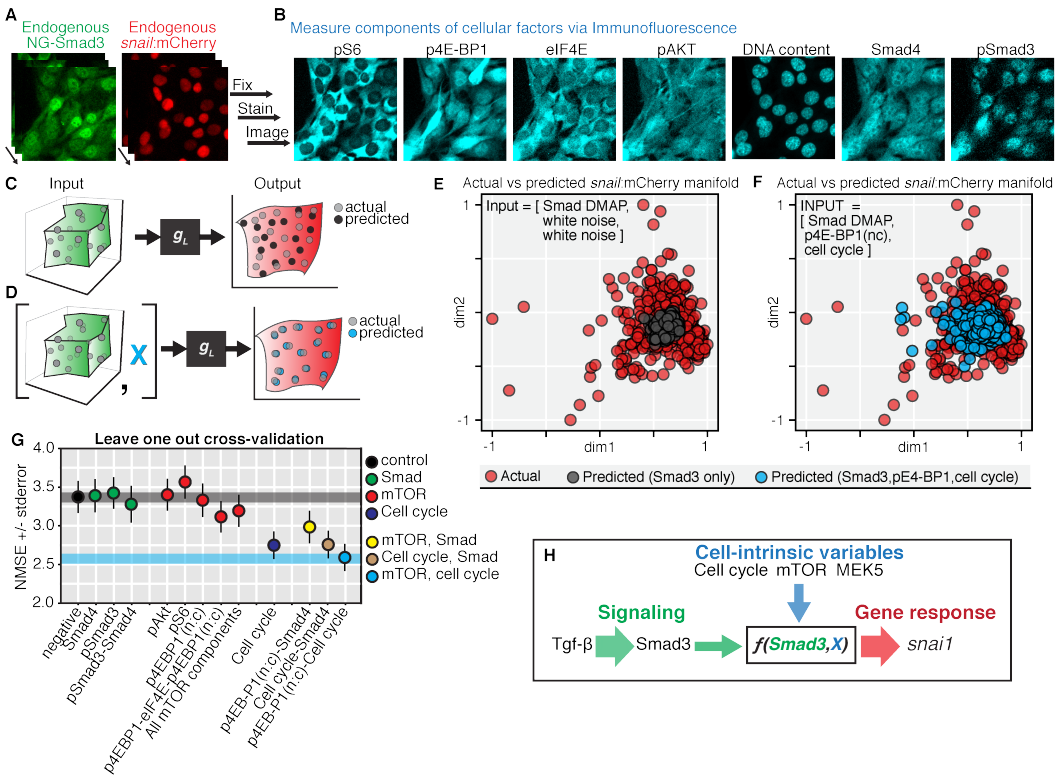

Figure 3.7: Combining intracellular signal and specific intrinsic variables lead to better prediction of gene output. (A-B) Combining live microscopy with sequential antibody staining enables measuring of intracellular signal, gene response, and cell intrinsic variables in the same cells. (A) Cells were stimulated with Tgf-β1 (2.4 ng/mL) for 4 hours and then fixed. (B) Following the protocol developed by Gut, Herrmann, and Pelkmans (2018) cells were iteratively stained and imaged with different antibodies. The different intrinsic variables were mTOR components (pS6, p4E-BP1, eIF4E, pAKT), Cell cycle (DNA content) and other Smad pathway measurements as controls (Smad4 and pSmad3). (C-D) Cartoon showing how if gene response depends on hidden variables, then addition of hidden variables to prediction input will result in improved manifold prediction. (E-F) Testing actual vs predicted snail:mCherry manifolds for different inputs: Smad3 dynamics only, *E*, and Smad3 dynamics and p4E-BP1 and cell cycle, *F*. (G) Quantification of prediction accuracy for different inputs: Smad3 only (black), or Smad3 and other variables—other Smad pathway measures (green), mTOR targets (red), Cell cycle (blue) as well as combinations (yellow, gold, cyan). The most improved prediction occurs using Smad3 dynamics together with p4E-BP1 localization and cell cycle stage. Normalized mean squared error, NMSE, reported is the mean NMSE determined after performing leave one out cross-validation. Error bars are standard error of the mean. (H) The gene response to Smad3 dynamics is modulated by cell intrinsic factors: cell cycle, mTORC1/2, and MEK5.

3.3 Discussion

In this study, we systematically set out to analyze how much does signal dictate gene output in single cells. We found surprisingly that signal dynamics only weakly predicts gene response. Rather, a cell's response to signal is significantly governed by its own internal state (Figure 3.7H).

The existence of factors that modulate gene response to a given signal is well known across tissues and cell-types (ref David and Massagué, 2018). The different expression levels of cell-type specific factors can change 1) which genes are activated in response to signal (Mullen et al., 2011) and 2) how strongly those genes are activated (Lamouille et al., 2012). These factors help explain the variation in outcomes across tissues and cell types. Our study finds that variability in gene expression response in single cells also depends cell-intrinsic factors that vary across a clonal population. The effect is significant to the point that even knowing the entire signal dynamics gives us poor prediction of the gene output.

We found here 3 cellular variables that dictate how cells interpret Smad3 signal: the activity of another signaling pathway (MEK5), growth state of the cells (mTOR), and cell cycle state. That these factors interface with Tgf-β/Smad signaling is not new. Previous studies have identified that MEK5 activity and mTOR activity are required for Tgf-β/Smad induced Snail expression (Marchetti et al., 2008; Lamouille et al., 2012; Pavan et al., 2018; S. Kim, Lim, and Woo, 2013), and it was shown that a cell either undergoes epithelial-to-mesenchymal transition (EMT) or apoptosis depending on cell cycle in AML-12 murine hepatocytes (Yang et al., 2006). However, the extent to which these effects were mediated by altering Smad dynamics was not fully known. Our work directly shows that these factors act not by modulating Tgf-β/Smad signaling dynamics, but rather by modulating how Smad3 dynamics are transduced into gene response. Further, our work extends the effect of these factors to variability across a clonal population of cells.

It is critical to note that we are not trying to assert we have discovered all of the hidden variables that explain Smad/Snail input/output relationship. Rather here our goal was to identify and confirm that there exist factors whose variability underlies the variability in snail response to Smad3. It will therefore be important to examine the effect of other factors. A primary candidate to test are master transcription factors, which affect where Smad3 binds in the genome (Mullen et al., 2011). These master transcription factors are known to vary across cell-types but the effect of their variation within cells is not known. A major challenge in addressing this is

determining which master transcription factor is dominant in a given cell type, as the identity of the master transcription factors affecting Smad3 binding has only been characterized for 4 cell types (Mullen et al., 2011).

Importantly, this work brings new insights to assessing clinical treatments for Tgf-β signaling in disease. Tgf-β signaling is known to induce epithelial to mesenchymal transition (EMT), a cell fate decision associated with metastasis in cancer, by upregulating the expression of Snail, (J. Xu, Lamouille, and Derynck, 2009). Because the functioning of each of mTOR, MEK5 and cell cycle has been determined to affect Tgf-β induced EMT (Lamouille et al., 2012; Pavan et al., 2018; Yang et al., 2006), it is postulated that cancer treatments could be better targeted by assessing the level of activity of Tgf-β and these other factors in a tissue. Our work suggests the need to not just assess the level of activity of MEK5 and mTOR activity in a tissue, but also heterogeneity of these factors as well. Indeed, heterogeneity in Tgf-β-induced EMT is frequently observed in patients, where cells adopt a hybrid epithelial mesynchymal state, which is associated with greater metastatic and tumorigenic capacity and poorer patient prognosis (Sulaiman, Z. Yao, and L. Wang, 2018). Thus in assessing targeting for therapies it will be essential to integrate information from the heterogeneity of Smad dynamics as well as the heterogeneity of factors associated with modulating the response to Smad dynamics.

Finally, this work opens a new door for analysis of signaling dynamics. The approach presented here, applying manifold learning to endogenous signaling dynamics, has uncovered a significant role for cell intrinsic factors in determining how cells respond to Smad dynamics. The Smad pathway is set apart from other pathways by how many different cellular functions it regulates (David and Massagué, 2018). It will be interesting to see if other pathways exhibit a similar dependency on cell intrinsic factors for how signaling dynamics are interpreted. In the Tgf-β/Smad pathway, to determine how a cell decodes the instruction encoded by signaling dynamics, one must know the cell itself.

References

Balleza, Enrique, J Mark Kim, and Philippe Cluzel (2018). "Systematic characterization of maturation time of fluorescent proteins in living cells". In: *Nature methods* 15.1, p. 47.

Battich, Nico, Thomas Stoeger, and Lucas Pelkmans (2015). "Control of transcript variability in single mammalian cells". In: *Cell* 163.7, pp. 1596–1610.

Brandl, Martina et al. (2010). "IKKα controls canonical TGFβ–SMAD signaling to regulate genes expressing SNAIL and SLUG during EMT in Panc1 cells". In: *J Cell Sci* 123.24, pp. 4231–4239.

Chen, Hong Bing et al. (2005). "Nuclear targeting of transforming growth factor-β-activated Smad complexes". In: *Journal of Biological Chemistry* 280.22, pp. 21329–21336.

Coifman, Ronald R and Stéphane Lafon (2006). "Diffusion maps". In: *Applied and computational harmonic analysis* 21.1, pp. 5–30.

David, Charles J, Yun-Han Huang, et al. (2016). "TGF-β tumor suppression through a lethal EMT". In: *Cell* 164.5, pp. 1015–1030.

David, Charles J and Joan Massagué (2018). "Contextual determinants of TGFβ action in development, immunity and cancer". In: *Nature Reviews Molecular Cell Biology*, p. 1.

Derynck, Rik and Erine H Budi (2019). "Specificity, versatility, and control of TGF-β family signaling". In: *Sci. Signal.* 12.570, eaav5183.

Frick, Christopher L, Clare Yarka, Harry Nunns, and Lea Goentoro (2017). "Sensing relative signal in the Tgf-β/Smad pathway". In: *Proceedings of the National Academy of Sciences*, p. 201611428. DOI: 10.1073/pnas.1611428114.

Gillies, Taryn E et al. (2017). "Linear integration of ERK activity predominates over persistence detection in Fra-1 regulation". In: *Cell systems* 5.6, pp. 549–563.

Gut, Gabriele, Markus D Herrmann, and Lucas Pelkmans (2018). "Multiplexed protein maps link subcellular organization to cellular states". In: *Science* 361.6401, eaar7042.

Gut, Gabriele, Michelle D Tadmor, et al. (2015). "Trajectories of cell-cycle progression from fixed cell populations". In: *Nature methods* 12.10, p. 951.

Hill, Caroline S (2009). "Nucleocytoplasmic shuttling of Smad proteins". In: *Cell research* 19.1, p. 36.

Hoot, Kristina E et al. (2008). "Keratinocyte-specific Smad2 ablation results in increased epithelial-mesenchymal transition during skin cancer formation and progression". In: *The Journal of clinical investigation* 118.8, pp. 2722–2732.

Inman, Gareth J et al. (2002). "SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7". In: *Molecular pharmacology* 62.1, pp. 65–74.

Ju, Wenjun et al. (2006). "Deletion of Smad2 in mouse liver reveals novel functions in hepatocyte growth and differentiation". In: *Molecular and cellular biology* 26.2, pp. 654–667.

Kim, Suji, Jae Hyang Lim, and Chang-Hoon Woo (2013). "ERK5 inhibition ameliorates pulmonary fibrosis via regulating Smad3 acetylation". In: *The American journal of pathology* 183.6, pp. 1758–1768.

Lamouille, Samy et al. (2012). "TGF-β-induced activation of mTOR complex 2 drives epithelial–mesenchymal transition and cell invasion". In: *J Cell Sci* 125.5, pp. 1259–1273.

Lee, John A and Michel Verleysen (2007). *Nonlinear dimensionality reduction*. Springer Science & Business Media.

Lee, Robin EC et al. (2014). "Fold change of nuclear NF-κB determines TNF-induced transcription in single cells". In: *Molecular cell* 53.6, pp. 867–879.

Li, Xianqiang et al. (1998). "Generation of destabilized green fluorescent protein as a transcription reporter". In: *Journal of Biological Chemistry* 273.52, pp. 34970–34975.

Li, Yuchao et al. (2018). "Spatiotemporal control of TGF-β signaling with light". In: *ACS synthetic biology* 7.2, pp. 443–451.

Lin, Huei-Min et al. (2009). "Transforming growth factor-β/Smad3 signaling regulates insulin gene transcription and pancreatic islet β-cell function". In: *Journal of Biological Chemistry* 284.18, pp. 12246–12257.

Lin, Xia et al. (2006). "PPM1A functions as a Smad phosphatase to terminate TGFβ signaling". In: *Cell* 125.5, pp. 915–928.

Liu, Dong, Brian L Black, and Rik Derynck (2001). "TGF-β inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3". In: *Genes & development* 15.22, pp. 2950–2966.

Liu, Xuedong et al. (1997). "Transforming growth factor β-induced phosphorylation of Smad3 is required for growth inhibition and transcriptional induction in epithelial cells". In: *Proceedings of the National Academy of Sciences* 94.20, pp. 10669–10674.

Lo, Roger S and Joan Massagué (1999). "Ubiquitin-dependent degradation of TGF-β-activated Smad2". In: *Nature cell biology* 1.8, p. 472.

Lönn, Peter et al. (2010). "PARP-1 attenuates Smad-mediated transcription". In: *Molecular cell* 40.4, pp. 521–532.

Malhotra, Nidhi and Joonsoo Kang (2013). "SMAD regulatory networks construct a balanced immune system". In: *Immunology* 139.1, pp. 1–10.

Marchetti, Alessandra et al. (2008). "ERK5/MAPK is activated by TGFβ in hepatocytes and required for the GSK-3β-mediated Snail protein stabilization". In: *Cellular signalling* 20.11, pp. 2113–2118.

Mullen, Alan C et al. (2011). "Master transcription factors determine cell-type-specific responses to TGF-β signaling". In: *Cell* 147.3, pp. 565–576.

Nicolás, Francisco J et al. (2004). "Analysis of Smad nucleocytoplasmic shuttling in living cells". In: *Journal of cell science* 117.18, pp. 4113–4125.

Osborn, Mark J et al. (2005). "A picornaviral 2A-like sequence-based tricistronic vector allowing for high-level therapeutic gene expression coupled to a dual-reporter system". In: *Molecular Therapy* 12.3, pp. 569–574.

Paek, Andrew L et al. (2016). "Cell-to-cell variation in p53 dynamics leads to fractional killing". In: *Cell* 165.3, pp. 631–642.

Pavan, Simona et al. (2018). "A kinome-wide high-content siRNA screen identifies MEK5-ERK5 signaling as critical for breast cancer cell EMT and metastasis." In: *Oncogene* 37.31, pp. 4197–4213.

Purvis, Jeremy E et al. (2012). "p53 dynamics control cell fate". In: *Science* 336.6087, pp. 1440–1444.

Rand, Ulfert et al. (2012). "Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response". In: *Molecular systems biology* 8.1, p. 584.

Rasmussen, Carl Edward (1999). *Evaluation of Gaussian processes and other methods for non-linear regression.* University of Toronto.

Rechsteiner, Martin and Scott W Rogers (1996). "PEST sequences and regulation by proteolysis". In: *Trends in biochemical sciences* 21.7, pp. 267–271.

Reyes, José et al. (2018). "Fluctuations in p53 signaling allow escape from cell-cycle arrest". In: *Molecular cell* 71.4, pp. 581–591.

Sato, Misako et al. (2003). "Targeted disruption of TGF-β1/Smad3 signaling protects against renal tubulointerstitial fibrosis induced by unilateral ureteral obstruction". In: *The Journal of clinical investigation* 112.10, pp. 1486–1494.

Selimkhanov, Jangir et al. (2014). "Accurate information transmission through dynamic biochemical signaling networks". In: *Science* 346.6215, pp. 1370–1373.

Sero, Julia E et al. (2015). "Cell shape and the microenvironment regulate nuclear translocation of NF-κB in breast epithelial and tumor cells". In: *Molecular systems biology* 11.3, p. 790.

Shaner, Nathan C et al. (2013). "A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum". In: *Nature methods* 10.5, p. 407.

Strasen, Jette et al. (2018). "Cell-specific responses to the cytokine TGFβ are determined by variability in protein levels". In: *Molecular systems biology* 14.1, e7733.

Sulaiman, Andrew, Zemin Yao, and Lisheng Wang (2018). "Re-evaluating the role of epithelial-mesenchymal-transition in cancer progression". In: *Journal of biomedical research* 32.2, p. 81.

Sung, Myong-Hee et al. (2014). "Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF-κB signaling". In: *Sci. Signal.* 7.308, ra6–ra6.

Suter, David M et al. (2011). "Mammalian genes are transcribed with widely different bursting kinetics". In: *Science* 332.6028, pp. 472–474.

Tenenbaum, Joshua B, Vin De Silva, and John C Langford (2000). "A global geometric framework for nonlinear dimensionality reduction". In: *science* 290.5500, pp. 2319–2323.

Thuault, Sylvie et al. (2008). "HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition". In: *Journal of Biological Chemistry* 283.48, pp. 33437–33446.

Tidin, Onur et al. (2019). "Quantitative relationships between SMAD dynamics and target gene activation kinetics in single live cells". In: *Scientific Reports* 9.1, p. 5372.

Uttamsingh, S et al. (2008). "Synergistic effect between EGF and TGF-β1 in inducing oncogenic properties of intestinal epithelial cells". In: *Oncogene* 27.18, p. 2626.

Warmflash, Aryeh et al. (2012). "Dynamics of TGF-β signaling reveal adaptive and pulsatile behaviors reflected in the nuclear localization of transcription factor Smad4". In: *Proceedings of the National Academy of Sciences* 109.28, E1947–E1956.

Wong, Victor C et al. (2018). "NF-κB-chromatin interactions drive diverse phenotypes by modulating transcriptional noise". In: *Cell reports* 22.3, pp. 585–599.

Xu, Fengyun et al. (2016). "TGF-β/SMAD pathway and its regulation in hepatic fibrosis". In: *Journal of Histochemistry & Cytochemistry* 64.3, pp. 157–167.

Xu, Jian, Samy Lamouille, and Rik Derynck (2009). "TGF-β-induced epithelial to mesenchymal transition". In: *Cell research* 19.2, p. 156.

Yagi, Ken et al. (1999). "Alternatively spliced variant of Smad2 lacking exon 3 Comparison with wild-type Smad2 and Smad3". In: *Journal of Biological Chemistry* 274.2, pp. 703–709.

Yang, Y et al. (2006). "Transforming growth factor-β1 induces epithelial-to-mesenchymal transition and apoptosis via a cell cycle-dependent mechanism". In: *Oncogene* 25.55, p. 7235.

Yao, Jason, Anna Pilko, and Roy Wollman (2016). "Distinct cellular states determine calcium signaling response". In: *Molecular systems biology* 12.12, p. 894.

Yingling, Jonathan M et al. (1996). "Mammalian dwarfins are phosphorylated in response to transforming growth factor beta and are implicated in control of cell growth". In: *Proceedings of the National Academy of Sciences* 93.17, pp. 8940–8944.

Zavadil, Jiri et al. (2004). “Integration of TGF-β/Smad and Jagged1/Notch signalling in epithelial-to-mesenchymal transition”. In: *The EMBO journal* 23.5, pp. 1155–1165.

Zhang, Ying E (2017). “Non-Smad signaling pathways of the TGF-β family”. In: *Cold Spring Harbor perspectives in biology* 9.2, a022129.

3.4 Supplementary Figures

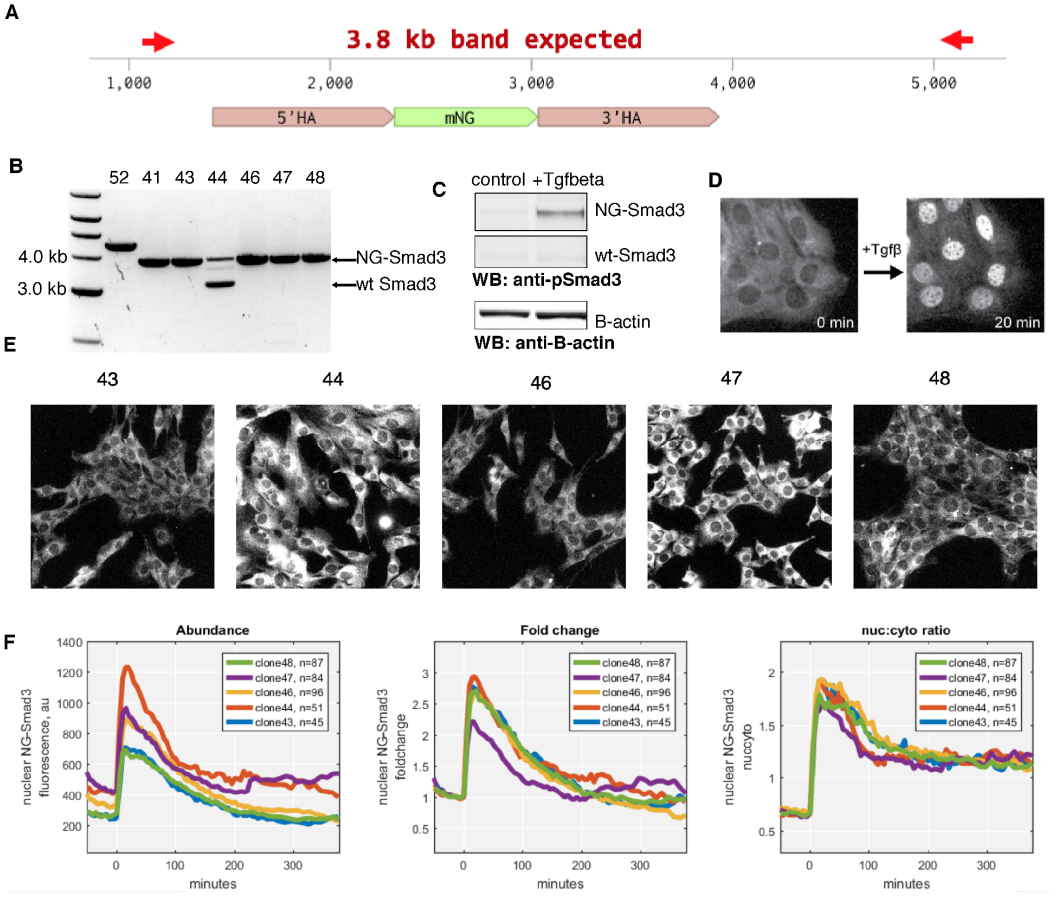

Figure 3.8: Characterization of NG-Smad3 clones. (A). Each clone is analyzed by genomic PCR to assess whether both alleles are edited (homozygous) or only one is edited (heterozygous). The primers are positioned outside of the homology arms to ensure amplification is of target region of genome and not residual HDR template plasmid. (B) Multiple clones showed homozygous incorporation of mNeonGreen into the *smad3* locus. Clone 44 is heterozygous, given the presence of a wild-type smad3 band. Clone 52 was rejected because the PCR band size was incorrect. (C) Western blotting against anti-phospho-Smad3 shows NG-Smad3 is phosphorylated in response to Tgf-β additon. Antibody is 1:500 dilution of Phospho-SMAD3 (Ser423, Ser425) Antibody (16H5L12), ABfinity Rabbit Monoclonal (Invitrogen). β-Actin was stained as a loading control. Antibody is 1:1000 β-Actin (8H10D10) Mouse mAb (#3700 Cell Signaling Technology). Secondary antibodies are IRDye 680LT Goat (polyclonal) anti-Rabbit IgG (H+L) Highly Cross-Adsorbed (925-68021 Licor), 1:5,000 dilution and IRDye 800CW Goat (polyclonal) anti-Mouse IgG (H+L) Highly Cross-Adsorbed (926-32210 Licor), 1:10,000 dilution. (E) Representative fluorescence images of clones 43, 44, 46, 47 and 48. The clones show similar morphology and localization. (F). The dynamics of each clone in response to Tgf-β1 addition (2.4 ng/mL, added immediately prior to t=0). The population median is plotted.

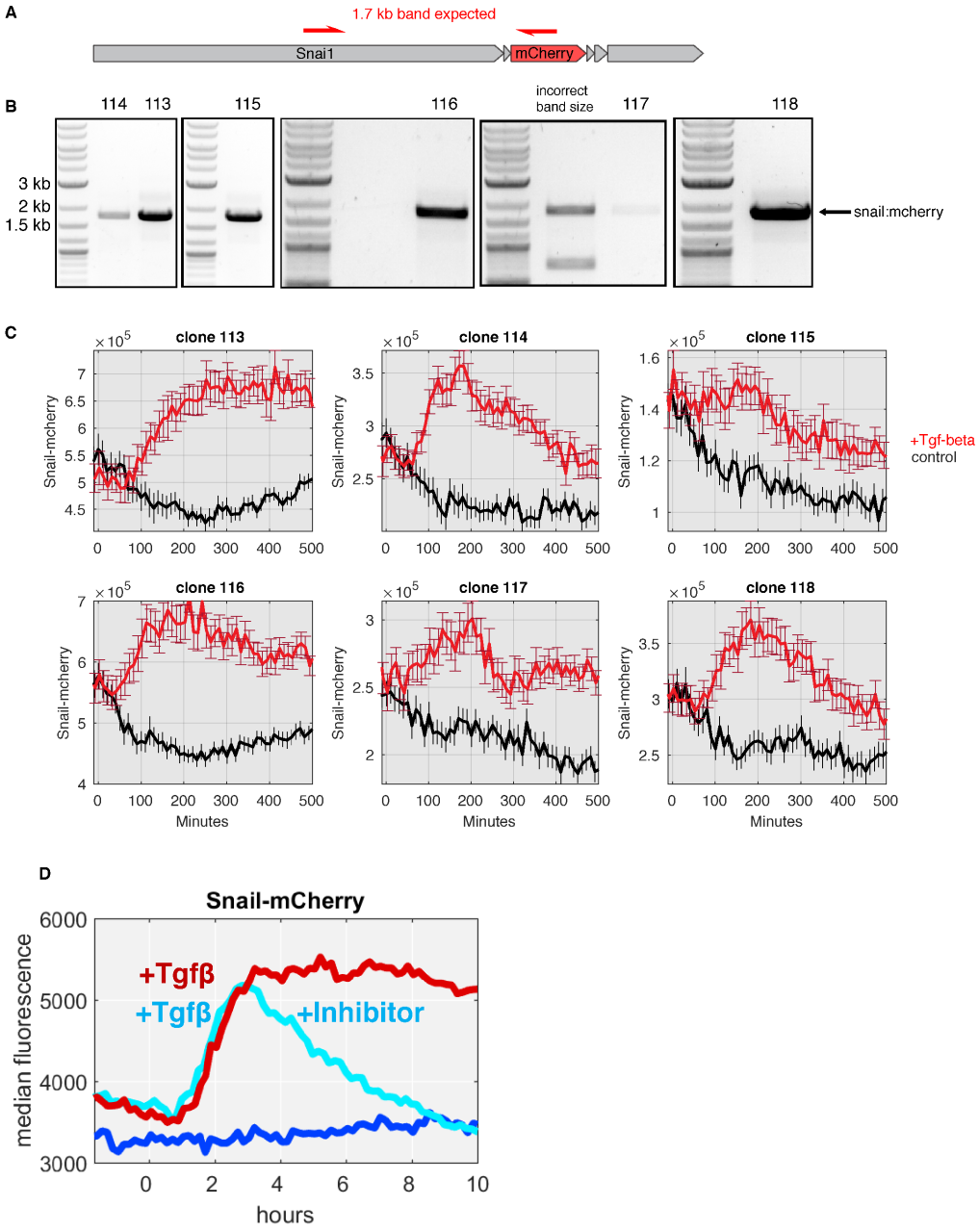

Figure 3.9: Characterization of snail:mCherry in candidate double reporter clones. (A). Each clone is analyzed by genomic PCR to assess whether both alleles to ensure the correct genomic edit has occured for at least one allele (heterozygous). (B) Multiple clones showed correct incorporation of mCherry into the *snail* locus. Clones with no band or incorrect band sizes were rejected.(C) The dynamics of each clone in response to Tgf-β1 addition (2.4 ng/mL, added immediately prior to t=0).(D) Addition of SB 431-542 ("inhibitor"), which inhibits phosphorylation of Smad2/3 terminates the induction of *snail*:mCherry.

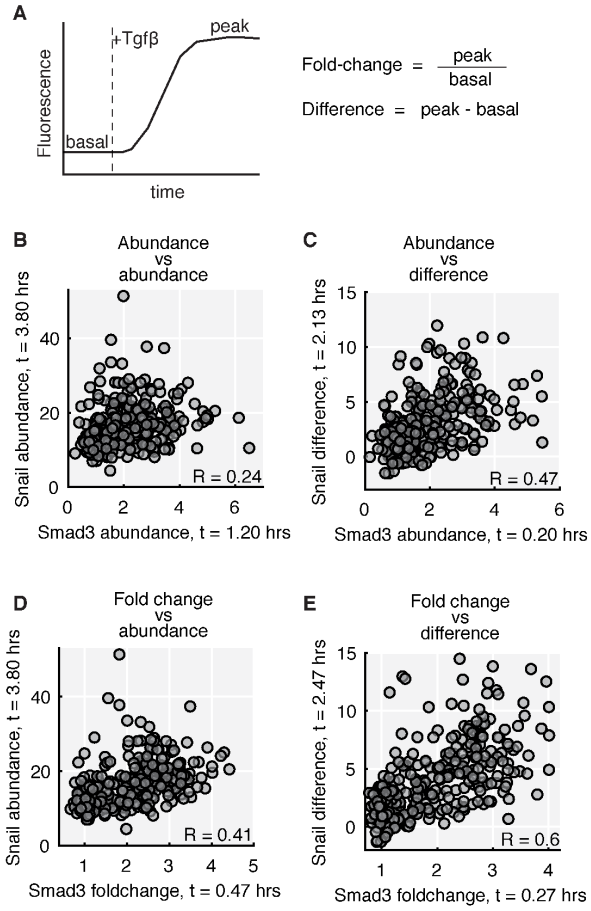

Figure 3.10: Testing for correlation between dynamics of NG-Smad3 and *snail*:mCherry. (A) The different dynamic features tested illustrated. (B-E) The different correlations tested. (B) Correlation between NG-Smad3 abudnance and *snail*:mCherry abundance. (C) Correlation between NG-Smad3 abudnance and *snail*:mCherry difference. (D) Correlation between NG-Smad3 fold-change and *snail*:mCherry abundance. (E) Correlation between NG-Smad3 fold-change and *snail*:mCherry difference.

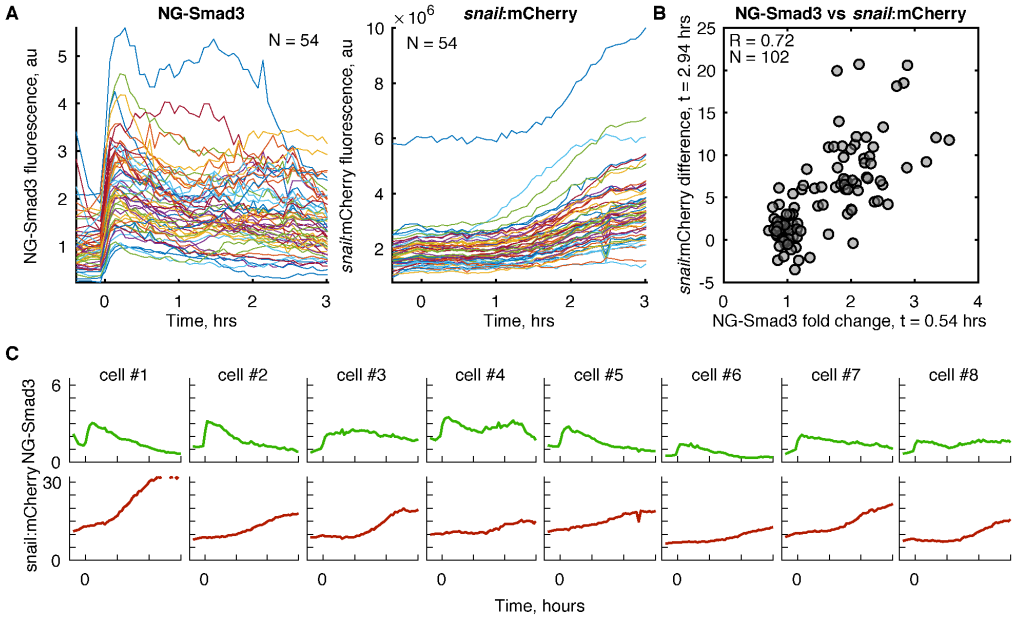

Figure 3.11: A second clone also shows weak correlation between dynamics of NG-Smad3 and *snail*:mCherry. (A) Single cell trajectories of endogenous NG-Smad3 fluorescence in response to Tgf-β1, added at t=0. Fluorescence is quantified as the median nuclear fluorescence, *left*. Single cell trajectories of endogenous snail:mCherry fluorescence in response to Tgf-β1, added at t=0. Fluorescence is quantified as the total nuclear fluorescence to capture total production of snail, *right*. (B) snail:mCherry expression correlates most strongly with fold change in NG-Smad3. (C) Quantitation of endogenous NG-Smad3 and snail:mCherry dynamics within single cells.

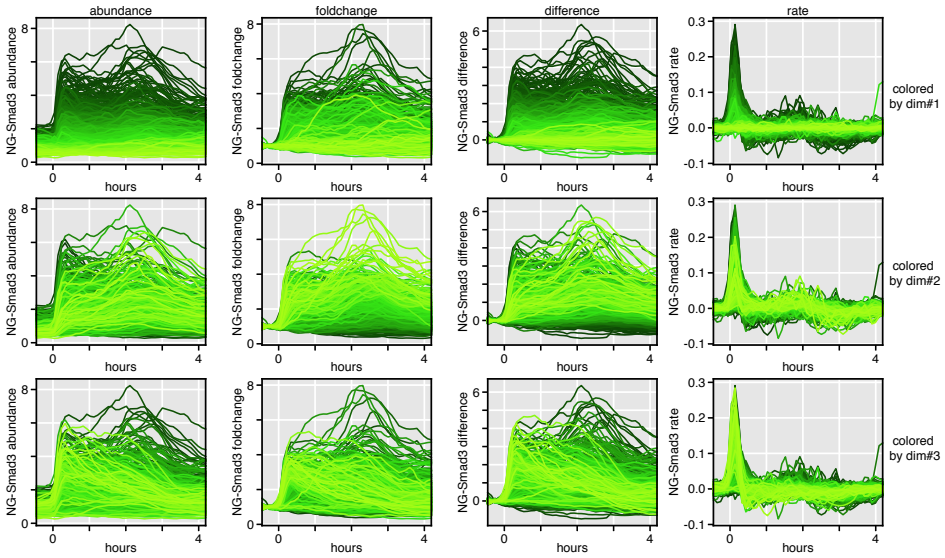

Figure 3.12: The trajectories of NG-Smad3 colored by their manifold coordinate values. Top panels: the single cell trajectories colored according to their location in dimension 1. Middle panels: The trajectories colored according to their location in dimension 2. Bottom panels: The trajectories colored according to their location in dimension 3. Moving from left to right, the panels plot different functions of snail:mCherry. Raw fluorescence values ("abundance"), *left*; Fold change in fluorescence, *left-middle*; Increase in fold-change fluorescence ("difference"), *right-middle*; instantaneous rate of change ("rate"), *right*.

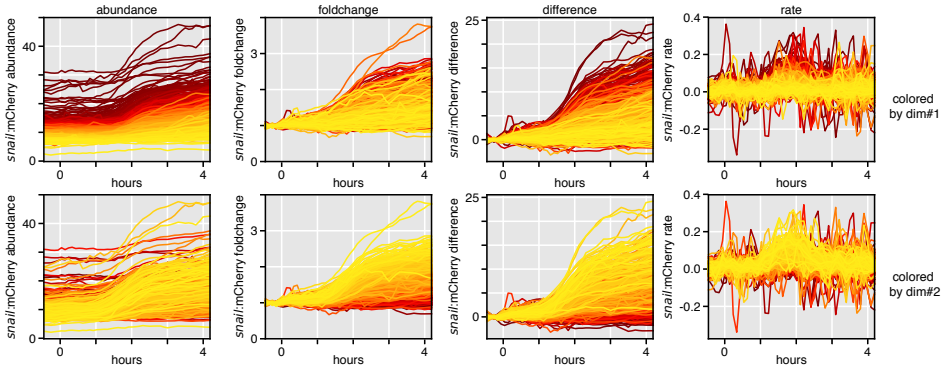

Figure 3.13: The trajectories of snail:mCherry colored by their manifold coordinate values. Top panels: the single cell trajectories colored according to their location in dimension 1. Bottom panels: The trajectories colored according to their location in dimension 2. Moving from left to right, the panels plot different functions of snail:mCherry. Raw fluorescence values ("abundance"), *left*; Fold change in fluorescence, *left-middle*; Increase in fold-change fluorescence ("difference"), *right-middle*; instantaneous rate of change ("rate"), *right*.

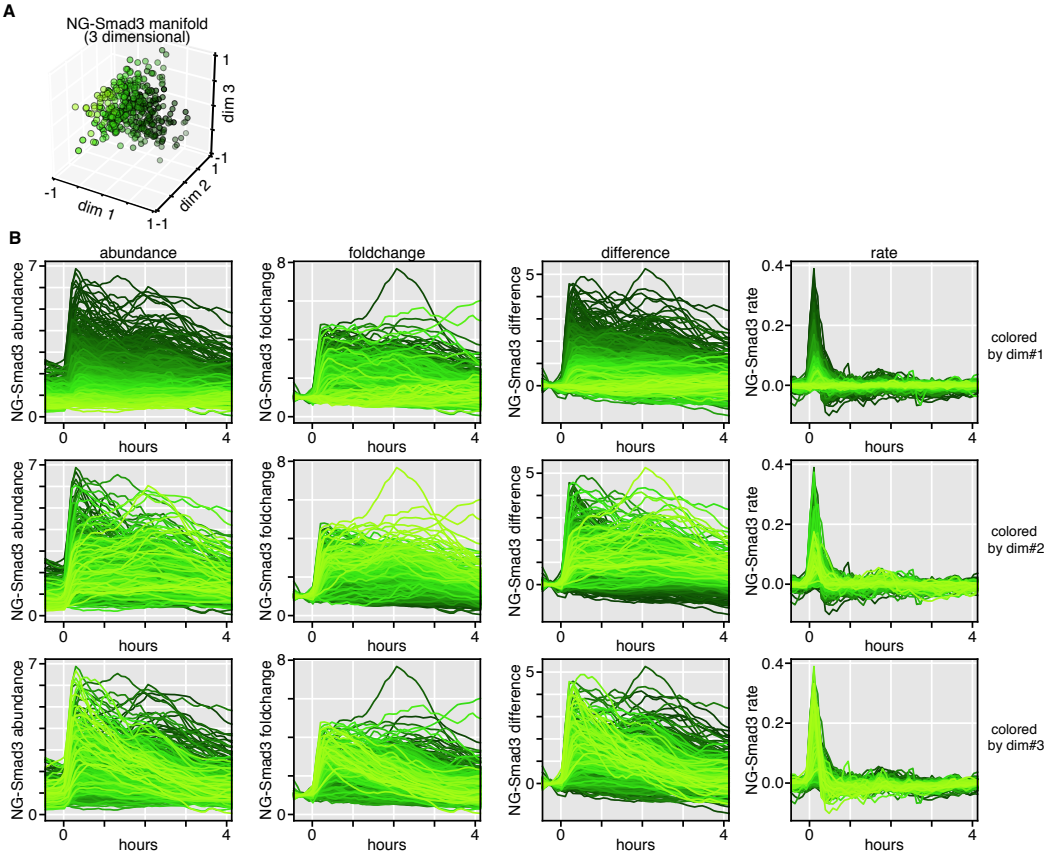

Figure 3.14: Diffusion Maps analysis obtains a good manifold for NG-Smad3 in second experiment. (A) The Smad manifold obtained for the second independent experiment is also 3 dimensional. (N=375 cells) (B) The trajectories of NG-Smad3 colored by their manifold coordinate values. Top panels: the single cell trajectories colored according to their location in dimension 1. Middle panels: the trajectories colored according to their location in dimension 2. Bottom panels: The trajectories colored according to their location in dimension 3. Moving from left to right, the panels plot different functions of NG-Smad3. Raw fluorescence values ("abundance"), *left*; Fold change in fluorescence, *left-middle*; Increase in fold-change fluorescence ("difference"), *right-middle*; instantaneous rate of change ("rate"), *right*.

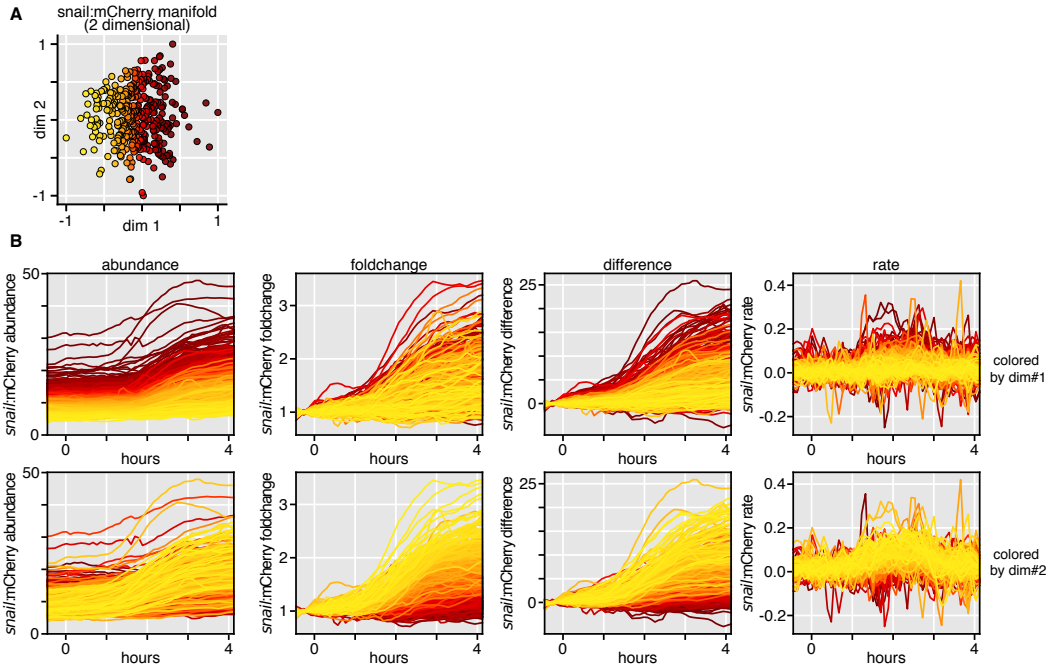

Figure 3.15: Diffusion Maps analysis obtains a good manifold for *snail*:mCherry in second experiment. (A) The *snail*:mCherry manifold obtained for the second independent experiment is also 2 dimensional. (N=375 cells) The trajectories of *snail*:mCherry colored by their manifold coordinate values. Top panels: the single cell trajectories colored according to their location in dimension 1. Bottom panels: the trajectories colored according to their location in dimension 2. Moving from left to right, the panels plot different functions of *snail*:mCherry. Raw fluorescence values ("abundance"), *left*; Fold change in fluorescence, *left-middle*; Increase in fold-change fluorescence ("difference"), *right-middle*; instantaneous rate of change ("rate"), *right*.

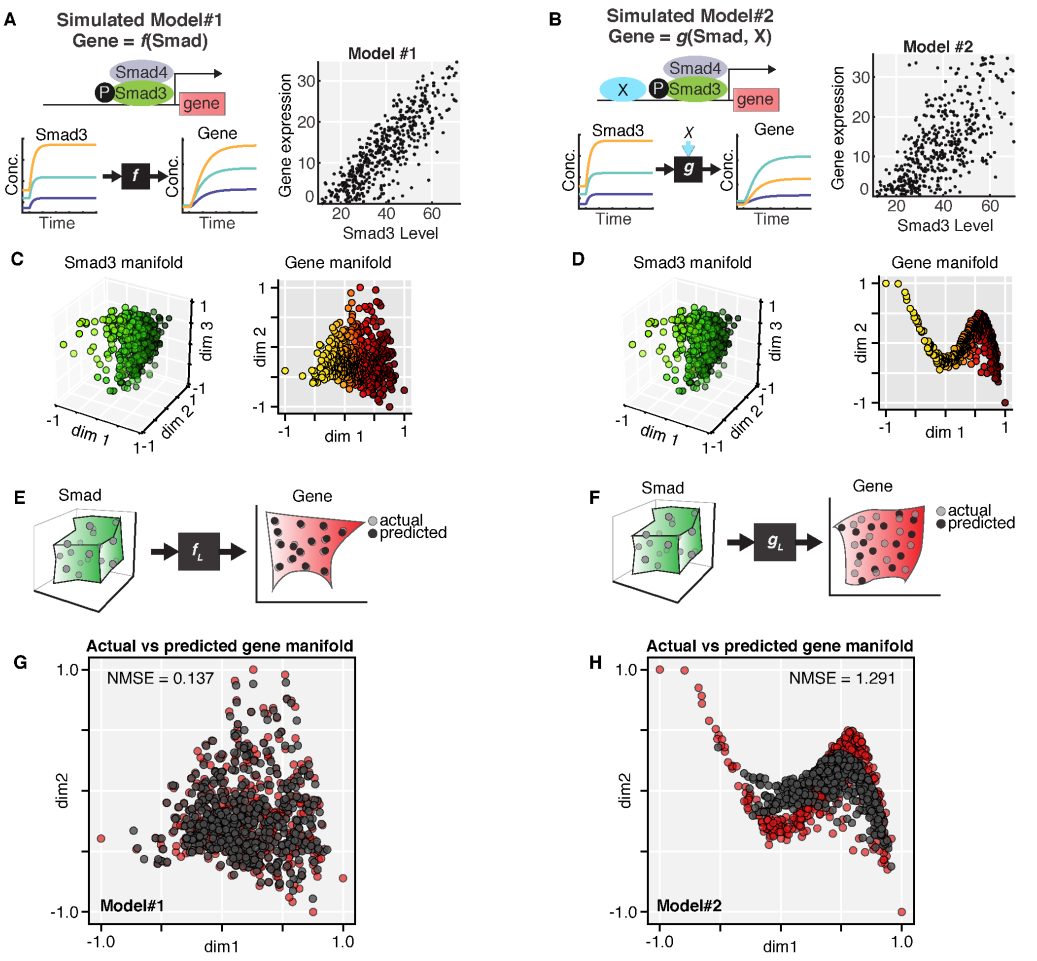

Figure 3.16: Simulations of Tgf-β/Smad model to validate our manifold learning and gaussian process approach. (A-B) Illustrations of the two models being simulated, including representative traces of Smad3 and Gene expression from the simulations, and a plot showing that the best correlation between any time point measures of Smad3 and Gene expression is weak for both models. (A) In model#1, gene expression depends only on Smad3:Smad4 complex. (B) IN model#2, gene expression depends on Smad3:Smad4 complex *and* an unknown cellular factor, X. (C-D) The Smad3 and Gene manifolds obtained from the simulations (N=500 simulations for both models). (E-F) Illustration showing the expected outcome of applying Gaussian Process Regression to the two models. (E) The Gene manifold should be completely predictable by the Smad3 manifold for model#1. (F). The Gene manifold should not be able to be recovered from Smad manifold alone in the case of model#2. (G-H) Actual vs predicted gene manifolds after applying Gaussian Process Regession to the simulaiton manifolds. (G) In model#1, it is possible to fully predict the Gene manifold. (H) In model#2, the prediciton of Gene manifold is inaccurate. NMSE is normalized mean squared error. The mathematical model used for all simulations is from (Schmierer et al., 2008). Simulations were carried out with random parameter variation, to mimic natural cell-cell variatbility, as in Frick et al., 2017.

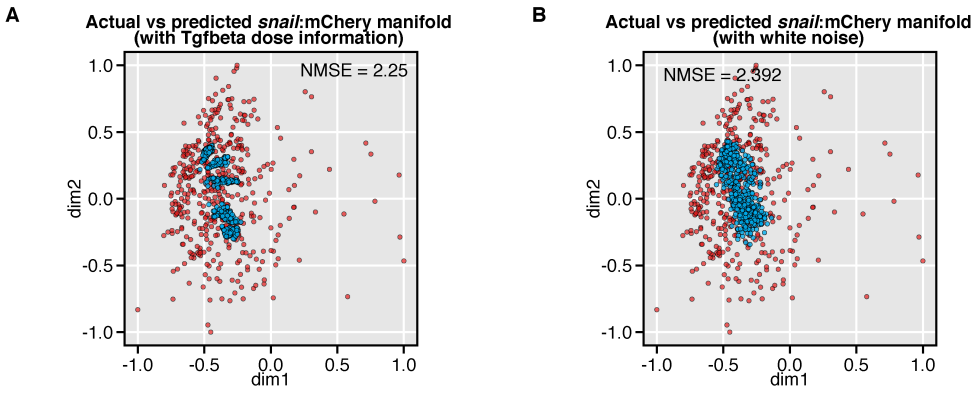

Figure 3.17: Including Tgf-β dose information in GP regression does not improve prediciton of Snail (A) Actual vs predicted plot for snail:mCherry manifold when prediction includes information about Tgf-β dose. (B) Actual vs predicted plot for prediction using white noise instead of Tgf-β dose. NMSE, normalized mean squared error.

Name	Target	Dose	Effect on Smad3	Effect on Snail
3BP	glycolysis	100 uM	cannot analyze, significant morphology change	cannot analyze, significant morphology change
A-485	p300/CBP	3 uM	Strong inhibition of Smad3	Strong inhibition of Smad3
Antimycin	oxidative phosphorylation	10 uM	negligible effect on Smad3	negligible effect on Snail
BIX 02189	MEK5 inhibitior	10 uM	Negligible effect on Smad3	strong inhibition of Snail
BMP-2	BMP pathway	100, 10, or 1 ng/mL	transient, adaptive response	strong sustained response
CHIR99021	Gsk3-beta inhibitor Wnt pathway	3 uM	Negligible effect on Smad3	negligible effect on Snail
EGF	Mapk pathway	100 ng/mL	imperceptible	imperceptible
GANT61	Gli1/2 [hedgehog pathway]	20 uM	cannot analyze, significant cytotoxic	cannot analyze, significant cytotoxic
Insulin	Insulin/mTor pathway	5 ug/mL	no effect	perhaps slight increase
LDN193189	BMPRI (alk2, alk3)	1 uM	slightly decreased basal smad	decreased basal expression
Netropsin	HMGA2 (and prolings G cycle)	1 uM	negligible effect on Smad3	negligible effect on Snail
PD0325901	MEK 1/MEK2	1 uM (10, 1, 0.1 0.01)	slight increase in smad abundance stronger increase in activated level	increase in total snail (basal) weakens snail activation
Rapamycin	mTor	1 nM to 1000 nM	Negligible effect on Smad3	Significant inhibition of Snail
Romidepsin	HDAC1, HDAC2	100 nM	strong increase in Smad3 activation	increase in snail expression in absence of Tgf-β
Sapanisertib	mTorc1/2 kinase domain	25 nM	negligible effect on Smad3 at moderate doses	strong inhibition of Snail
SB431542	TGFBRI (alk5)	10 uM	lower basal and stimulated level lower nuclear localization	decreased basal expression slightly weakened ligand response
U0126	MEK1/MEK2	10 uM, (10, 1, 0.1, 0.01)	slight increase in smad abundance slight increase in activated level	slightly stronger stimulated activation noticeable basal increase
Wortmannin	covalent inhibitor phosphoinositide 3-kinases (PI3Ks)	1 uM	Negligible effect on Smad3	negligible effect on Snail

Figure 3.18: Complete list of drugs and agonists tested.

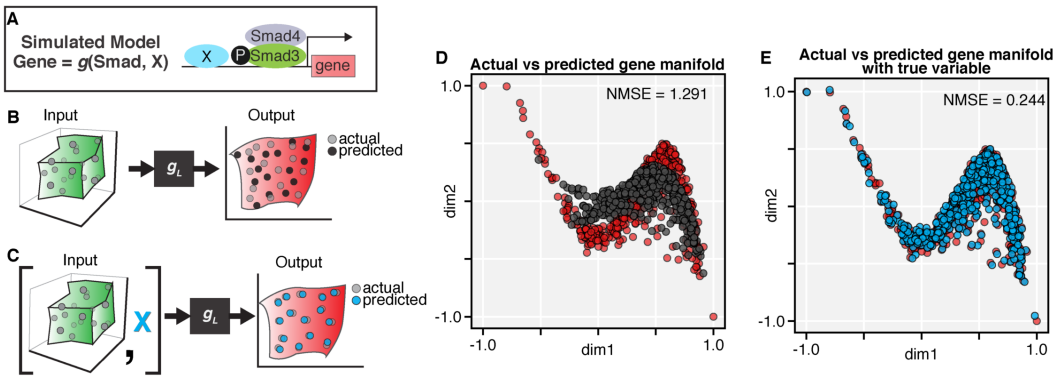

Figure 3.19: Correct identification of hidden variable restores manifold prediction accuracy. (A) The scenario being analyzed is model#2. (B-C) The expected outcome of trying to predict the Gene manifold while only using Smad3 manifold as input *(B)* or when predicting Gene manifold using a matrix of both the Smad3 manifold and measurements X *(C)*. (D-E) Actual vs predicted gene manifolds after applying Gaussian Process Regession to the simulation manifolds. (D) The Gene manifold cannot be predicted with only the Smad manifold as an input. (E) Including measurements of X in the Gaussian Process input enables a recovers an accurate prediction of the Gene manifold. NMSE is normalized mean squared error.

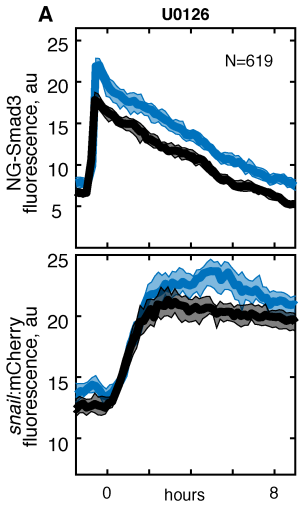

Figure 3.20: (A) Dynamics of NG-Smad3, *top*, and snail:mCherry, *bottom* in response to Tgf-β1 (2.4 ng/mL, added at t=0) with or without addition of drug. NG-Smad3 fluorescence is median nuclear fluorescence intensity. snail:mCherry fluorescence is total nuclear fluorescence intensity. Lines and areas are median and 95% confidence interval of median, respectively.

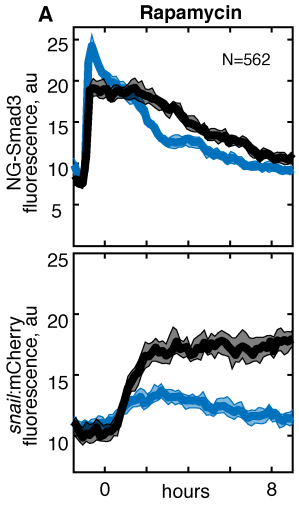

Figure 3.21: (A) Dynamics of NG-Smad3, *top*, and snail:mCherry, *bottom* in response to Tgf-β1 (2.4 ng/mL, added at t=0) with or without addition of drug. NG-Smad3 fluorescence is median nuclear fluorescence intensity. snail:mCherry fluorescence is total nuclear fluorescence intensity. Lines and areas are median and 95% confidence interval of median, respectively.

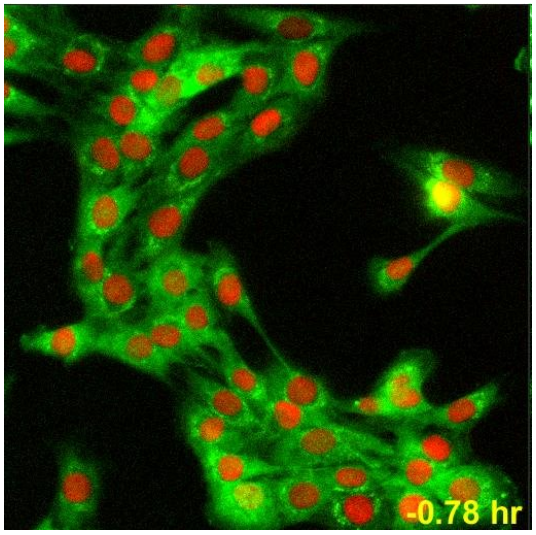

Figure 3.22: MOVIE: Dynamics of endogenous NG-Smad3 and snail:mCherry in single cells. Cells were treated with Tgf-β1. Tgf-β1 is added immediately before t = 0 to a final concentration of 2.4 ng/mL. The time of image acquisition is listed in the bottom right corner. Fluorescence from NG-Smad3 is in green, and fluorescence from snail:mCherry is in red.

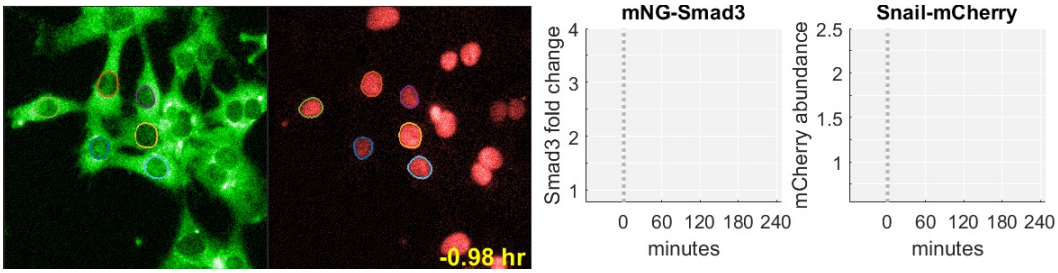

Figure 3.23: MOVIE: Example of quantification of NG-Smad3 and snail:mCherry dynamics in the same cells. Cells were Tgf-β1 (2.4 ng/mL) immediately before t = 0. The time of image acquisition is listed in the bottom right corner. NG-Smad3 fluorescence is colored green (far left panel) and snail:mCherry fluorescence is colored red (second to left panel).

3.5 Materials and Methods

Expression Construct.

The human Smad3 cDNA was a gift from Joan Massague (Addgene; plasmid 27010). Human Smad3 and mouse Smad3 contain 100% sequence identity. The mNeonGreen (NG) gene was obtained from Allele Biotechnology (ABP-FP-MNEONSA). The mCerulean3-C1 cDNA was a gift from Klaus Hahn (Addgene; plasmid 22030). The NG-Smad3 construct was placed downstream of a CMV promoter, and the mCerulean3 gene was fused with a 3x nuclear localization sequence (NLS) and placed downstream of an SV40 promoter.

Cell Culture.

C2C12 cells (American Type Culture Collection, CRL-1772) were cultured at 37 °C and 5% (vol/vol) CO2 in DMEM (ThermoFisher Scientific; 11995) supplemented with 10% (vol/vol) FBS (Invitrogen; A13622DJ), 100 U/mL penicillin, 100 μg/mL streptomycin, 0.25 μg/mL amphotericin, and 2 mM GlutaMAX (Gibco). Recombinant human Tgf-β1 (PeproTech; 100-21) was used at concentrations indicated in figures or text.

CRISPR-Cas9 mediated genome editing of Smad3

We designed a homology directed repair (HDR) template plasmid to insert mNeonGreen coding region at the immediate N-terminus of Smad3 (removing the Smad3 ATG). The plasmid had 900 base pairs of homology to smad3 immediately upstream of the smad3 ATG (5' Homology Arm) and 900 base pairs of homology to smad3 immediately downstream of the smad3 ATG (3' Homology Arm). The HDR mediated editing of genomic smad3 results in expression of a functional, fluorescently labeled Smad3 that is identical to the construct used (Frick et al., 2017) To perform the homology directed repair mediated knock in of mNeonGreen, cells were transfected with HDR template plasmid, guideRNA synthesized by IDT, and Cas9 protein from IDT (Alt-R S.p. Cas9) using Lipofectamine LTX (Life Technologies). We generated 4 homozygous, and 1 heterozygous NG-Smad3 clonal cell lines. One homozogous clonal cell line (clone #43) was chosen for generation of two color reporter cell lines (see below).

CRISPR-Cas9 mediated genome editing of Snail

We designed a homology directed repair (HDR) template plasmid to insert a p2a-mCherry-3xNLS at the immediate 3' end of genomic snail. The plasmid had 900

base pairs of homology to snail immediately upstream of the snail stop codon (5' Homology Arm) and 900 base pairs of homology to snail immediately downstream of the snail stop codon (3' Homology Arm). The stop codon is removed upon successful homology directed repair. The p2a linker results in cleavage of mCherry protein from Snail protein cotranslationally (Osborn et al., 2005). The mCherry contains a 3x repeat of a nuclear localization sequence at its C-terminus to localize mCherry fluorescence to the nucleus to facilitate segmentation, tracking, and quantification. To perform the homology directed repair mediated knock in of mCherry-3xNLS-PEST, cells were transfected with HDR template plasmid, guideRNA synthesized by IDT, and Cas9 protein from IDT (Alt-R® S.p. Cas9) using Lipofectamine LTX (Life Technologies). We generated 6 heterozygous Snail-p2a-mCherry clonal cell lines, in a parental cell line homozygously expressing NG-Smad3 (see above). Analysis shown here is from clone #116.

Single cell cloning for Smad3

To obtain clones with successfully edited genomic Smad3, the cells were sorted for mNeonGreen fluorescence and then plated as single cells in 96 well plates via limiting dilution. The clonal populations that grew up were screened for mNeonGreen fluorescence, NG-Smad3 nuclear accumulation upon Tgf-β treatment, characteristic localization of Smad3 (absence of puncta, absence of membrane blebbing, correct subcellular localization, etc), and genomic PCR to check for number of alleles edited (heterozygosity or homozygosity) as well as sequencing. Genomic PCR was performed on whole cell lysates using KOD Xtreme Hot Start DNA Polymerase (Millipore-Sigma).

Single cell cloning for double reporter cells To obtain cells expressing both endogenous NG-Smad3 and endogenous snail:mCherry cloning was done as described for Smad3 except cell sorting was performed based on mCherry signal, and cells were screened for nuclear localization of mCherry, increase in mCherry expression upon Tgfβ treatment and PCR analysis. Genomic PCR was performed on whole cell lysates using KOD Xtreme Hot Start DNA Polymerase (Millipore-Sigma).

Live-Cell Imaging.

Cells were grown on 96-well or 24-well glass-bottomed plates (Griener Bio-One; 662892) and cultured in imaging media – FluoroBrite DMEM (Gibco-Life Technologies; A18967) containing 10% (vol/vol) FBS, 1 mM glutamax (Gibco-Life Technologies; 35050), 1 mM sodium pyruvate, 100 U/mL penicillin, 100 μg/mL

streptomycin, and 0.25 μg/mL amphotericin. Time-lapse imaging was performed with a Zeiss Axio Observer.Z1 inverted fluorescence microscope under incubation [37 °C and 5% (vol/vol) CO2, with humidification] on a motorized stage, using a 40×, 1.4-N.A. Plan Apo Oil Objective with Immersol 518F (Zeiss; 444970-9010) and an Orca Flash 4.0 V sCMOS camera. Images were acquired at 4-10 minute intervals. Cells were imaged for at least 1 h before stimulation with Tgf-β1 (PeproTech; 100-21). When drug or Tgf-β was added to cells, 20 μL of stock reagent was added, followed by mixing with a 50 μL pipette. Buffer-only media were added in 0 ng/mL experiments to control for any effects of adding liquid (e.g., shear stimulation).

Time-lapse movies were quantified after flat-field correction, background subtraction and bleaching correction. We followed the standard methods described by Waters, 2009. Segmentation, tracking, and quantification were done using custom written Matlab scripts (see ref Frick et al., 2017).

Image Analysis, Cell Tracking, and Fluorescence Quantification.

Time-lapse movies were quantified after flat-field correction, bleaching correction, and background subtraction. We followed the standard protocol described by Waters (Waters, 2009). In flat-field correction, to capture the shape of fluorescence illumination, we imaged a well containing media only. We imaged five different positions within the well, and computed the median of the images. Flat-field correction was performed by dividing each experimental image by this media-only image. This procedure was repeated for each fluorescence channel. Bleaching correction was performed for each fluorescence channel. The bleaching profile used for corrections was determined in unstimulated cells. For background correction, images were segmented such that the entireties of cells were broadly outlined, and fluorescence signal from the background was then averaged and subtracted from the image. This procedure was repeated for all images at each time frame.

Fluorescence Quantification. We report the median fluorescence intensity of NG-Smad3 fluorescence in the nuclei and the total fluorescence intensity of snail:mCherry fluorescence in the nuclei. The nuclei of cells were first segmented based on the fluorescence of mCherry. Next, segmented nuclei were tracked across all time frames. Finally, the fluorescence data from the segmented nuclei were extracted. We only tracked and quantified fluorescence from cells that maintained consistent morphology for at least 4.5 h of imaging (30 minutes before ligand addtion and 4 h after ligand addition). Cells that divided, balled up, left the imaging field of view, or displayed some abnormality (e.g., double-nucleated, abnormally large) during

this time frame were excluded. We performed all segmentation, tracking, and fluorescence quantitation steps using custom MATLAB (MathWorks) scripts (available upon request).

Diffusion Maps

We analyzed our data using the Difussion Maps algorithm developed by Coifman and Lafon 2006. Our input is a matrix of the raw fluorescence trajectories that is m cells by n time points. We then compute the time derivatives and integrals over all time points for our data and include these in our matrix, giving a new matrix that is m cells by 3xn time points. The fluorescence values for each trajectory are then shifted and scaled into the range of 0 to 1.

This matrix is then input into the diffusion maps algorithm, which is described here in brief for reference. Given m data points $D = y_1, ..., y_m$ embedded in an ambient Euclidean space E = $\mathbb{R}^m$ close to a smooth manifold M, we construct a graph between points, where the connectivity is based on a Gaussian kernel similarity measure with respect to the Euclidean distance d in e. For a certain scale parameter $\epsilon > 0$, the connectivity between two points $y_i, y_k \in E$ is defined through the kernel matrix $K_{ik} = k(y_i, y_k) = exp(-\frac{d(y_i,y_k)^2}{\epsilon})$. The Diffusion Maps concept is based on the convergence of the normalized graph Lapalacian on this graph to the Laplace-Beltrami operator on the manifold, in the limit $n \to \infty$, $\epsilon \to 0$. For non-uniform time-points, the matrix K will be normalized by an estimation of the density. $P_{ii} = \sum_{k=1}^{n} K_{ik}, \widetilde{K} = P^{-a} K P^{-a}$. We use $a = 1$ for non uniform sampling as in (Coifman, Stephane Lafon, et al., 2005). The kernel matrix $\widetilde{K}$ is then subsequently normalized by the diagonal matrix $D_{ii} = \sum_{k=1}^{n} \widetilde{K}_{ik}$. A non-linear embedding of the manifold is then given by a certain number l of eigenvectors of $A = D^{-1}\widetilde{K}$, scaled by their respective eigenvalue. The eigenvectors of the manifold were determined by the method developed by (Dsilva et al., 2018), which identifies and excludes repeated eigendirections by applying a local linear regression algorithm.

To ensure the best possible manifold was obtained we repeated this process for a range of possible ϵ values, which is the one most significant parameter of the algorithm. Finally, from the 50 manifolds generated, we selected the one that is best able to predict the original fluorescence trajectories from which it was derived. The predictions of the original trajectories are performed using Gaussian Process Regression.

Gaussian Process Regression

We perform Gaussian Process Regression using GPy, Gaussian process framework in Python (GPy, since 2012). We use the Mat52 kernel + white noise kernel, and we specify a non-isotropic GP, that allows one lengthscale paramter per dimension. We input a manifold, X, along with an output manifold, Y. And then run Gpy.model.optimize. We then use the optimized regression to attempt to predict the original output manifold, Y. To do this we run Y2 = model.predict(X), where X is the input manifold and Y2 is the predicted output manfiold. We then rescale the given output Y and the predicted output Y2 in the following way. Ys = (Y-mean(Y)) / std(Y) , Y2s = (Y2-mean(Y2)) / std(Y2). We then make a plot of Ys vs Y2s and determine the normalized mean squared error defined as: NMSE = mean($(Ys - Y2s)^2$).

Iterative indirect immunofluorescence

Following live cell imaging performed, cells were fixed with 4% Paraformaldehyde (Electron Microscopy Sciences) for 15 minutes. Cells were then permeablized with 0.5% Triton X-100 (Sigma) for 15 min and washed 5 times with 1X PBS. The following steps are then performed iteratively to achieve sequential antibody staining:

(1) *Antibody Elution*. Sample was washed 4 times with ddH_2O. Residual ddH_2O was aspirated to minimal volume. Subsequent actions are repeated 3 times: 300 μL Elution Buffer (EB) was added to sample and shaken at 290 rpm for 10 min. Then EB was aspirated to minimal volume possible. Sample is then washed once with ddH_2O, and then once with PBS.

(2) *Blocking*. 300 μL 4i blocking solution (sBS) was added to sample and shaken at 290 revolutions per minutes (rpm) for 1 hour. After 1 h sample was washed 6 times with PBS.

(3) *Indirect immunofluorescence, primary antibody stain*. Primary antibody solution, 150 μL, was added to sample and shaken at 290 rpm for either a) 1-2 hr; or b) 12-18 hr at 4°C with rocking. The sample was then washed 4 times with PBS.

(4) *Indirect immunofluorescence, secondary antibody stain*. Secondary antibody solution, 150 μL, was added to sample and shaken at 100 rpm for 1-2 hours at room temperature. The sample was then washed 4 times with PBS.

(5) *Nuclear staining*. 400 μL DNA Stain Solution was added to sample and shaken at 290 rpm for 10 min. After 10 min sample was washed 4 times with ddH_2O. Residual

ddH_2O was aspirated to minimal volume. Our recipe for DNA Stain solution was modified from original in that we used Hoechst rather than DAPI.

(6) *Imaging*. 300 μL Imaging Buffer was added to sample and sample was imaged.

Cell imaging was performed on 24-well glass-bottomed plates (Griener Bio-One; 662892). All steps were carried out at room temperature unless otherwise specified. All wash volumes are 0.5 mL. Buffer formulations can be found in Gut, Herrmann, and Pelkmans (2018). Shaking was performed with a GeneMate Orbital Shaker MP4 (BioExpress). All liquid dispensing and washing steps were performed manually.

Antibodies for immunofluorescence

Primary Antibodies:
SMAD4 (D3R4N) XP Rabbit mAb, Cell Signaling Technology (46535T), 1:800 dilution
Phospho-4E-BP1 (Thr37/46) (236B4) Rabbit mAb, Cell Signaling Technology (2855S), 1:200 dilution
Phospho-SMAD3 (Ser423, Ser425) Antibody (16H5L12), ABfinity Rabbit Monoclonal, Invitrogen (702292) 1:250 dilution.
Phospho-S6 Ribosomal Protein (Ser235/236) (D57.2.2E) XP Rabbit mAb, Cell Signaling Technology (4858S), 1:100 dilution.
EIF4E, Cell Signaling Technology (9742S), 1:100 dilution
Phospho-Akt (Ser473) (D9E) XP Rabbit mAb, Cell Signaling Technology (4060T), 1:400 dilution

Secondary Antibody:
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647, Invitrogen (A-31573), 1:500 dilution.

References

Coifman, Ronald R and Stéphane Lafon (2006). "Diffusion maps". In: *Applied and computational harmonic analysis* 21.1, pp. 5–30.

Coifman, Ronald R, Stephane Lafon, et al. (2005). "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods". In: *Proceedings of the National Academy of Sciences* 102.21, pp. 7432–7437.

Dsilva, Carmeline J et al. (2018). "Parsimonious representation of nonlinear dynamical systems through manifold learning: A chemotaxis case study". In: *Applied and Computational Harmonic Analysis* 44.3, pp. 759–773.

Frick, Christopher L, Clare Yarka, Harry Nunns, and Lea Goentoro (2017). "Sensing relative signal in the Tgf-β/Smad pathway". In: *Proceedings of the National Academy of Sciences*, p. 201611428. DOI: 10.1073/pnas.1611428114.

GPy (since 2012). *GPy: A Gaussian process framework in python*. http://github.com/SheffieldML/GPy.

Gut, Gabriele, Markus D Herrmann, and Lucas Pelkmans (2018). "Multiplexed protein maps link subcellular organization to cellular states". In: *Science* 361.6401, eaar7042.

Osborn, Mark J et al. (2005). "A picornaviral 2A-like sequence-based tricistronic vector allowing for high-level therapeutic gene expression coupled to a dual-reporter system". In: *Molecular Therapy* 12.3, pp. 569–574.

Schmierer, Bernhard et al. (2008). "Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system". In: *Proceedings of the National Academy of Sciences* 105.18, pp. 6608–6613.

Waters, Jennifer C (2009). *Accuracy and precision in quantitative fluorescence microscopy*.

Chapter 4

CONCLUSIONS

This work uncovers how information is transferred from Tgf-β to Smads to gene expression. In Chapter 2, using a live-cell imaging approach, I show that the intracellular signal in the pathway is carried by fold change in Smad3—it is the precise feature of the Smad3 dynamics, and it is the feature that determines target gene expression. Transmitting signal in the relative dynamics of Smad3 enables cells to overcome extrinsic variability. In Chapter 3, by introducing an approach for testing the relationship between input and output dynamics of signaling, I show that the variability in gene expression dynamics are not fully dictated by variability in Smad3 dynamics. Instead, how target gene responds to Smad3 dynamics is modulated by cell intrinsic factors, which in this context were shown to be cell cycle, mTORC1/2, and MEK5. Thus, even in a clonal population of cells, a cell's response to signal depends on its internal state.

Finding fold change detection in the Tgf-β pathway further establishes widespread use of this strategy in cell signaling. The challenge facing signaling pathways is how to accurately transmit signals despite high levels of cell-cell variability in protein expression. Out of the numerous possible ways a cell can overcome this problem—integrating signaling responses from multiple pathways (Cheong et al., 2011), utilizing compensatory crosstalks (Uda et al., 2013), averaging responses with neighboring cells (Cheong et al., 2011), or implementing negative feedbacks (Voliotis et al., 2014)—it is emerging that a primary way cells overcome variability is by sensing signal relative to background. Finding fold-change detection in Tgf-β/Smad signaling brings the list of pathways where fold-change detection has been observed or proposed to six, four of which are major developmental pathways. (Lee et al., 2014; Goentoro and Kirschner, 2009; Cohen-Saidon et al., 2009; Frick et al., 2017; Thurley, Tovey, et al., 2014; Thurley, Gerecht, et al., 2015).

There remain two primary questions about fold-change detection for researchers to answer in these pathways: 1) what properties of the network give rise to precise fold-change, and 2) what is the mechanism by which fold-changes in transcription factor activation are decoded into gene expression. In regards to the first question, in a paper exploring the precise fold-change dynamics observed here and elsewhere,

Nunns and Goentoro (Nunns and Goentoro, 2018; Nunns, 2019) identified that the architectural features that give rise to the linearity underyling fold-change in four different pathways. For the Tgf-β pathway they identified two important features for robust encoding of fold-change: First, the constitutitve nucleocytoplasmic cycling process must be fast. Second, cells must maintain an excess of unphosphoryated R-Smad. Curiously, the expression level of Smad3 in mESCs is extremely low, perhaps to the point where they would be outside of the robust fold-change regime. It will be interesting to test whether these cells exhibit substantial Smad3 signaling variability as a result.

The mechanism by which fold-changes are detected, however, remains an open question. There are multiple simple circuits that can achieve fold-change detection (Goentoro, Shoval, et al., 2009) and it has been proposed that fold-change detection is mediated by type 1 incoherent feed forward loop (IFFL) in the Wnt Pathway (Kim et al., 2017) and NF-κB pathway (Lee et al., 2014). This fold-change detection circuit requires the transcriptional regulator to activate the target gene itself and activate repression of that target gene (either directly or indirectly). In the Wnt/β-Catenin pathway it is proposed that this repressive step is accomplished by β-Catenin itself via cis-regulation—β-Catenin directly binds to a short 11 base pair negative response element in the promoter of target genes which it is activating. In the NF-κB pathway, modeling analysis suggests that the repressive step is carried out by competitor proteins.

Future work will be necessary to determine how fold-change detection is accomplished by Smad3 target genes. To test whether Smad fold-change detection requires cis-regulatory elements, I propose to examine how Smad dynamics are transduced into activation of a synthetic Smad responsive promoter (CAGA12) that contains only repeats of Smad3/4 binding sites (Dennler et al., 1998). Because this synthetic construct lacks regulatory sequences, If expression of this gene correlates with Smad3 fold-change, then the fold-change detection mechanism must be mediated outside of cis-regulatory sequences. I believe the other most compelling candidate to test is Tgf-β/Smad induced microRNAs. Smad3 activation both induces miRNA expression and directly interacts with the microRNA processing protein DROSHA (Davis et al., 2008). How Tgf-β/Smad induced microRNAs affect Smad dynamics or dynamics of Smad target genes, however, is not well understood. Interestingly, microRNAs have been shown to be capable of mediating IFFL FCD in mammalian cells (Strovas et al., 2014). Bioinformatic work will be necessary to determine if

Tgf-β/Smad activated miRNAs also target Smad3 activated target genes, followed by biochemical validation.

In addition, it will be of great interest to determine the percentage of genes that are responsive to fold-change. It remains a possibility that different genes can sense different features of signaling dynamics, and thus a subset of genes may not respond to fold-changes at all. To address this it will be necessary to devise RNA-seq experiments in which distinct features of the signaling response are independently perturbed. An obvious starting point will be to perturb the concentration of Smad3 and the fold-change of Smad3 independently to test whether a fraction of targets are dependent on the absolute concentration of Smad3 rather than the level of Smad3.

Finding fold-change detection in this pathway also reinforces the principle that signal is encoded in transcription factor dynamics. The dependency of cellular outcome on transcription factor dynamics is widespread in cellular systems (Purvis and Lahav, 2013; Adler and Alon, 2018). Indeed, the most robust signaling response in some pathways has been shown to be the full dynamics, rather than any single measured feature of the dynamics (Selimkhanov et al., 2014).

The observation that signaling dynamics often display weak correlation with target gene expression, led us to ask the question in Chapter 3 of how strongly target gene dynamics depend on signaling dynamics. The current approaches to studying transcription factor dynamics and gene expression dynamics (see ref Martin and Sung, 2018), including the approach in Chapter 2, have been sufficient to address the question asked within the respective studies (e.g., does this dynamic more strongly determine this gene outcome) but are insufficient for probing how strongly a target gene's dynamics depend on the full dynamics of transcription factor activation.

One of the primary challenges is finding a low-dimensional representation of the global structure and variation of dynamic trajectories. The approach in Chapter 3 overcomes this by introducing manifold learning as a method for achieving a parsimonious parameterization of dynamics. The small set of learned parameters describe the full dynamics of a single trajectory (a high dimensional set of time point measurements). Any feature of the dynamics is a function of this small set of learned parameters.

Combining this approach with live cell imaging of endogenous Tgf-β/Smad signaling, we find evidence that individual cells vary interpretation of signaling dynamics. While it has long been appreciated that ligand-induced gene response can be dif-

ferent depending on the tissue or cell-type context, this work clearly demonstrates that individual cells within a clonal population of cells modulate their responses to signal.

A major implication from this work is the design of assays and reporters for pathway activity. Often times a reporter of signaling dynamics is used as a proxy for cellular outcome and vice versa. In these studies it is assumed that whether the signaling dynamics are observed to be precise or heterogeneous determines whether the gene output is precise or heterogenous. However, our work has shown directly that cells with identical signaling dynamics can give different gene responses and vice versa. Thus to predict a cell's response to signal, it is imperative to know the cell itself.

The work in this thesis has determined how signal is transmitted through Tgf-β/Smad signaling pathway in single cells. The finding of fold-change detection reinforces that signals are carried by dynamics, and that sensing dynamics relative to background activity is a favored strategy in pathways for overcoming cell-cell variability. In addition, we find that cell's make individualized decisions on how to respond to the information conveyed by signaling dynamics. Thus while cells make great effort to encode signal robustly in pathway dynamics, the signaling outcome is not dictated by these dynamics, but rather is based on a cell's own internal state.

References

Adler, Miri and Uri Alon (2018). "Fold-change detection in biological systems". In: *Current Opinion in Systems Biology* 8, pp. 81–89.

Cheong, Raymond et al. (2011). "Information transduction capacity of noisy biochemical signaling networks". In: *science* 334.6054, pp. 354–358.

Cohen-Saidon, Cellina et al. (2009). "Dynamics and variability of ERK2 response to EGF in individual living cells". In: *Molecular cell* 36.5, pp. 885–893.

Davis, Brandi N et al. (2008). "SMAD proteins control DROSHA-mediated microRNA maturation". In: *Nature* 454.7200, p. 56.

Dennler, Sylviane et al. (1998). "Direct binding of Smad3 and Smad4 to critical TGFβ-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene". In: *The EMBO journal* 17.11, pp. 3091–3100.

Frick, Christopher L, Clare Yarka, Harry Nunns, and Lea Goentoro (2017). "Sensing relative signal in the Tgf-β/Smad pathway". In: *Proceedings of the National Academy of Sciences*, p. 201611428. DOI: 10.1073/pnas.1611428114.

Goentoro, Lea and Marc W Kirschner (2009). "Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling". In: *Molecular cell* 36.5, pp. 872–884.

Goentoro, Lea, Oren Shoval, et al. (2009). "The incoherent feedforward loop can provide fold-change detection in gene regulation". In: *Molecular cell* 36.5, pp. 894–899.

Kim, Kibeom et al. (2017). "Two-element transcriptional regulation in the canonical Wnt pathway". In: *Current Biology* 27.15, pp. 2357–2364.

Lee, Robin EC et al. (2014). "Fold change of nuclear NF-κB determines TNF-induced transcription in single cells". In: *Molecular cell* 53.6, pp. 867–879.

Martin, Erik and Myong-Hee Sung (2018). "Challenges of Decoding Transcription Factor Dynamics in Terms of Gene Regulation". In: *Cells* 7.9, p. 132.

Nunns, Harry (2019). "Linearity in Cell Signaling Pathways". PhD thesis. California Institute of Technology.

Nunns, Harry and Lea Goentoro (2018). "Signaling pathways as linear transmitters". In: *eLife* 7, e33617.

Purvis, Jeremy E and Galit Lahav (2013). "Encoding and decoding cellular information through signaling dynamics". In: *Cell* 152.5, pp. 945–956.

Selimkhanov, Jangir et al. (2014). "Accurate information transmission through dynamic biochemical signaling networks". In: *Science* 346.6215, pp. 1370–1373.

Strovas, Timothy J et al. (2014). "MicroRNA-based single-gene circuits buffer protein synthesis rates against perturbations". In: *ACS synthetic biology* 3.5, pp. 324–331.

Thurley, Kevin, Daniel Gerecht, et al. (2015). "Three-dimensional gradients of cytokine signaling between T cells". In: *PLoS computational biology* 11.4, e1004206.

Thurley, Kevin, Stephen C Tovey, et al. (2014). "Reliable encoding of stimulus intensities within random sequences of intracellular Ca2+ spikes". In: *Sci. Signal.* 7.331, ra59–ra59.

Uda, Shinsuke et al. (2013). "Robustness and compensation of information transmission of signaling pathways". In: *Science* 341.6145, pp. 558–561.

Voliotis, Margaritis et al. (2014). "Information transfer by leaky, heterogeneous, protein kinase signaling systems". In: *Proceedings of the National Academy of Sciences* 111.3, E326–E333.

BIBLIOGRAPHY

Azeloglu, Evren U and Ravi Iyengar (2015). "Signaling networks: Information flow, computation, and decision making". In: *Cold Spring Harbor perspectives in biology* 7.4, a005934.

Barkai, Naama and Stan Leibler (1997). "Robustness in simple biochemical networks". In: *Nature* 387.6636, p. 913.

Cheong, Raymond et al. (2011). "Information transduction capacity of noisy biochemical signaling networks". In: *science* 334.6054, pp. 354–358.

Cohen-Saidon, Cellina et al. (2009). "Dynamics and variability of ERK2 response to EGF in individual living cells". In: *Molecular cell* 36.5, pp. 885–893.

David, Charles J et al. (2016). "TGF-β tumor suppression through a lethal EMT". In: *Cell* 164.5, pp. 1015–1030.

Flanders, Kathleen C, Gerson Ludecke, et al. (1991). "Localization and actions of transforming growth factor-beta s in the embryonic nervous system". In: *Development* 113.1, pp. 183–191.

Flanders, Kathleen C, Nancy L Thompson, et al. (1989). "Transforming growth factor-beta 1: histochemical localization with antibodies to different epitopes." In: *The Journal of Cell Biology* 108.2, pp. 653–660.

Goentoro, Lea and Marc W Kirschner (2009). "Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling". In: *Molecular cell* 36.5, pp. 872–884.

Green, Jeremy BA, Helen V New, and JC Smith (1992). "Responses of embryonic Xenopus cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm". In: *Cell* 71.5, pp. 731–739.

Heine, UI et al. (1991). "Localization of transforming growth factor-beta 1 in mitochondria of murine heart and liver." In: *Cell regulation* 2.6, pp. 467–477.

Hough, Chris, Maria Radu, and Jules JE Doré (2012). "Tgf-beta induced Erk phosphorylation of smad linker region regulates smad signaling". In: *PloS one* 7.8, e42513.

Lee, Robin EC et al. (2014). "Fold change of nuclear NF-κB determines TNF-induced transcription in single cells". In: *Molecular cell* 53.6, pp. 867–879.

Lo, Roger S and Joan Massagué (1999). "Ubiquitin-dependent degradation of TGF-β-activated Smad2". In: *Nature cell biology* 1.8, p. 472.

Massague, Joan (1990). "The transforming growth factor-beta family". In: *Annual review of cell biology* 6.1, pp. 597–641.

Mullen, Alan C et al. (2011). "Master transcription factors determine cell-type-specific responses to TGF-β signaling". In: *Cell* 147.3, pp. 565–576.

Nelson, DE et al. (2004). "Oscillations in NF-κB signaling control the dynamics of gene expression". In: *Science* 306.5696, pp. 704–708.

Schmierer, Bernhard et al. (2008). "Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system". In: *Proceedings of the National Academy of Sciences* 105.18, pp. 6608–6613.

Selimkhanov, Jangir et al. (2014). "Accurate information transmission through dynamic biochemical signaling networks". In: *Science* 346.6215, pp. 1370–1373.

Shinar, Guy and Martin Feinberg (2010). "Structural sources of robustness in biochemical reaction networks". In: *Science* 327.5971, pp. 1389–1391.

Sigal, Alex et al. (2006). "Variability and memory of protein levels in human cells". In: *Nature* 444.7119, p. 643.

Spemann, Hans and Hilde Mangold (1924). "Über induktion von Embryonalanlagen durch Implantation artfremder Organisatoren". In: *Development Genes and Evolution* 100.3, pp. 599–638.

Sporn, Michael B et al. (1983). "Polypeptide transforming growth factors isolated from bovine sources and used for wound healing in vivo". In: *Science* 219.4590, pp. 1329–1331.

Tay, Savaş et al. (2010). "Single-cell NF-κB dynamics reveal digital activation and analogue information processing". In: *Nature* 466.7303, p. 267.

Uttamsingh, S et al. (2008). "Synergistic effect between EGF and TGF-β1 in inducing oncogenic properties of intestinal epithelial cells". In: *Oncogene* 27.18, p. 2626.

Voliotis, Margaritis et al. (2014). "Information transfer by leaky, heterogeneous, protein kinase signaling systems". In: *Proceedings of the National Academy of Sciences* 111.3, E326–E333.

Wahl, Sharon M et al. (1987). "Transforming growth factor type beta induces monocyte chemotaxis and growth factor production". In: *Proceedings of the National Academy of Sciences* 84.16, pp. 5788–5792.

Xing, Yingying et al. (2015). "hCLP46 increases Smad3 protein stability via inhibiting its ubiquitin-proteasomal degradation". In: *Protein & cell* 6.10, pp. 767–770.

Zieba, Agata et al. (2012). "Intercellular variation in signaling through the TGF-β pathway and its relation to cell density and cell cycle phase". In: *Molecular & Cellular Proteomics* 11.7, pp. M111–013482.