Memory and Decoding in Signaling Transduction Pathways

Thesis by
Kibeom Kim

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
Defended May 23, 2019

Kibeom Kim
ORCID: 0000-0002-0764-4875

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Lea, who has guided me with patience and inspired with her creative energy and dedication to science. My deep appreciation goes out to our fantastic lab members, especially Harry, Chris, and Noah from the signaling team, who provided many thoughtful discussions and encouragement. I wish the best for Andrew, who has worked closely with me in my projects and is now pursuing his own graduate studies. I am grateful for the genuine friendships made throughout my graduate study, both within lab and outside, that have enriched my journey. Lastly, I would like to thank my family: my parents for their unconditional love and support, and Kay and Liz, who I can always count on for encouragement and advice.

ABSTRACT

Intercellular communication allows cells to broadcast and receive necessary information for decision making, and is essential for development, growth, and maintenance of a community of cells in a multicellular organism. Signaling pathways are highly conserved systems of communication between cells, each composed of a distinct network of protein interactions that detect extracellular signal and transduce the signal information for cellular response. A signaling pathway typically encodes information from signaling events into dynamics of second messengers, intracellular molecules in the signaling pathway that activate in response to signal and initiate cellular response. Therefore, understanding how information is encoded in second messenger dynamics, and how transcriptional machinery decode and generate output response is an important aspect in investigating how signaling information is transduced inside a cell. In the first chapter, we investigate the timescales of memory in endogenous β-catenin and Smad3, second messengers in the Wnt and Tgf-β pathways, through single cell timelapse microscopy. The findings demonstrate that both second messengers have short memory and high cell-to-cell variability, and that their memory is tunable through modulating cellular contexts. In the second chapter, we investigate decoding of information from β-catenin in the Wnt pathway. We identify a novel 11-bp DNA element that recruit β-catenin for transcriptional suppression. This negative regulatory element is shown to act in conjunction with the canonical Wnt responsive element to form an incoherent feedforward loop (IFFL). Through mathematical simulations, we present how the IFFL circuit can generate complex output functions in decoding β-catenin dynamics, which include those that confer robustness against perturbations in signaling response such as band-pass filtering and fold change detection.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. Kim, K. & Goentoro, L. Choosing the right input in cell signaling. *Science* **361,** 643–644 (2018).
https://doi.org/10.1126/science.aau6457

   K.K participated in writing of the perspective.

2. Kim, K. *et al.* Two-element transcriptional regulation in the canonical Wnt pathway. *Current Biology* **27,** 2357–2364 (2017).
https://doi.org/10.1016/j.cub.2017.06.037

   K.K participated in design and execution of experiments, figure development, and writing of the manuscript. For a full acknowledgement of contributions, please refer to 3.7: Work contributions.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

INTRODUCTION

Decision making and information transfer in biology

Life adds an interesting layer to the physical world with its teleonomic property, which can be best described as 'purpose-driven behavior that emerges from physical law'[1]. Any physical reaction that is touched by life can be weighed in relation to a subjective purpose and thus gains a layer of consequence. An otherwise indifferent mass emission of thermal energy from combustion of hydrocarbons could be viewed as costly for survival of organisms in the immediate vicinity of the reaction, but beneficial for other organisms such as pine trees that require such energy for reproduction [2]. The consequences can therefore be measured in cost and benefit, and can further decompose into specific values such as risk, utility, and efficiency.

The consequences of physical reactions that an organism could encounter or embody are closely intertwined with behaviors of the organism, as these behaviors could harness, avoid, or cope with the reactions. In this light, a set of behaviors that triggers specific events with costs or benefits for an organism could be considered a decision. Decision making generally requires information for the biological actor. For a mouse, the flight and freeze response is triggered by visual stimulus of an approaching, looming shadow [3]. For *E.coli*, a decision to synthesize proteins required to import and digest lactose require the cell to sense paucity of glucose and abundance of lactose [4]. Indeed, analyzing how specific information is sensed, transduced, and interpreted is essential in investigating the design and evolution of decision-making modules in biology.

Although decision-making encompasses a vast range of scale and form, here we narrow down the scope to cellular decision-making for multicellular organisms. At a cellular level, decision-making modules are mediated by molecular interactions. A module could be composed of macromolecules like proteins, lipids, and functionally relevant parts of DNA/RNA, and interactions that arise within the architecture bring about decision-making functions. Take, for a simple example, the aforementioned E.coli's decision-making module for lactose digestion mediated by the lac-operon. The combination of constant repression by Lac repressor protein in the absence

of allolactase (metabolite of lactose) and activation requiring catabolite activator protein (triggered in low cAMP levels) at the transcriptional site of *lac* genes gives rise to a transcriptional 'AND' gate that initiates only when glucose levels are low and lactose levels are abundant. As we will soon cover, architecture of metazoan decision making modules can be far more complex in structure and functions.

In my thesis, I focus on information transfer in metazoan signaling pathways that enable a community of cells to coordinate decision making. In the first part of the thesis, in collaboration with Chris Frick, I characterized variability in the temporal dynamics of signaling proteins, and investigating how memory may be encoded by the dynamics. In the second part of the thesis, in collaboration with Jaehyoung Cho and Thomas Hilzinger, I uncovered a novel mechanism of transcriptional regulation in the Wnt pathway that enables decoding of dynamic information encoded in signal transduction. Before unwrapping these topics, I would first like to cover relevant background on information transfer in signaling pathways that motivate the research in my thesis.

Multi-cellular decision making and signaling pathways

A multicellular organism requires communication within its members for regulation of roles and coordination of activity to facilitate shared goals of survival and reproduction as a whole. Consider development, an indispensable process for multicellular life: each cell in the developing embryo must express a proper set of genes and mobilize based on its relative location and time point, thus deciding cell fate for itself and its daughters. Precise coordination is required for a single fertilized egg to develop into a mature organism such as a person, where trillions of cells have determined their respective role among hundreds of possible cell types encoded by the genome [5, 6]. Each cell requires significant contextual information to achieve this extraordinary feat of spatiotemporal coordination. Much of the information is already encoded within the cell, with memory of past decisions and context manifest in forms such as epigenetic regulation and existing configuration of gene expression [7]. This encoding by intracellular modules allow organization of cell fates into a hierarchical tree of development [8], or - in another perspective- flexibly defined trajectories in Waddington's landscape [9], in which past decisions alter the probability of potential fates.

However, extracellular information is also essential for a cell to hone down its positional context and timing of its decision-making: cells must communicate with

other cells nearby and over long distances. Seven metazoan signaling pathways have emerged early in evolutionary history to fulfill this need of intercellular information transfer: Tgf-4β, canonical Wnt, nuclear receptors (NR), receptor tyrosine kinase (RTK), the Notch/Delta, Hedgehog, and Janus kinase/signal transducer and activator of transcription (JAK/STAT) pathways [10, 11]. Functional pathways in the list can be observed in organisms as ancient in the phylogenetic tree as cnidarians (jellyfish) and ctenophores (comb jellies). Moreover, core components of the pathways are highly conserved from cnidarians to vertebrates.

How signaling pathways function in a cell

How do signaling pathways carry out intercellular communication? Signaling pathways are composed of signaling ligands that travel in extracellular space, receptors that detect the ligands, and intracellular network of proteins that transduce the signal for response within the cell. (Figure 1.1)

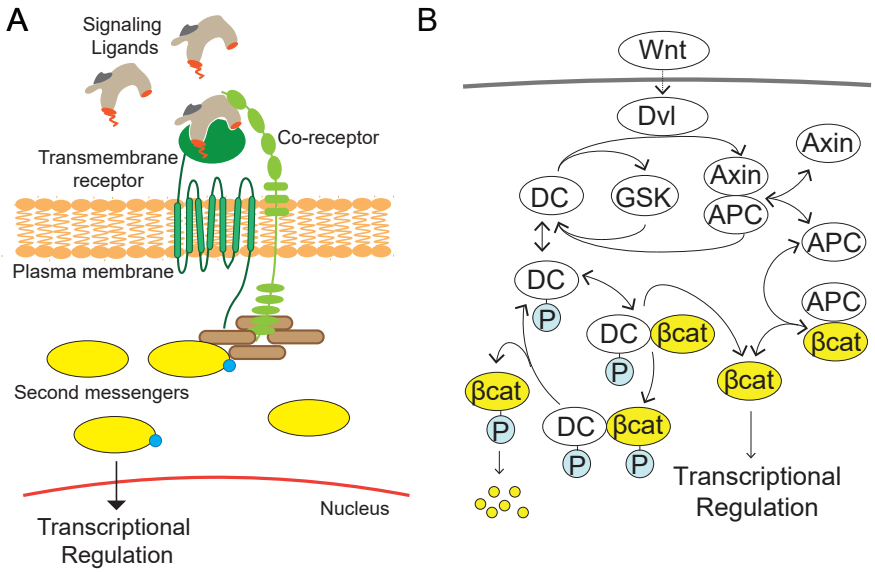


Figure 1.1: Structure of signaling pathways
(A) Example illustration of signalosome at the membrane- composed of signal ligands, receptors, transduction components, and second messengers.
(B) Example illustration intracellular signaling transduction network: Wnt transduction pathway

With the exception of nuclear receptor pathway, the receptors of signaling ligands reside in the cell's plasma membrane, spanning the membrane in such way that ligand-sensing domains are exposed to extracellular space while signal transducing domains project into the cytoplasm [12]. Receptor binding with specific ligands

triggers oligomerization and conformational change that activate the intracellular domains of the receptors. Once activated, the intracellular signaling domains of the receptors recruit a variety of factors such as kinases and scaffolding proteins necessary for signal transduction, forming a signalosome complex (Figure 1.1A). From the signalosome complex, a cascade of biochemical reactions propagates, which ultimately leads to accumulation of pathway specific transcription factors, also known as **second messengers/ signal proteins**, that regulate gene expression in the nucleus (Figure 1.1B).

While following the aforementioned generality, the molecular architecture of signal transduction are quite divergent in signaling pathways. For example, the signalosome activation in the Wnt pathway recruits and inactivates main components of β-catenin degradation complex, allowing β-catenin to stabilize, complex with Tcf transcription factors, and translocate to the nucleus for gene regulation [13]. in the Tgf-β pathway, Smad, a second messenger, is activated by the signalosome through phosphorylation, associates with common partner Smad, and translocates into the nucleus [14]. In contrast, in the Notch/Delta pathway, Notch has a dual role as both signaling receptor for Delta ligand and second messenger after cleavage following ligand binding [15].

Based on the general structure, information transfer through signaling pathway seems quite straightforward: information in the form of signaling ligands is detected by receptors, transduced by activation and accumulation of second messengers, and decoded into transcriptional response in the nucleus. Modulation at different steps of signaling, however, allows for high complexity in information transfer.

Information transfer through dynamics of second messengers

The evolution of Metazoa saw fit to reuse the same pathways again and again with rising complexity in an organism's form and functions. Wide range of roles that involve signaling pathways include specialized processes of development such as digit formation or ocular development, organ homeostasis, immunity regulation, and control of cell proliferation [11]. How are these signaling pathways able to support information transfer in such diverse and dissimilar settings, varying in timescales, informational context, desired output response, and etc.? A simple answer to this question is that the structure of signaling pathways is modular and adaptable in such a way that there is ample room for evolving complexity and tuning of signaling functions without changing the core mechanisms of the pathway.

The extent of complexity in signaling pathways is staggering. Each pathway has multiple homologous signaling ligands and receptors allowing combinatorial specificity of ligand-receptor complexes: varying cellular response depending on specific ligand-receptor signalosome is ubiquitous in signaling pathways. The receptors have ancillary chaperones and co-receptors that affect ligand binding and selectivity. Crosstalk introduces combinatorial effect of information transfer between signaling pathways.

In the decoding layer of signaling pathways, second messengers can interact with multitudes of transcription factors (TFs)- activators, repressors, epigenetic factors, inhibitors, etc. These interactions are controlled spatiotemporally, such as timing of specific TF expression or locales of TF binding sites in the genome. Further, the target gene output from signaling pathways can be linked to negative or positive regulation of pathway activation, creating feedback control.

Transfer of complex information from the signalosome to the transcriptional response in the nucleus depends on second messengers. Richness of information seems to inpart derive from dynamics of signal proteins. Dynamics can be defined as 'spatiotemporal patterns of activity' [16]: in signaling, it describes behavior of proteins over time in a set spatial arrangement, *e.g.* accumulation in the nucleus, export into extracellular space, etc. Aspects of dynamics can be characterized in terms such as amplitude, duration, and frequency of activity.

One of the earliest demonstrations of how information is encoded into signaling dynamics was presented in the involvement of Erk signaling in PC12 differentiation. Prior investigations had demonstrated that stimulation of PC12 cells by signaling ligands NGF (nerve growth factor) and EGF (epidermal growth factor) led to varying outcomes; differentiation with the former, and proliferation with the latter ligand. The ligands were found to induce different dynamics in ERK, one sustained and the other transient response, and this difference in the dynamics drove the distinct differentiation outcome [17]. Since then, there has been a plethora of evidence presenting that dynamics of signaling proteins dictate signal outcome, such as rate-sensing in the Tgf-β [18], frequency modulation in Calcium [19] and Erk [20], and oscillatory dynamics in Nf-κB [21] and Notch [22] pathways. Fidelity of information transfer through dynamics of signal proteins require dealing with a ubiquitous feature in cellular processes: variability.

Noise and variability in second messenger dynamics

That genetically identical cells in homogeneous conditions can have variability in their decision-making has been an often-repeated observation made as early as the 1970s [23]. An exemplary case of this phenomenon in metazoan system is presented in decision making for apoptosis in cells exposed to TRAIL (TNF-related apoptosis inducing ligand) [24]. In the study, the investigators reported heterogeneous response in clonal cell lines, with variance in time to death after TRAIL exposure, where some cells did not die at all. Upon analysis of relevant factors in sister cells, the authors concluded that variability in several proteins exert combinatorial control in an important kinetic rate in the apoptotic decision. Further, they demonstrated that there is a degree of heritability from the mother cell that decays over time, suggesting a time frame of memory that carries over a cell generation.

At the biochemical level, protein levels between cells may vary due to inherent stochasticity in diverse processes such as transcription, translation, and degradation. For second messengers bound within the network of the pathway, propagation of noise from other proteins in the network could contribute to the variability. For example, analysis of the mathematical model of the Wnt pathway [25], generated from measured pathway kinetics *in vivo*, revealed that β-catenin is highly sensitive to variation in in the factors in the pathway such as scaffolds and kinases in the β-catenin degradation complex [26]. Certain features of the network can also reduce variability: negative feedback reduces noise at the cost of limiting the dynamic range of the second messenger [27, 28].

The implications of variability on signaling proteins have cast a question on how signaling pathways can reliably transmit information, particularly on how receptor activated stimulation can lead to robust cellular response in cells. Underlying this question was evidence of significant variability in signal proteins from the aforementioned analysis of the Wnt pathway, as well as single cell measurements of signal proteins in the Tgf-β [29], ERK [30], and NF-κB pathways [31]. Several strategies have been proposed that enable cells to address variability in signaling, including compensation by cross-talk with other pathways [32] or averaging cellular response of multiple cells [28], but one of the proposed strategies utilize information within the dynamics: In each of the aforementioned pathways, theoretical and experimental evidence demonstrated that cells could be sensing relative levels of amplitude in signal dynamics than absolute levels during signal activation [29–

31]. In other words, the dynamic output of signaling molecules were much more consistent when calculating fold change of activated level over basal level, and transcriptional output of some target genes correlated more closely to the fold change than to the absolute levels.

Memory and temporal profiles of variability

Closely related to phenomenon of cell-to-cell variability is the temporal profile of variability within each cell. Population variability arises as protein levels or reaction rates in each cell to deviate from the population norm. As important to how the variation is manifest in the population is how the variability behaves across time through fluctuations in individual cells. For example, protein levels in two populations may have identical variability in a static measure such as standard deviation, but frequency and amplitude of fluctuations in protein levels between cells of the two populations could be drastically different.

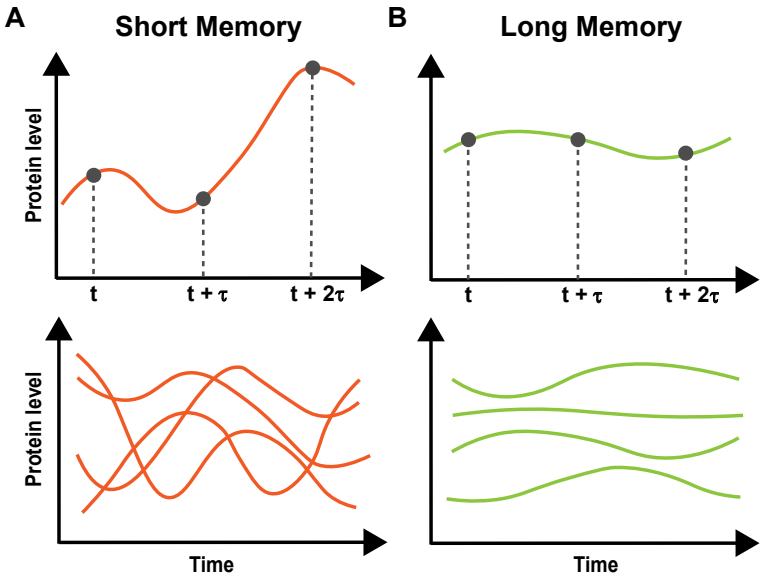


Figure 1.2: Memory of protein levels
(A) Top: An example trace of protein levels over time with short memory. Fluctuations cause correlation between time points to decay quickly with increasing time interval (τ)
Bottom: How protein levels in a population may behave with short memory
(B) Top: An example trace of protein levels over time with long memory. Stable levels cause correlation between time points to decay slowly with increasing time interval (τ)
Bottom: How protein levels in a population may behave with long memory

Memory is an important concept in characterizing temporal dynamics of fluctu-

ations, defined as 'the rate of decay of statistical dependence of two points with increasing time interval between the two points' (Figure 1.2) [33]. In the aforementioned case of cells' apoptotic decision-making, memory of factors that determine apoptotic outcome is shorter than a cell cycle, leading sister cells to diverge in their decisions within one generation.

Using single cell analysis, two studies have focused on investigating memory and variability of protein levels: Austin *et al.* (2006) investigated memory of destabilized GFP expressed by synthetic gene circuits in *E.coli*, demonstrating strong influence of degradation and dilution on memory [34]. The authors further showed that introduction of negative autocorrelation led to dampening of noise amplitude and shorter memory profiles. Sigal *et al.* (2006), investigated variability and memory of endogenous housekeeping genes in human cell lines, demonstrating that proteins have distinct memory profiles and that memory has close correlation with variability [35].

Memory in signaling second messengers

Thus far, memory of Metazoan second messengers of signaling pathways has not been characterized except in cases that exhibit oscillations and frequency modulation in signal dynamics [36, 37]. Although both Wnt and Tgf-β signaling pathways have negative feedback components that could theoretically enable periodicity in their respective second messengers, evidence from the dynamics do not suggest oscillations nor frequency modulation in either pathways [38–40]. Because signaling dynamics carry important weight in information transfer, we propose to investigate how memory and temporal variability operates in second messengers and affects cellular decision making. In Chapter 2 of this thesis, I cover a collaborative effort between Chris Frick and myself that establishes a system to characterize memory of β-catenin and Smad3 at a single cell level, and investigates how the property is modulated in various contexts.

Decoding information from signaling dynamics

There is ample evidence that gene expression can discriminate specific information encoded in dynamics of signal molecules and generate contextually appropriate output. For many of the phenotypes, clear understanding of molecular mechanisms behind the decoding yet remains elusive. In the mechanisms that have been elucidated to decode dynamics, main transcriptional strategies include modulation of

kinetics and and feedforward regulations.

In several studies, binding and promoter kinetics involved with transcription have been demonstrated to decode specific information from signaling dynamics. One such example is from investigation of stress-response signaling in budding yeast S. cerevisae, which involves secondary messenger Msn2. Hao and Shae (2011) observed that varying stress stimuli evoked different modes of Msn2 response dynamics: glucose limitation stimulated duration and frequency modulation, oxidative stress modulated duration of initial peak, and oxidative stress primarily affected amplitude of the dynamics [41]. To investigate how genes could decode Msn2 dynamics and thereby differentiate varying stresses, the authors focused on modeling expression based on modulating kinetics. Both theoretical and experimental analysis demonstrated that tuning binding kinetics of transcription factors influenced sensitivity of response to amplitude modulation, and tuning promoter kinetics (switching on-off states) influenced sensitivity to duration and frequency modulation.

In a similar vein, dynamics of calcium, a secondary messenger with many regulatory roles including mitochondrial and immune control, is known to be decoded through binding kinetics of its partner secondary messengers [42]. JNK and NF-κB have slow degradation and low affinity for calcium, while NFAT has fast degradation and high affinity for the ion. The kinetics therefore bias secondary messenger activation such that JNK and NF-κB activate in strong but transient calcium pulses, and NFAT activates with low, sustained calcium dynamics.

Coherent feedforward regulation also has a role in decoding dynamics. One of the transcriptional products in Erk signaling is c-fos, a regulatory transcription factor. Erk phosphorylates c-fos and therefore stabilizes it. Thus, new c-fos products rapidly degrade in transient Erk signaling as Erk is unavailable, but stabilize with sustained Erk dynamics [43]. Feedforward regulation to sense sustained dynamics is also observed with NF-κb. In host defense in response to LPS, Nf-κb translocates to the nucleus to acts as a weak activator in the transcription of Il6. Transient activation of Nf-κb thus leads to weak expression of Il6 that is quickly repressed. However, Nf-κB also induces expression of C/EBPδ, which cooperates with Nf-δB for strong activation of Il6. This feedforward regulation involving C/EBPγ allows for Il6 expression during sustained activation of Nf-κB.

Incoherent feedforward loop (I1-FFL), in which an input activates and concomitantly represses a target, has also been implicated in decoding dynamics in second messengers. With increasing cases that fold change detection occurs in multiple

signaling pathways, I1-FFL has been theorized as a possible transcriptional circuit that could detect fold change dynamics of second messengers [29, 44]. Indeed, evidence suggests that transcription of Il8 detects fold change through activation by Nf-κB and 'repression' through promoter competition of transcriptionally inactive p50 homodimer, another target of NF-κB [31].

Decoding in the Wnt pathway

Previous examples of decoding second messenger dynamics demonstrate general strategies and tuning that cells incorporate to sense specific information. Chapter 3 of my thesis focuses specifically on how signal in the Wnt pathway is decoded. *Siamois* and *Xnr3*, Wnt target genes in development of *Xenopus laevis*, exhibit particular behavior. The induction of these developmental genes in insensitive to perturbations that alter absolute levels of β-catenin, but responding robustly to ligand-receptor activation [26]. The evidence suggests that the genes have transcriptional mechanisms that could filter variability in the absolute levels, possibly responding to fold change response of β-catenin. Building on this information, my collaborators and I discover of a new transcriptional regulation operating in the Wnt pathway, thereby adding to the understanding of how signal dynamics can be decoded.

1.1 Bibliography

1. Olsman, N. A. *Architecture, Design, and Tradeoffs in Biomolecular Feedback Systems* PhD thesis (California Institute of Technology, 2019).
2. Ne'eman, G., Goubitz, S. & Nathan, R. Reproductive traits of Pinus halepensis in the light of fire–a critical review. *Plant Ecology* **171,** 69–79 (2004).
3. Yilmaz, M. & Meister, M. Rapid innate defensive responses of mice to looming visual stimuli. *Current Biology* **23,** 2011–2015 (2013).
4. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology* **3,** 318–356 (1961).
5. Bianconi, E. *et al.* An estimation of the number of cells in the human body. *Annals of human biology* **40,** 463–471 (2013).
6. Carroll, S. B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature* **409,** 1102 (2001).
7. Cavalli, G. *Chromatin and epigenetics in development: blending cellular memory with cell fate plasticity* 2006.
8. Zhou, J. X. & Huang, S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends in genetics* **27,** 55–62 (2011).
9. Waddington, C. *The strategy of the genes: a discussion of some aspects of theoretical biology. 1957* (London: Allen & Unwin).
10. Nichols, S. A., Dirks, W., Pearse, J. S. & King, N. Early evolution of animal cell signaling and adhesion genes. *Proceedings of the National Academy of Sciences* **103,** 12451–12456 (2006).
11. Babonis, L. S. & Martindale, M. Q. Phylogenetic evidence for the modular evolution of metazoan signalling pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20150477 (2017).
12. Gerhart, J. 1998 Warkany lecture: signaling pathways in development. *Teratology* **60,** 226–239 (1999).
13. Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149,** 1192–1205 (2012).
14. Moustakas, A. & Heldin, C.-H. The regulation of TGFβ signal transduction. *Development* **136,** 3699–3714 (2009).
15. Mumm, J. S. & Kopan, R. Notch signaling: from the outside in. *Developmental biology* **228,** 151–165 (2000).
16. Behar, M. & Hoffmann, A. Understanding the temporal codes of intra-cellular signals. *Current opinion in genetics & development* **20,** 684–693 (2010).

17. Marshall, C. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80,** 179–185 (1995).

18. Warmflash, A. *et al.* Dynamics of TGF-β signaling reveal adaptive and pulsatile behaviors reflected in the nuclear localization of transcription factor Smad4. *Proceedings of the National Academy of Sciences* **109,** E1947–E1956 (2012).

19. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455,** 485 (2008).

20. Albeck, J. G., Mills, G. B. & Brugge, J. S. Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Molecular cell* **49,** 249–261 (2013).

21. Nelson, D. *et al.* Oscillations in NF-κB signaling control the dynamics of gene expression. *Science* **306,** 704–708 (2004).

22. Nandagopal, N. *et al.* Dynamic ligand discrimination in the notch signaling pathway. *Cell* **172,** 869–880 (2018).

23. Spudich, J. L. & Koshland, D. E. Non-genetic individuality: chance in the single cell. *Nature* **262,** 467 (1976).

24. Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459,** 428 (2009).

25. Lee, E., Salic, A., Krüger, R., Heinrich, R. & Kirschner, M. W. The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS biology* **1,** e10 (2003).

26. Goentoro, L. & Kirschner, M. W. Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling. *Molecular cell* **36,** 872–884 (2009).

27. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405,** 590 (2000).

28. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *science* **334,** 354–358 (2011).

29. Frick, C. L., Yarka, C., Nunns, H. & Goentoro, L. Sensing relative signal in the Tgf-β/Smad pathway. *Proceedings of the National Academy of Sciences* **114,** E2975–E2982 (2017).

30. Cohen-Saidon, C., Cohen, A. A., Sigal, A., Liron, Y. & Alon, U. Dynamics and variability of ERK2 response to EGF in individual living cells. *Molecular cell* **36,** 885–893 (2009).

31. Lee, R. E., Walker, S. R., Savery, K., Frank, D. A. & Gaudet, S. Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Molecular cell* **53,** 867–879 (2014).

32. Uda, S. *et al.* Robustness and compensation of information transmission of signaling pathways. *Science* **341,** 558–561 (2013).

33. Beran, J. *Statistics for long-memory processes* (Routledge, 2017).

34. Austin, D. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439,** 608 (2006).

35. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643 (2006).

36. Kellogg, R. A. & Tay, S. Noise facilitates transcriptional control under dynamic inputs. *Cell* **160,** 381–392 (2015).

37. Heltberg, M., Kellogg, R. A., Krishna, S., Tay, S. & Jensen, M. H. Noise induces hopping between NF-κB entrainment modes. *Cell systems* **3,** 532–539 (2016).

38. Gouzé, J.-L. Positive and negative circuits in dynamical systems. *Journal of Biological Systems* **6,** 11–15 (1998).

39. Zi, Z., Chapnick, D. A. & Liu, X. Dynamics of TGF-β/Smad signaling. *FEBS letters* **586,** 1921–1928 (2012).

40. Aulehla, A. & Pourquie, O. Oscillating signaling pathways during embryonic development. *Current opinion in cell biology* **20,** 632–637 (2008).

41. Hao, N. & O'shea, E. K. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nature structural & molecular biology* **19,** 31 (2012).

42. Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca2+ response amplitude and duration. *Nature* **386,** 855 (1997).

43. Murphy, L. O., Smith, S., Chen, R.-H., Fingar, D. C. & Blenis, J. Molecular interpretation of ERK signal duration by immediate early gene products. *Nature cell biology* **4,** 556 (2002).

44. Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell* **36,** 894–899 (2009).

Chapter 2

INVESTIGATING MEMORY IN β-CATENIN AND SMAD3 DYNAMICS

2.1 Introduction

Cells respond to intercellular signals and transduce information to the nucleus by signaling pathways. Information from extracellular signals is encoded through dynamics of second messengers, intracellular molecules that activate and respond to the signals- often with direct roles in transcriptional regulation. One of the foci in investigating signaling pathways is how the pathways operate given cell-to-cell variability. Indeed, single cell studies have demonstrated significant variability in levels of second messengers in multiple signaling pathways (Figure 2.1A).

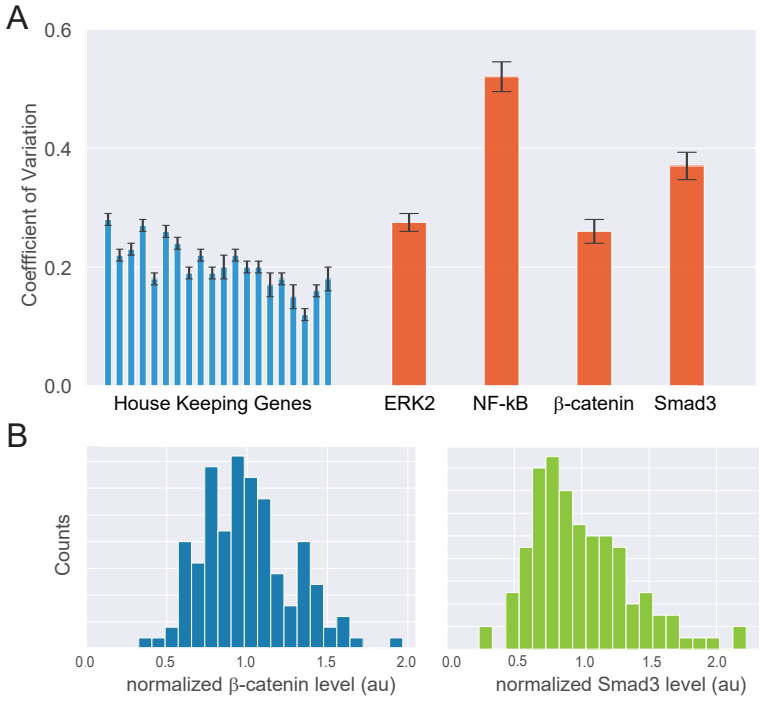


Figure 2.1: Variability of second messengers in signaling pathways
(A) Variability of proteins in mammalian context. Blue: CV of 19 nuclear house-keeping genes in human H1299 cells [1]. Orange: CV of Erk2 [2], NF-kB [3], β-catenin, and Smad3. CV from β-catenin and Smad3 were obtained from single cell data in β-catenin-mCitrine and NG-Smad3 cell lines, respectively.
(B) Distribution of β-catenin (left) and Smad3 (right). The distributions were sampled from snapshots of of cells.

In regards to detection of signal via ligand-receptor activation, emerging findings have demonstrated that fold change of second messengers could be precise despite the variability. [2–5]. However, there is ample evidence that, even without upstream ligand-receptor activation or inhibition, perturbations of second messenger dynamics can bring about significant changes in cells [6–9]. Further, variability of signaling states, reported by transcription output of synthetic promoters, underlies heterogeneity in important cellular responses such as cell-fate decision or response to drug treatment [10, 11].

In this study, in collaboration with Chris Frick, we ask how variability of second messengers in signaling pathways is characterized over time, and how this characteristic may affect cellular processes. Aside from investigations of known oscillatory dynamics [12, 13], how second messengers fluctuate over time has not been well characterized. We use the concept of memory, often represented by autocorrelation functions and defined as 'the rate of decay of statistical dependence of two points with increasing time interval between the two points,' to ground our investigation [14]. Simply put, cells with second messengers that have long memory would have stable levels for longer time, while cells with shorter memory would have quick fluctuations with high amplitude (Figure 1.2, reproduced below).

Memory of protein levels at single cell level has been investigated in two previous studies: Austin et al. (2006) characterized memory of destabilized GFP controlled by synthetic circuits in E.coli [15] and Sigal et al. (2016) did so in 19 nuclear housekeeping proteins in human cells [1]. Both studies use autocorrelation function, which is correlation of a time series with a delayed copy of itself as a function of delay (τ). While both studies use normalized autocorrelation for which data in each time point is normalized to population average, Austin et al. calculates the autocorrelation function from protein levels while Sigal et al. calculates it from ranks of the protein levels. In effect, the autocorrelation function from Sigal et al. informs how quickly a cell's protein level fluctuate relative to those of other cells in a population. By calculating the time at which the autocorrelation function decays to 0.5, one can obtain a standard measure to compare the timescale of memory. This value is referred to as 'noise frequency range' by Austin et al. and 'mixing time' by Sigal et al.

In their study, Sigal et al. find that mixing times of proteins can range from below 1 cell generation to nearly 3 cell generations, demonstrating that proteins fluctuate at different rates. Further, they report high correlation of variability (measured by CV)

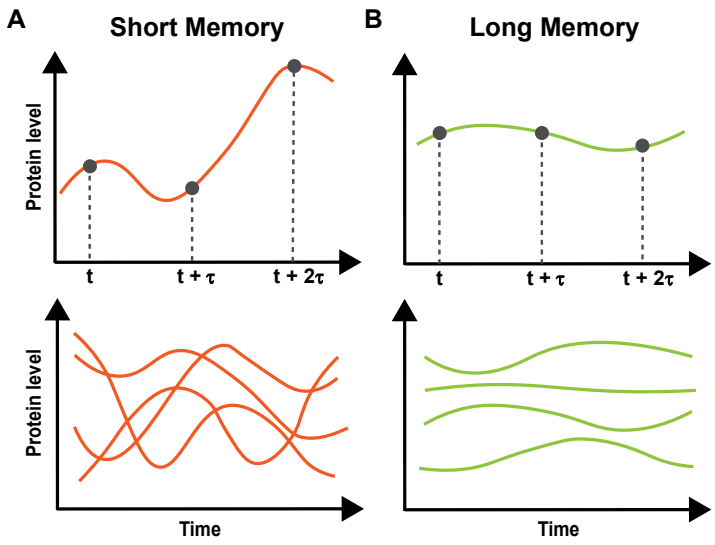
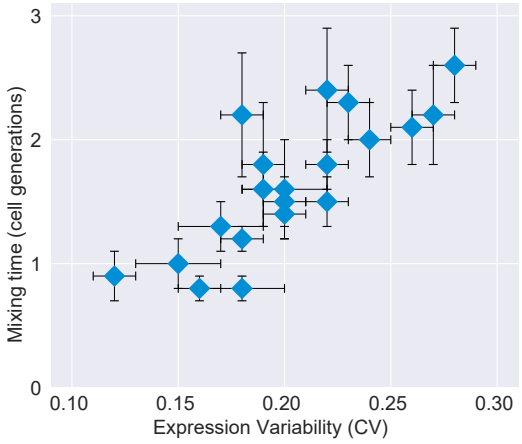


Figure 1.2 (reproduced): Memory of protein levels
Top: An example trace of protein levels over time with short **(A)** or long **(B)** memory. Fluctuations cause correlation between time points to decay quickly **(A)**/ slowly **(B)** with increasing time interval (τ)
Bottom: How protein levels in a population may behave with short **(A)** or long **(B)** memory

and mixing time of a protein- a protein within a population of high CV retained its relative level for a longer time compared to that in a population of low CV (illustrated below).



Mixing time (T_m) vs CV in human housekeeping genes in H1299 cell line, reproduced from original data in [1]

We aim to characterize memory of second messengers in signaling pathways, given the importance of their roles in transduction of signal. For this purpose, we find the

dataset of human proteins from Sigal *et al.* a useful standard for which to compare these dynamic properties. Thus, we calculate memory with mixing time and auto-correlation function used by Sigal *et al.*, shown below:

$$A(\tau) = \frac{\left\langle \left\langle \left(R_i(t) - \left\langle R_i(t) \right\rangle_i\right) \cdot \left(R_i(t+\tau) - \left\langle R_i(t+\tau) \right\rangle_i\right) \right\rangle_i \right\rangle_t}{\left\langle \left\langle \left(R_i(t) - \left\langle R_i(t) \right\rangle_i\right)^2 \right\rangle_i \right\rangle_t}$$

Here, we characterize memory of β-catenin and Smad3, second messengers in two major Metazoan signaling pathways, the Wnt and the Tgf-β pathway. In the canonical Wnt pathway, β-catenin gets constantly degraded by cytoplasmic degradation complex composed of scaffold protein Axin, APC, and kinases CK1 and GSK3-β [16], (Figure 2.2A). During Wnt ligand-receptor activation, the destruction complex is recruited to the receptor and degradation of β-catenin is inhibited, allowing β-catenin to accumulate and translocate to the nucleus.

Stimulation by Tgf-β ligands primarily transduces information through two receptor activated Smads (R-Smads), Smad2 and Smad3, although contexts in which Smad1 and Smad5 could also be activated have been discovered [17], (Figure 2.2B). Upon receptor activation, R-Smads are phosphorylated in the C-terminus and form a complex with common Smad4. The resulting Smad heteromer alters the normal flux of Smads shuttling in and outside of the nucleus with slower export and faster import rates, resulting in nuclear localization [18].

To investigate memory in the β-catenin and Smad3, we used single-cell time-lapse microscopy with fluorescent reporters tagged to endogenous proteins. This system allows real-time quantification of protein levels at a single cell level over extended time periods, and have been utilized previously to capture signaling dynamics in other signaling pathways.[2, 3, 5, 19] However, past reports of signaling dynamics using this technique either did not have homozygous alleles of fluorescent second messenger, or utilized exogenous protein constructs. To comprehensively and faithfully capture the dynamics of second messengers, we aimed to develop clonal cell lines with homozygous fusion of fluorescent reporters to endogenous β-catenin and Smad3 using Crispr-Cas9.

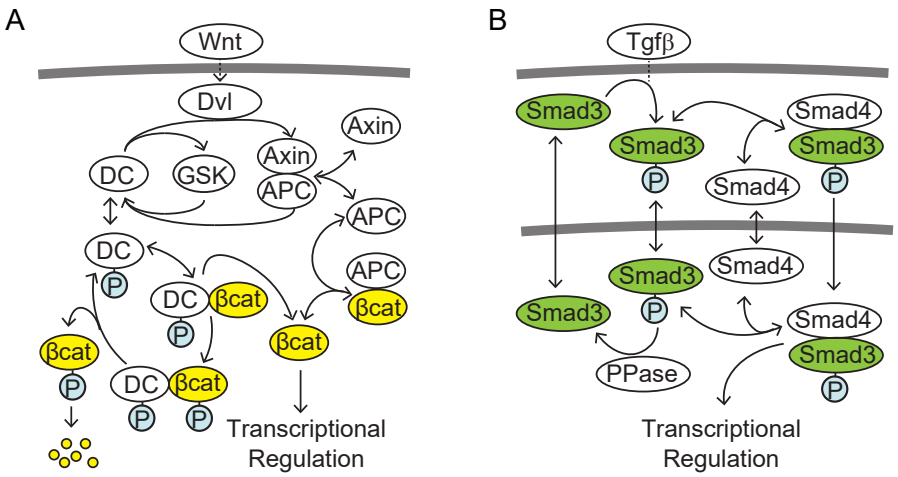


Figure 2.2: Wnt and Tgf-β signaling pathways.
(A) Illustration of Smad3 activation and nucleocytoplasmic shuttling in the Tgf-β pathway
(B) Illustration of rapid degradation of β-catenin and receptor activated inhibition of the degradation in the Wnt pathway.

2.2 Results

Design and Characterization of Endogenous Fluorescent β-catenin and Smad3

Previously, ectopic β-catenin-GFP fused at both N or C-terminus has been utilized in various contexts to investigate Wnt, from sea urchin and *Xenopus* embryos to human cell lines. The fusion protein retained proper localization, phosphorylation, nuclear shuttling, as well as response to Wnt ligand stimulation [20–22]. We have successfully generated homozygous clonal cell lines with mCitrine yellow fluorescent protein tagged at the β-catenin C-terminus (β-catenin-mC) (Figure 2.4A) in mESCs and nMuMGs, and heterozygous clones in C2C12s. Both PCR of β-catenin's genomic locus and Western blot against β-catenin reflected homozygous integration of the fluorescent protein (Figure 2.4B, S2.1B)

In the clonal mESC β-catenin-mC line, functional assays were performed to validate that the fusion protein retained its signaling capacity. RT-PCR of Wnt target genes *Brachyury* and *Axin2* in the β-catenin-mC cell line revealed target activations upon stimulation with recombinant Wnt-3a protein (Figure S2.1C). Phosphorylation of β-catenin is transiently inhibited with ligand-receptor activation[23], and Western blot against phospho-β-catenin in the β-catenin-mC clone reflects strong reduction in the levels at 15-60 minutes after ligand stimulation (Figure 2.4C). Fluorescent imaging

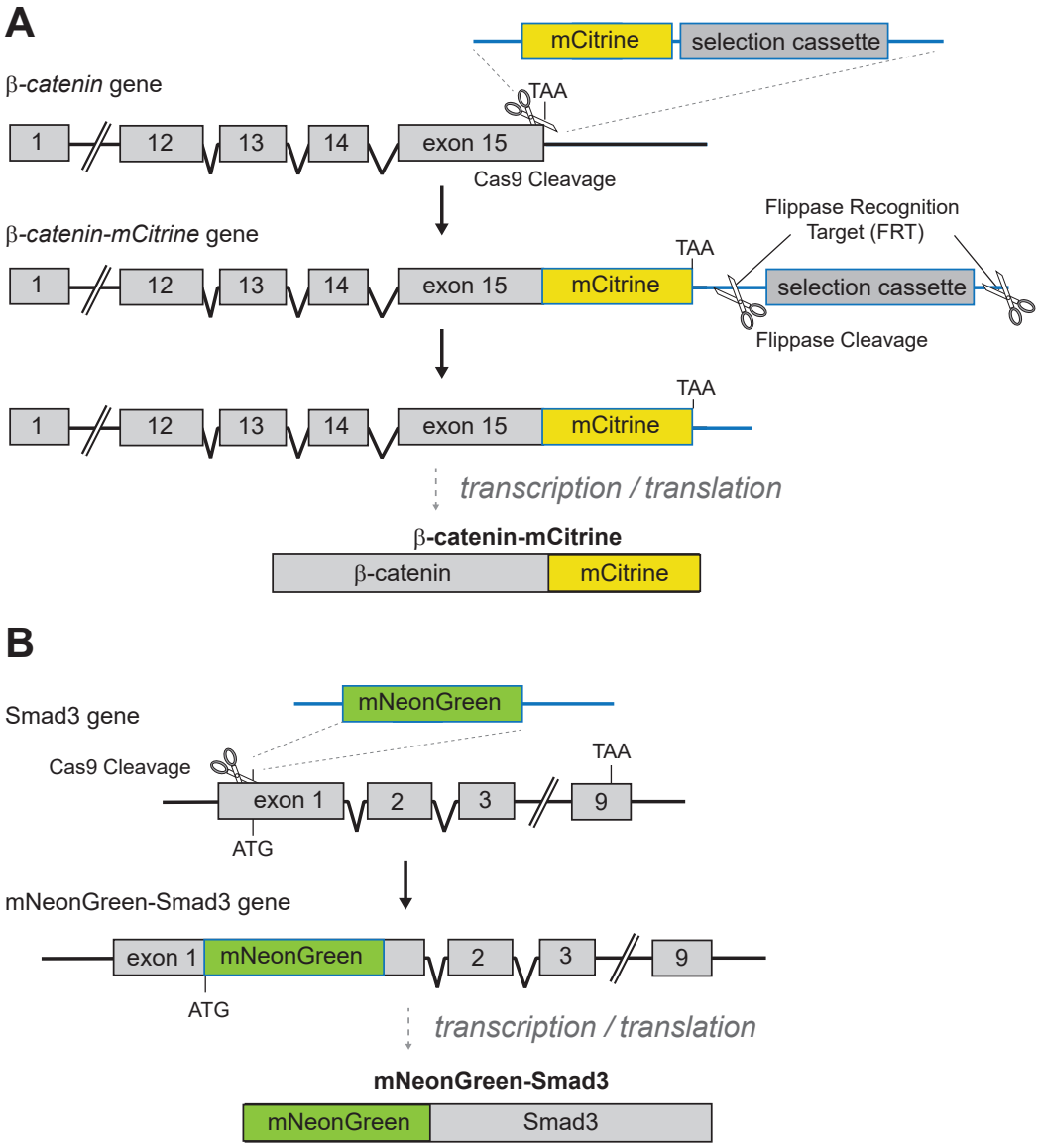


Figure 2.3: Crispr-Cas9 mediated genomic fusion of β-catenin and Smad3 with fluorescent reporters.
(A) Design of β-catenin-mCitrine fusion protein. Antibiotic resistance cassette (geneticin) was used for selection, then excised using Flp-Frt recombination.
(B) Design of mNeonGreen-Smad3 fusion protein. Clonal selection was achieved through sequential sorting by fluorescence.

of the cells reveal β-catenin signal across the whole cell, with especially strong signal at the membrane in cell-cell junctions (Figure 2.4D). This is consistent with known β-catenin interactions with E-Cadherins, which mediate formation of cell-cell junctions [24]. Finally, time-lapse imaging reveals that stimulation of Wnt3a

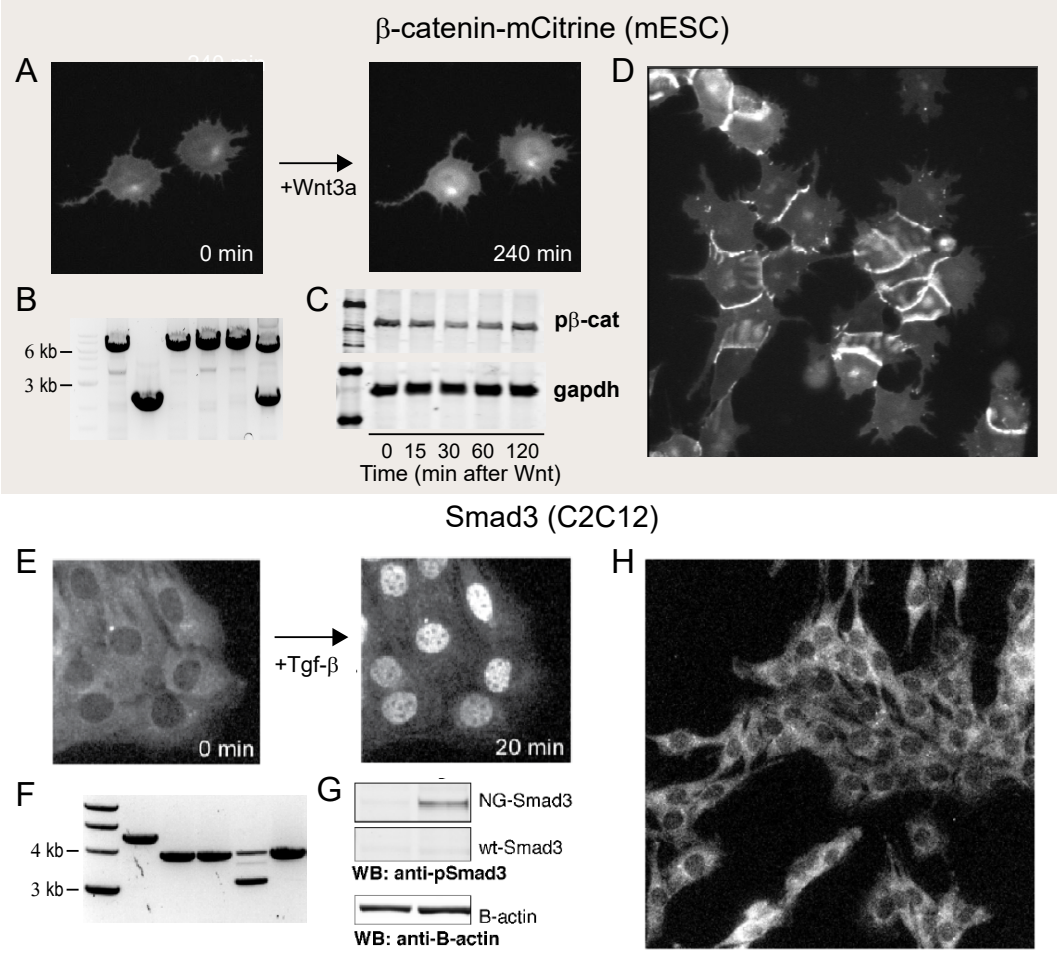


Figure 2.4: Characterization of fluorescence-fusion β-catenin and Smad3
(A) Live imaging the response of β-catenin-mCitrine cells to 150 ng/mL recombinant Wnt3A through mCitrine fluorescence.
(B) PCR of genomic β-catenin locus to validate homozygous integration of mCitrine. Expected band for successful integration is 6.4 kb, wildtype band is 2.3kb. Lanes 1,3,4,5 show homozygous integration, Lane 2 shows no integration, and lane 6 shows heterozygous integration.
(C) Western blot against phospho-β-catenin shows transient inhibition of phosphorylation after addition of 150 ng/mL Wnt3a.
(D) Representative fluorescence image of mESC β-catenin-mCitrine clonal line.
(E) Live imaging the response of NG-Smad3 to Tgf-β addition.
(F) Expected band for successful integration is 3.8 kb. Lanes 2,3,5 show homozygous integration, 4 shows heterozygous integration, and 1 shows incorrect integration.
(G) Western blot against phospho-Smad3 shows phosphorylation of Smad3 in response to Tgf-β addition.
(H) Representative fluorescence image of C2C12 NG-Smad3 clonal line

induce increase in overall level and nuclear translocation of β-catenin (Figure 2.4A, S2.2A)

The two principal second messengers in the Tgf-β pathway, Smad2 and Smad3, are structurally similar but regulate distinct genes. Moreover, while Smad3 binds directly to DNA, the dominant isoform of Smad2 does not. In this study, we chose to fuse fluorescent reporter at the N-terminus of endogenous Smad3. Smad3 tagged at the N-terminus has been shown to retain its signaling functions - phosphorylation at C-terminus, nuclear shuttling, and transcriptional regulation, and previously utilized to investigate dynamics with ectopic expression [5, 25]. Using Crispr-Cas9 directed genome modification, we successfully generated homozygous clonal cell lines of Smad3 fused mNeonGreen fluorescent protein (NG-Smad3) at the N-terminus (Figure 2.3B) in C2C12 muscle progenitor cells (Figure 2.4B) and nMuMG epithelial cell lines. We also generated clones of heterozygous NG-Smad3 in mouse embryonic stem cells (mESCs), but were not able to get clones with homozygous integration.

NG-Smad3 in clonal C2C12 lines were demonstrated to retain capacity for signal transduction. Blotting for phospho-Smad3 reveals that the fusion protein is readily phosphorylated with Tgf-β stimulation (Figure 2.4G). In time-lapse imaging, the cells show response to Tgf-β ligand stimulation as cytoplasmic Smad3 is rapidly localized to nucleus (Figure 2.4E). Finally, dynamics of Snail activation, a known target gene of Tgf-β [26], reported by endogenous Snail-p2a-mKate integrated into the same cell lines, demonstrate activation with Tgf-β and correlation with Smad3 dynamics (Figure S2.1A).

Endogenous β-catenin and Smad3 fluctuate quickly

For initial characterization of variability and memory, we used NG-Smad3 clone from C2C12 and β-catenin-mC clone from mESCs. We captured the dynamics of the fluorescence-tagged second messengers by time-lapse microscopy, under consistent temperature, humidity, and CO^2 levels. The cells were kept in normal growing conditions without any ligand stimulation or perturbations. The cells were imaged for 30 hours at 10 minutes per frame, covering over two cell generations. From the images, cells were segmented and tracked by nuclear signal of either mCerulean or mKate tagged with nuclear localization signals. Levels of fluorescence from both second messengers and nuclear markers were quantified, and tracked cells were digitally synchronized by cell cycle (Figure S2.2).

In the control conditions, CV of both Smad3 and β-catenin remained consistent

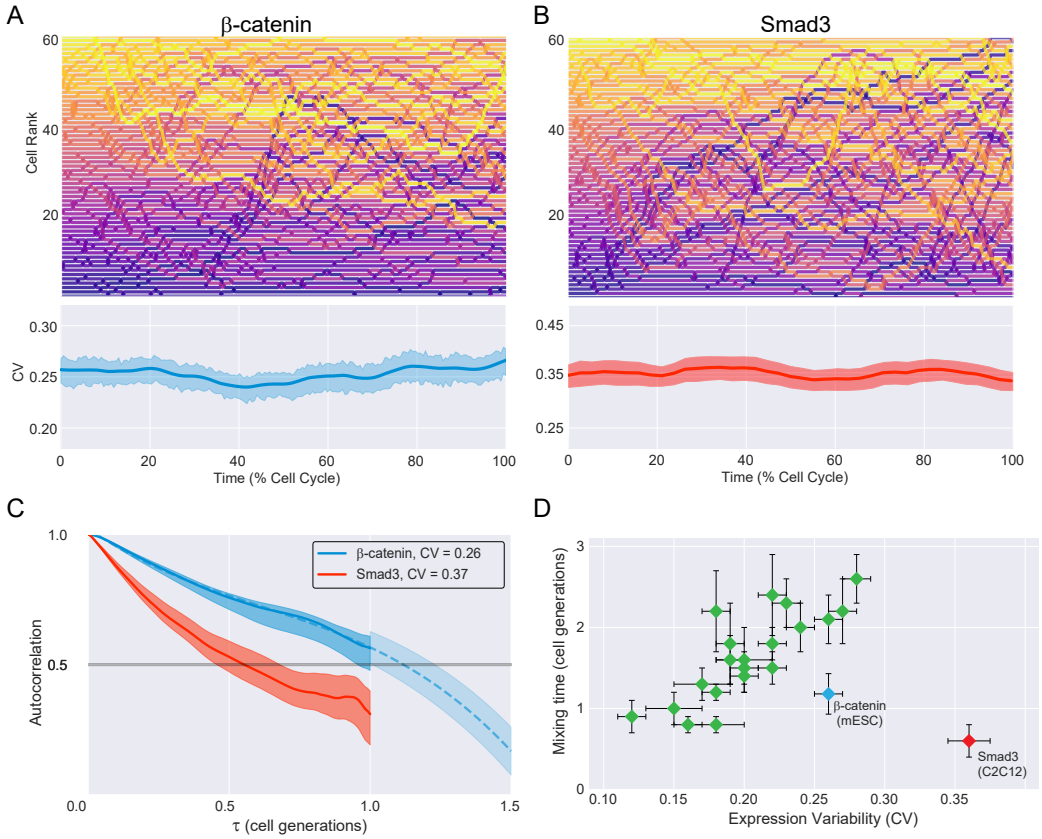


Figure 2.5: CV and memory of β-catenin and Smad3
(A,B) (Top) Ranks β-catenin (A) / Smad3 (B) levels in 60 random cells through one cell cycle. Color for each cell was determined at timepoint 0 by the fluorescence rank. (Bottom) Coefficient of Variation for the population of cells (β-catenin: N=194, Smad3: N= 114). ±1 bootstrapped SE shown.
(C) Mean-normalized rank autocorrelation functions of β-catenin and Smad3, ±1 bootstrapped SE shown. For β-catenin, polynomial fitting was used to extend the function (dashed line) to estimate mixing time, where A(τ) crosses 0.5
(D) Mixing time (τ_m) vs CV in human housekeeping genes in H1299 (green [1], β-catenin in mESCs (blue), and Smad3 in C2C12s (red).

throughout a cell cycle (Figure 2.5A,B). Compared to the range of CVs of housekeeping proteins measured by Sigal *et al.* (0.13 to 0.28) [1], the second messengers exhibited CV values on the high end (0.28 for β-catenin, 0.37 for Smad3) of the distribution (Figure 2.5A,B, 2.1). To calculate the normalized autocorrelation of the rank of protein levels, protein ranks were calculated and tracked for each cell (Figure 2.5A,B). The rank plots reveal that Smad3 fluctuates much faster than β-catenin, reflected by the faster decay in normalized rank autocorrelation functions (Figure 2.5C). Mixing time of β-catenin was calculated to be ~1.3 cell cycle, and that of

Smad3 was much faster at ~0.6 cell cycle (Figure 2.5D).

It is known that stem cells secrete Wnt proteins in human ES culture that introduce downstream variability [10]. To check whether endogenous Wnt secretions could affect variability and memory of β-catenin in mESCs, we treated the cell line with IWP2, Porcupine inhibitor that blocks secretion of Wnt proteins [27]. The variability and autocorrelation calculated from β-catenin dynamics in the treated cells did not show significant change from the control conditions (Figure S2.3)

Given previous findings correlating CV with mixing time, these results are surprising: in comparison to the mixing times of human housekeeping genes, both β-catenin and Smad3 had much shorter memories relative to their CV.

Wnt stimulation reduces variability and memory of β-catenin

As mentioned, Wnt stimulation leads to inhibition of degradation complex, leading to stabilization and nuclear translocation of β-catenin. In Gillespie simulations [28] and mass action models [1], increased protein degradation rates are predicted to shorten the timescales of memory. As Wnt signaling primarily inhibits degradation rate of β-catenin, we predicted that signaling activation could lengthen the memory of the second messenger.

We tracked and quantified the β-catenin levels of cells treated with 150 ng/mL of recombinant Wnt3a. The ligand stimulation led to 1.8 fold elevation of nuclear β-catenin concentration over the first 3 hours, and reduction of CV over the same duration (Figure 2.6A). Cell traces of ranked β-catenin revealed a surprising trend: fluctuations of the second messengers within the population sped up significantly (Figure 2.6B). The observation was confirmed in comparison of autocorrelations functions between traces without Wnt stimulation and those with stimulation, with Wnt-stimulated population exhibiting must faster decay of autocorrelation (Figure 2.6E). These results suggested that although degradation is the primary effector of β-catenin dynamics in the Wnt signaling, there may be other, more significant factors affecting memory of the second molecule.

Tgf-β stimulation raises variability and memory of Smad3

Next, we wondered whether reduction of memory in second molecules through receptor activation could be a general, preferred characteristic in signaling pathways despite varying transduction mechanisms. Indeed, the two pathways share certain qualities of information transfer such as fold change response of second messengers

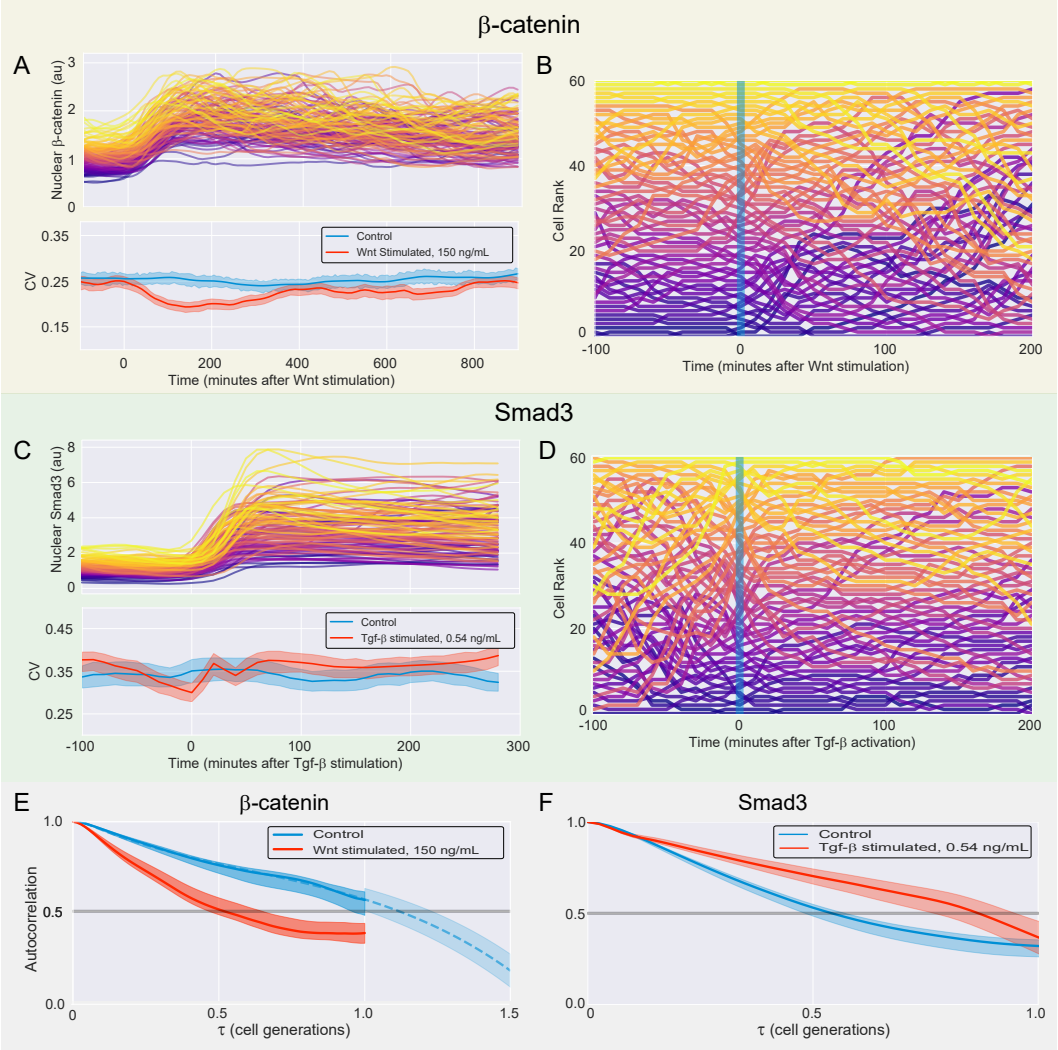


Figure 2.6: Signaling affects variability and memory of second molecules
(A,C) Top: Traces of normalized β-catenin **(A)** or Smad3 **(C)** levels following 150 ng/mL Wnt3a **(A)** or 0.54 ng/mL Tgf-β **(C)** stimulation. Bottom: CV of the traces over the same time period (red), contrasted with cells without ligand stimulation (blue). ±1 bootstrapped SE shown.
(B,D) Cell traces of β-catenin **(B)** or Smad3 **(D)** ranks before and after ligand addition. Color for each cell was determined based on the rank at time of ligand addition (blue bar).
(E,F) Normalized rank autocorrelation functions in β-catenin **(E)** and Smad3 **(F)**. Autocorrelations function in ligand-stimulated cells (Red) are plotted with cells in control conditions (Blue). Due to the timing of ligand addition, cells are not digitally synchronized in this data set. ±1 bootstrapped SE shown.

[4, 5] and linear input-output transmission [29]. Thus, we looked at how memory and variability of Smad3 in the Tgf-β pathway behaves in response to signal activation. Smad3 has 3 fold elevation in nuclear concentration following addition of 0.54

ng/mL Tgf-β ligand, coinciding with elevation of CV (Figure 2.6C). Plotting ranked cell traces reveal that, contrasting with findings from the Wnt pathway, Smad3 fluctuations stabilize and slow down with signal activation (Figure 2.6D)

It is interesting to note that signal activation affects memory in opposite manner in the Wnt and Tgf-β pathways. This disparity, as well as the difference in timescales of memory, could contribute to diversity in temporal complexity of information transduction in signaling. The rapid change in profiles of memory in response to signaling also suggest that variability and memory of the second messengers could be sensitive to cellular contexts.

Cellular context modulates CV and memory of β-catenin and Smad3

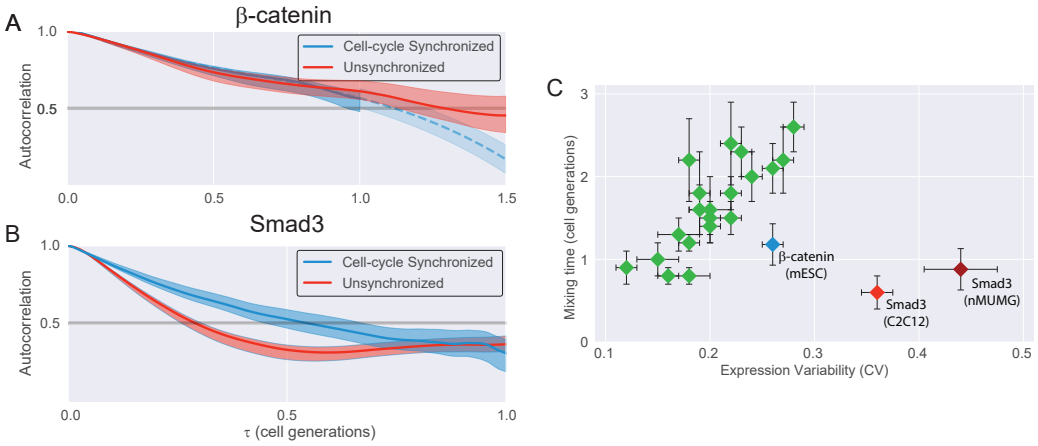


Figure 2.7: Cellular context modulates variability and memory of β-catenin and Smad3
(A,B) Normalized rank autocorrelation functions for β-catenin **(A)** or Smad3 **(B)** levels, comparing traces that were digitally synchronized for cell cycle (Blue) vs unsynchronized traces (Red). ± 1 bootstrapped SE shown.
(C) Mixing time (τ_m) vs CV in human housekeeping genes, updated showing Smad3 in two different cell lines, C2C12 (red) and nMUMG (dark red).

Previous analyses of variability and memory were performed with cell-cycle synchronized traces of the second messengers. Cells were digitally synchronized for two reasons: it may reduce confounding variables to use cells that behave collectively as one undergoing normal cell cycle, and synchronization allows a standard timeframe of comparison between varying pathways and cell lines. However, cell lines in culture often exhibit asynchronous divisions, suggesting that tissues are not composed of synchronized cells [30]. Synchronization is most frequently observed during rapid division of embryo development before midblastula transition (MBT),

in which cells skip G1/G2 phases during the synchronized divisions [31]. In many contexts in which cells are not synchronized, the variability in cell cycle phases could affect second messenger dynamics, as both Wnt and Tgf pathways have implicated roles in cell cycle control [32, 33].

We therefore analyzed second messenger dynamics from population of cells that were not signal activated, and compared the memory of unsynchronized cells vs digitally synchronized cells. β-catenin showed no difference between synchronized and unsynchronized cells, suggesting that cell cycle does not affect the dynamics of the second messenger (Figure 2.6A). Smad3, however, showed a significant reduction in memory, reflected by faster decay of the rank-autocorrelation function in unsynchronized cells compared to that of synchronized cells (Figure 2.7B). We find that such discrepancy arises from Smad3 concentration that continuously decreases over a cell cycle, while β-catenin concentrations remain steady (Figure S2.4).

Signaling pathways are used in varying contexts and purposes, and we wondered whether variability and memory of second messengers are contextually dependent given the apparent tuneability. Activation of Tgf-β pathway have stimulatory effects for differentiation and myogenesis in C2C12s [34, 35], while the pathway mediates epithelial-to-mesenchymal transdifferentiation in nMUMG cells [36, 37]. Comparing variability and memory of Smad3 in the two varying cell lines, we observed significant differences: Smad3 levels in nMUMG presented larger variability and longer memory (Figure 2.7C) The results suggest that memory of second messengers is indeed context specific, but still very short respective to the CV compared to housekeeping genes.

2.3 Discussion

We find, through single cell time-lapse experiments recording endogenous Smad3 and β-catenin levels, that second messengers have relatively short memory. Based on the rank-autocorrelation functions, it is likely that cells with relatively high concentrations of Smad3 or β-catenin in a population will not remain so after a cell cycle. This is quite surprising given the role of second messengers as transcription regulators, mediating expression of multiple target genes in a given context. For this reason, one may expect second messengers to have long timescales of memory for stable encoding and minimizing noisy transduction. Perhaps the short memory accommodates constantly changing environment and context during development, where these signaling pathways are reused frequently, so that second messenger

distributions are quickly reestablished after signaling events.

In any case, questions arise on whether target genes either filter or decode the fast, large amplitude fluctuations of second messengers. If indeed the latter is true for some target genes, it would be important to know the timescales of memory in applications such as drug treatment. Further, it would be useful to investigate decoding mechanisms that could differentiate signaling events from natural fluctuations, and classify genes that are sensitive to the short memory from those that are not.

The modulation of memory through signaling activation and cellular contexts demonstrate that memory of second messengers are rather malleable. The findings show that it would be possible to modulate memory in defined methods for further experimentation. As degradation rate and negative feedback have been shown to affect memory of proteins, tuning these properties could allow systematic modulation of memory [1, 15]. Surprisingly, inhibiting of degradation of β-catenin through Wnt ligand stimulation elicited an opposite effect on memory from expected, leading to faster fluctuations and shorter memory. To investigate further, small molecule activators of Axin were added to mESCs with expectations to speed up degradation of β-catenin, but the drugs did not bring about significant changes to the level of second messengers. However, there are still viable strategies to modulate degradation, including exogenous expression of degradation machinery. Further, as both Wnt and Tgf-β pathways have known negative feedback components [38–41], modulation of the negative feedback through techniques such as miRNA/siRNA could be further avenues in modulating memory in second messengers.

2.4 Materials and Methods

Cell Culture

mouse Embryonic Stem Cells (mESCs): E14 mESC (E14Tg2a.4) were obtained from Mutant Mouse Regional Resource Centers (015890-USCD). The mESC was cultured at 37C and 5% (vol/vol) CO2 on 0.1% gelatin coated plates with Glasgow's MEM supplemented with 1000 U/mL LIF (Millipore, ESG1106), 10% ES-certified FBS (Thermo Fisher 16141061), non-essential amino acids, and L-glutamine. To confirm the identity of the mESC, we performed differentiation assay, confirmed the expected changes in cell morphology, and verified expression of pluripotency and differentiation markers using qRT-PCR, including Nanog, Oct4, Brachyury, Flk1, Myf5, Gata4, Sox17, Pax6, Otx2. We confirmed the purity of the culture by performing mycoplasma testing.

C2C12 cells (American Type Culture Collection, CRL-1772) were cultured at 37C and 5% (vol/vol) CO2 in DMEM (ThermoFisher Scientific; 11995) supplemented with 10% (vol/vol) FBS (Invitrogen; A13622DJ), 100 U/mL penicillin, 100 ug/mL streptomycin, 0.25 ug/mL amphotericin, and 2 mM L-glutamine (Invitrogen)

nMUMG cells were cultured at 37C and 5% (vol/vol) CO2 in DMEM (ThermoFisher Scientific; 11995) supplemented with 10% (vol/vol) FBS (Invitrogen; A13622DJ), 1% non-essential amino acids, 100 U/mL penicillin, 100 ug/mL streptomycin, 0.25 ug/mL amphotericin, and 2 mM L-glutamine (Invitrogen)

Crispr-cas9 mediated genome editing

mNeonGreen-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).

β-catenin-mCitrine We designed HDR template plasmid to insert mCitrine with PGK drive puromycin casette and TK reverse selection casette at the immediate

C-terminus of β-catenin, removing the stop codon. The plasmid had 1000 basepairs of homology each on regions upstream and downstream of the stop codon. For delivery of gRNA, we obtained PX330 generously donated to Addgene by the Feng Zhang lab [42]. Into the PX330 plasmid, we inserted gRNA sequence targeting the T promoter and removed the Cas9 coding cassette. Cells were transfected with HDR template plasmid, modified PX330 plasmid, and Cas9 protein (IDT: Alt-R S.P. Cas9) using Lipofectamine LTX (Life Technologies). After clonal selection using puromycin, antibiotic cassette was removed from the genome with transfection of Flippase expression plasmid and selected with TK reverse-selection by 4 uM Ganciclovir solution (Life Technologies).

Single cell clonal selection

mNeonGreen-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).

β-catenin-mCitrine

Transfected cells from 24-well were plated in 10cm well and placed in puromycin selection (1.5 ug/mL). After antibiotic selection, single cells were plated in 96 well plates via limiting dilution. The clonal population were screend for characteristic β-catenin mCitrine fluorescence, with strong signal at cell-cell junctions. Genomic PCR was performed from clonal samples to check for integration as well as sequencing, with KOD Xtreme hot start DNA polymerase (Millipore-Sigma).

Western Blots

mESC cells at 70% confluency were scraped in PBS, pelleted, snapfrozen, and then thawed in NP-40 lysis buffer containing Halt inhibitor cocktail. Samples were spun down and supernatants were mixed with 1:1 vol Concanavalin A for removal of

cadherins. The solution was spun down again, and supernatants were transferred to Laemmli sample buffer and boiled. The samples were then run onto a Bolt™ 4-12% Bis-Tris Plus Gel (Thermofisher, NW04120BOX). Proteins were transferred onto nitrocellulose membranes, blocked for one hour at RT with blocking buffer (Odyssey® Blocking Buffer (TBS) (927-50000) and stained for 1 hr at RT with primary antibody diluted in blocking buffer. The membranes were then stained with fluorescent IR secondary antibodies diluted in blocking buffer for one hour at RT. The fluorescent signal was then imaged using the LiCOR Odyssey Imager and quantified using Odyssey Application software version 3.0. The background-subtracted intensity of the protein bands were normalized to GAPDH loading control.

RT-PCR.

Total RNA from mESC was isolated with RNeasy extraction kit (Qiagen 74104). cDNA was synthesized using Quantitect RT (Qiagen 205311). Real-time RT-PCR was performed using Quantitect SYBR-Green kit (Qiagen 204143) using StepOnePlus RT-PCR System (Applied Biosystems).

Live cell imaging

Cells were grown on 96-well glass-bottomed plates (Griener Bio-One; 655891) overnight before imaging in 200 uL of FluoroBrite DMEM (Gibco-Life Technologies; A18967) containing respective cell culture media. For mESC, the wells were pre-coated with laminin (Invitrogen; 23017-015) diluted into 10 ug/mL in dPBS solution overnight. Cells were imaged using a Zeiss Axio Observer.Z1 inverted fluorescence microscope under incubation [37C and 5% (vol/vol) CO2, with humidification] on a motorized stage. In each experiment, 25–80 positions were imaged and focus was maintained using Zeiss Definite Focus. Images were acquired at 10-min intervals with a 40×, 0.8-N.A. Plan Apo objective and Evolve 512 EM-CCD camera (Photometrics).

Imaging Analysis

Background subtraction and bleach correction: ime-lapse movies were quantified after flat-field correction, bleaching correction, and background subtraction. We followed the standard protocol described by Waters [43]. 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.

Segmentation, tracking, and quantification of signal in cells We quantified both median or total fluorescence intensity of NG-Smad3 or β-catenin-mCitrine fluorescence in the nuclei. The nuclei of cells were first segmented based on the fluorescence of the constitutively expressed mCerulean3-3NLS (3×NLS) or pSnail-mKate-3NLS. For images with mESC, Ilastik was used for nuclear segmentation [44]. Next, segmented nuclei were tracked across all time frames. Finally, the fluorescence data from the segmented nuclei were extracted. We performed segmentation, tracking, and fluorescence quantitation steps using the Lineage Tracker ImageJ (NIH) plug-in (36) and custom MATLAB (MathWorks) scripts (available upon request).

Digital synchronization Cells undergo general morphological change during cell division, balling up with nuclear disintegrated before dividing into two daughter cells. The division can be captured in time series fluorescence data by identifying when total level of fluorescent protein is drastically reduced (theoretically close to halved). Further, median fluorescence of second messengers also spike in levels when a cell balls up during division. Once division points are identified using these signatures, all time series between two cell divisions are collected for synchronization. Because cells have natural variations in duration of cell cycle, linear interpolation is used to fit all traces to the same timescale. All code for digital synchronization were performed in Python 3.6.

Bootstrapping Bootstrapping was used according to [45] in several measures in the study, e.g. CV, autocorrelation, to identify confidence intervals. In constructing the bootstrapped intervals, both N values and number of reps were tuned such that increasing each variable did not significantly alter the predicted intervals.

2.5 Supplementary Figures

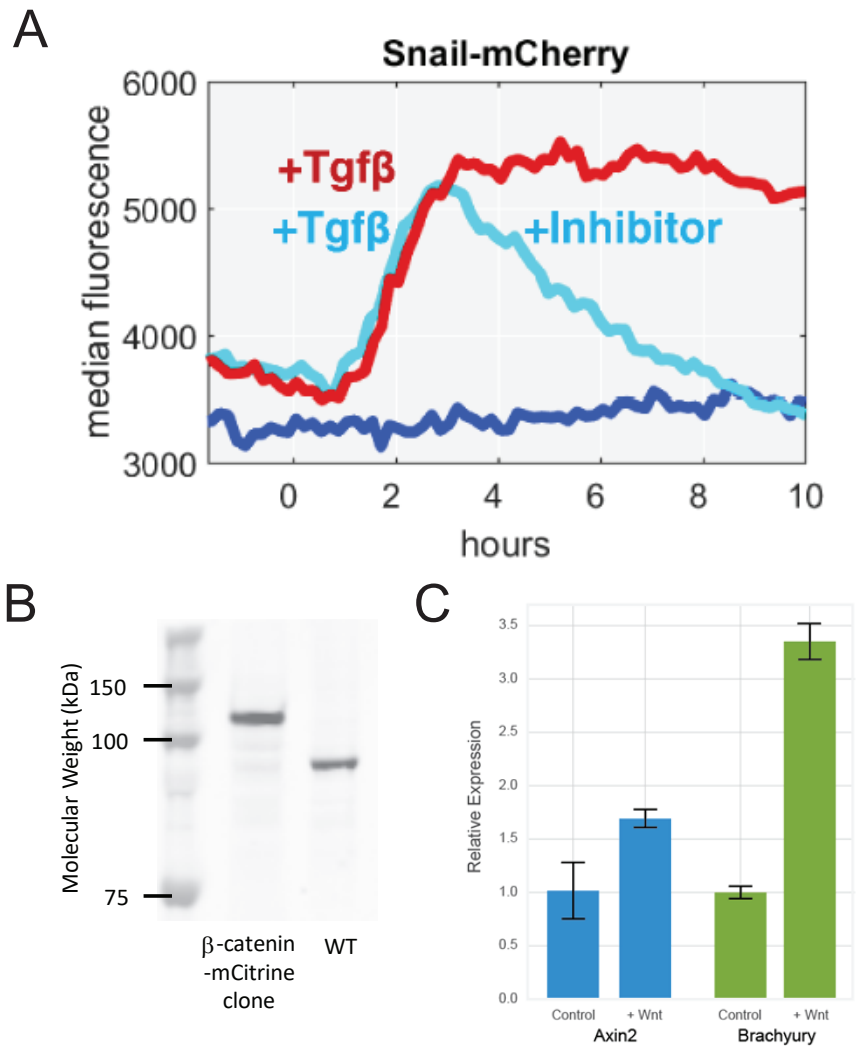


Figure S2.1: Fluorescence fusion second proteins activate target genes.
(A) Response of endogenous Snail-p2a-mCherry reporter to Tgf-β addition in NG-Smad3 clonal cell line. Blue: control, red: 2.4 ng/mL Tgf-β added, light blue: addition of SB431-542 inhibitor of Smad2/3 phosphorylation terminates induction of mCherry.
(B) Western blot against β-catenin reveals homozygous integration of mCitrine (left lane). Clonal line with Wildtype β-catenin (right lane) for comparison.
(C) RT-PCR demonstrating activation of Wnt target genes Axin2 and Brachyury with 150 ng/mL Wnt3a addition in the β-catenin-mCitrine clonal mESC line.

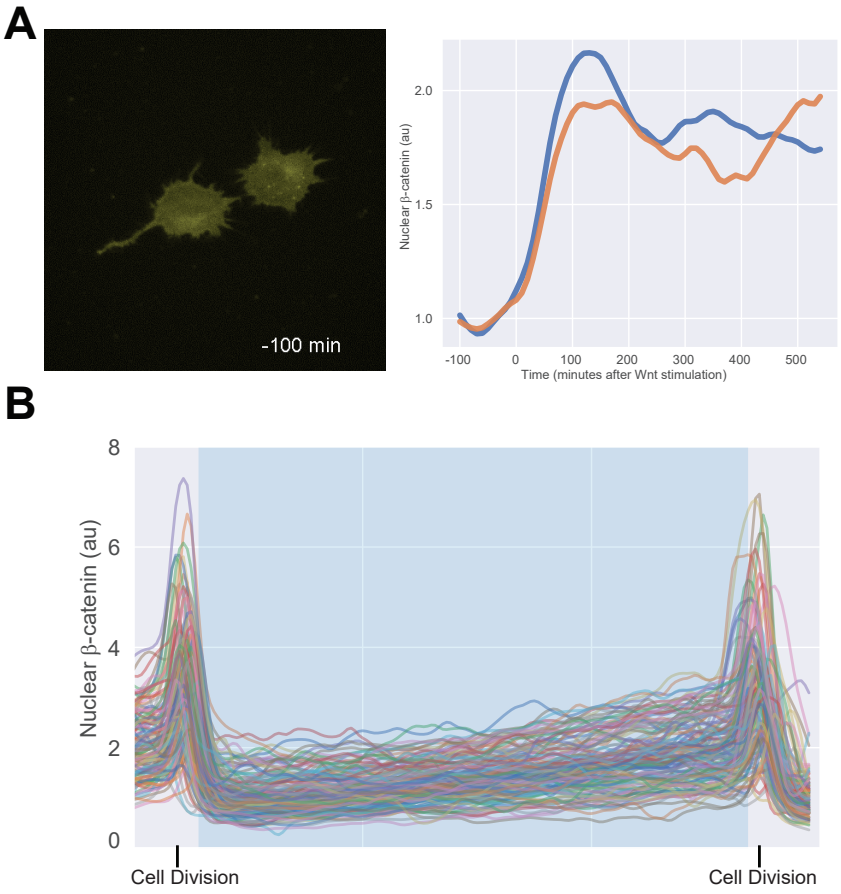


Figure S2.2: Fluorescence tracking and digital synchronization
(A) MOVIE: Example of quantification of β-catenin dynamics in sister cells. Cells were stimulated with 150 ng/mL Wnt3a at t=0 (time of acquisition listed in bottom right corner).
(B) Digital synchronization of β-catenin traces. See methods for details.

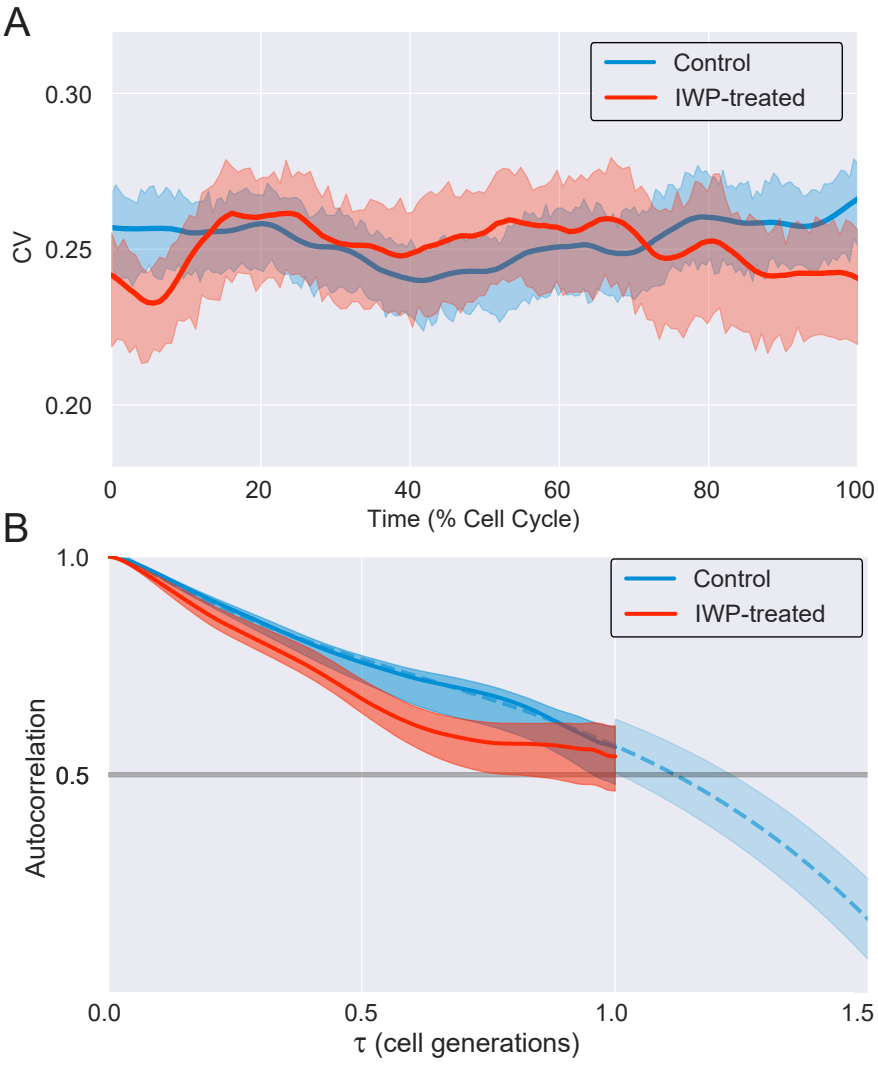


Figure S2.3: Variability and memory of β-catenin with blocking of Wnt secretion
(A) CV of β-catenin throughout a cell cycle, cells were either in control growing conditions or treated with 2.5 uM IWP2 for two days prior to imaging.±1 bootstrapped SE shown.
(B) Normalized rank autocorrelation functions of cells in control growing conditions or IWP2. ±1 bootstrapped SE shown.

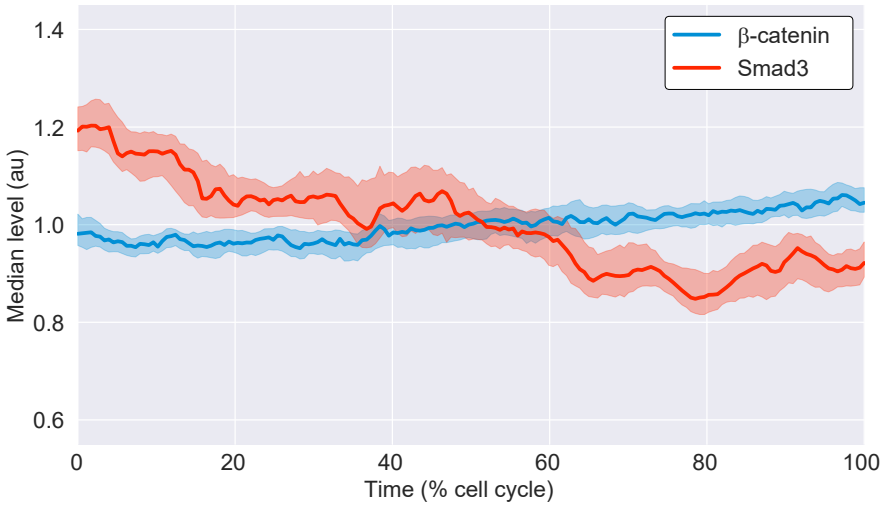


Figure S2.4: β-catenin and Smad3 concentration over a cell cycle]
Median nuclear levels of β-catenin and Smad3 over a duration of cell cycle. ± 1 bootstrapped SE shown.

2.6 Work Contributions

Chris Frick generated cell lines involving NG-Smad3, created code for image quantification (cell tracking), and participated in design and execution of experiments. K.K generated cell lines involving β-catenin-mCitrine, participated in design and execution of experiments, and data analysis.

2.7 Bibliography

1. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643 (2006).
2. Cohen-Saidon, C., Cohen, A. A., Sigal, A., Liron, Y. & Alon, U. Dynamics and variability of ERK2 response to EGF in individual living cells. *Molecular cell* **36,** 885–893 (2009).
3. Lee, R. E., Walker, S. R., Savery, K., Frank, D. A. & Gaudet, S. Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Molecular cell* **53,** 867–879 (2014).
4. Goentoro, L. & Kirschner, M. W. Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling. *Molecular cell* **36,** 872–884 (2009).
5. Frick, C. L., Yarka, C., Nunns, H. & Goentoro, L. Sensing relative signal in the Tgf-β/Smad pathway. *Proceedings of the National Academy of Sciences* **114,** E2975–E2982 (2017).
6. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature medicine* **10,** 55 (2004).
7. Marshall, C. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80,** 179–185 (1995).
8. Yap, J. L., Worlikar, S., MacKerell Jr, A. D., Shapiro, P. & Fletcher, S. Small-Molecule Inhibitors of the ERK Signaling Pathway: Towards Novel Anticancer Therapeutics. *ChemMedChem* **6,** 38–48 (2011).
9. Zhang, Y., Chang, C., Gehling, D. J., Hemmati-Brivanlou, A. & Derynck, R. Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase. *Proceedings of the National Academy of Sciences* **98,** 974–979 (2001).
10. Blauwkamp, T. A., Nigam, S., Ardehali, R., Weissman, I. L. & Nusse, R. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nature communications* **3,** 1070 (2012).
11. Oshimori, N., Oristian, D. & Fuchs, E. TGF-β promotes heterogeneity and drug resistance in squamous cell carcinoma. *Cell* **160,** 963–976 (2015).
12. Kellogg, R. A. & Tay, S. Noise facilitates transcriptional control under dynamic inputs. *Cell* **160,** 381–392 (2015).
13. Heltberg, M., Kellogg, R. A., Krishna, S., Tay, S. & Jensen, M. H. Noise induces hopping between NF-κB entrainment modes. *Cell systems* **3,** 532–539 (2016).

14. Beran, J. *Statistics for long-memory processes* (Routledge, 2017).

15. Austin, D. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439,** 608 (2006).

16. Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149,** 1192–1205 (2012).

17. Moustakas, A. & Heldin, C.-H. The regulation of TGFβ signal transduction. *Development* **136,** 3699–3714 (2009).

18. Schmierer, B., Tournier, A. L., Bates, P. A. & Hill, C. S. Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proceedings of the National Academy of Sciences* **105,** 6608–6613 (2008).

19. Warmflash, A. *et al.* Dynamics of TGF-β signaling reveal adaptive and pulsatile behaviors reflected in the nuclear localization of transcription factor Smad4. *Proceedings of the National Academy of Sciences* **109,** E1947–E1956 (2012).

20. Miller, J. R. & Moon, R. T. Analysis of the signaling activities of localization mutants of β-catenin during axis specification in Xenopus. *The Journal of cell biology* **139,** 229–243 (1997).

21. Weitzel, H. E. *et al.* Differential stability of β-catenin along the animal-vegetal axis of the sea urchin embryo mediated by dishevelled. *Development* **131,** 2947–2956 (2004).

22. Kafri, P. *et al.* Quantifying β-catenin subcellular dynamics and cyclin D1 mRNA transcription during Wnt signaling in single living cells. *Elife* **5,** e16748 (2016).

23. Hernández, A. R., Klein, A. M. & Kirschner, M. W. Kinetic responses of β-catenin specify the sites of Wnt control. *Science* **338,** 1337–1340 (2012).

24. Orsulic, S., Huber, O., Aberle, H., Arnold, S. & Kemler, R. E-cadherin binding prevents beta-catenin nuclear localization and beta-catenin/LEF-1-mediated transactivation. *Journal of cell science* **112,** 1237–1245 (1999).

25. Nicolás, F. J., De Bosscher, K., Schmierer, B. & Hill, C. S. Analysis of Smad nucleocytoplasmic shuttling in living cells. *Journal of cell science* **117,** 4113–4125 (2004).

26. Brandl, M. *et al.* IKKα controls canonical TGFβ–SMAD signaling to regulate genes expressing SNAIL and SLUG during EMT in Panc1 cells. *J Cell Sci* **123,** 4231–4239 (2010).

27. Chen, B. *et al.* Small molecule–mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nature chemical biology* **5,** 100 (2009).

28. Simpson, M. L., Cox, C. D. & Sayler, G. S. Frequency domain analysis of noise in autoregulated gene circuits. *Proceedings of the National Academy of Sciences* **100,** 4551–4556 (2003).

29. Nunns, H. & Goentoro, L. Signaling pathways as linear transmitters. *eLife* **7,** e33617 (2018).

30. Davis, P. K., Ho, A. & Dowdy, S. F. Biological methods for cell-cycle synchronization of mammalian cells. *Biotechniques* **30,** 1322–1331 (2001).

31. Kraeussling, M., Wagner, T. U. & Schartl, M. Highly asynchronous and asymmetric cleavage divisions accompany early transcriptional activity in pre-blastula medaka embryos. *PloS one* **6,** e21741 (2011).

32. Massagué, J., Blain, S. W. & Lo, R. S. TGFβ signaling in growth control, cancer, and heritable disorders. *Cell* **103,** 295–309 (2000).

33. Davidson, G. & Niehrs, C. Emerging links between CDK cell cycle regulators and Wnt signaling. *Trends in cell biology* **20,** 453–460 (2010).

34. Filvaroff, E. H., Ebner, R. & Derynck, R. Inhibition of myogenic differentiation in myoblasts expressing a truncated type II TGF-beta receptor. *Development* **120,** 1085–1095 (1994).

35. Budasz-Rwiderska, M., Jank, M. & Motyl, T. Transforming growth factor-beta1 upregulates myostatin expression in mouse C2C12 myoblasts. *Journal of physiology and pharmacology: an official journal of the Polish Physiological Society* **56,** 195–214 (2005).

36. Miettinen, P. J., Ebner, R., Lopez, A. R. & Derynck, R. TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *The Journal of cell biology* **127,** 2021–2036 (1994).

37. Piek, E., Moustakas, A., Kurisaki, A., Heldin, C.-H. & ten Dijke, P. TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. *J cell Sci* **112,** 4557–4568 (1999).

38. Stroschein, S. L., Wang, W., Zhou, S., Zhou, Q. & Luo, K. Negative feedback regulation of TGF-β signaling by the SnoN oncoprotein. *Science* **286,** 771–774 (1999).

39. Miyazono, K. Positive and negative regulation of TGF-beta signaling. *J Cell Sci* **113,** 1101–1109 (2000).

40. Jho, E.-h. *et al.* Wnt/β-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Molecular and cellular biology* **22,** 1172–1183 (2002).

41. Niida, A. *et al.* DKK1, a negative regulator of Wnt signaling, is a target of the β-catenin/TCF pathway. *Oncogene* **23,** 8520 (2004).

42. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (Feb. 2013).

43. Waters, J. C. *Accuracy and precision in quantitative fluorescence microscopy* 2009.

44. Sommer, C., Straehle, C., Koethe, U. & Hamprecht, F. A. *Ilastik: Interactive learning and segmentation toolkit* in *2011 IEEE international symposium on biomedical imaging: From nano to macro* (2011), 230–233.

45. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap (Chapman& Hall, CRC)* 1998.

Chapter 3

TWO ELEMENT TRANSCRIPTIONAL REGUATION IN THE CANONICAL WNT PATHWAY

3.1 Introduction

An important step in a cell responding to signal is the conversion of intracellular signal into transcription of particular genes. In a signal transduction pathway, this step is typically mediated by a transcription factor recognizing a specific cis-element in the DNA and initiating transcription. A typical signaling pathway, however, functions in thousands of different processes. These diverse outcomes may partly be explained through multiple complexities that have emerged at the transcriptional regulation. For instance, transcriptional outcomes could be tuned by the presence of different co-factors, or different chromatin modifications, or other transcription factors induced by signaling cross-talk. Here, we present another layer of complexity in which a transcription factor of a signaling pathway utilizes two distinct cis-elements in a gene's promoter.

The canonical Wnt pathway provides a representative model system to investigate how signal in converted into transcriptional response. The Wnt pathway regulates numerous fundamental processes throughout development and adult physiology, and is often disrupted in diseases [1–3]. Signal activation in the pathway is transduced by the β-catenin protein (Figure 3.1A). In the absence of Wnt stimulation, β-catenin is rapidly phosphorylated and degraded by a large destruction complex, built of multiple scaffolds (APC and Axin1) and kinases (GSK3β and CKIα). Wnt stimulation inhibits the destruction complex, leading to accumulation of β-catenin. Together with the Tcf/Lef transcription factors, β-catenin regulates a large cassette of genes by binding to a specific DNA sequence, CTTT GAWS [4]. This motif, known as the Wnt-Responsive Element (WRE), was identified through directed and random screen studies [5–8]. The prevalence of WRE in the Wnt target genes was observed in chromatin immunoprecipitation studies [9–12]. In some contexts, Tcf binding to WRE is potentiated by a nearby, within 10-bp proximity, Helper site, GCSGS [8].

Given the diverse roles of the pathway, it is not surprising that complexity in gene regulation is observed at various levels [4]. For instance, dominant negative forms of

Tcf/Lef proteins are produced by alternative promoters and mRNA splicing, which allows for diversification and sub-specialization of the Tcf isoforms in different tissues. Also emerging is a growing list of co-activators, co-repressors, and chromatin modifying factors recruited by β-catenin and Tcf. In some contexts, Wnt stimulation also induces β-catenin-mediated phosphorylation of some Tcf proteins, regulating their ability to bind to WRE [13]. But through all these years, the WRE(-Helper) site had remained the major cis-element by which signal from Wnt/β-catenin is channeled to gene regulation.

Recent evidence, however, suggests the inadequacy of this present model. In invertebrates, an armadillo/Pangolin-binding repressive site WGAWAW, in place of WRE, was found to regulate *Ugt36Bc* and *Tig* genes in Drosophila [14], raising questions if there are distinct Wnt-responsive elements yet to be found in vertebrates. Our own study in a vertebrate system indicated a suppressive regulation in endogenous Wnt targets that was not captured by the TopFlash reporter consisting of WRE sites alone [15]. Although endogenous genes are regulated by multiple factors and signaling pathways, we wondered if a more complex cis-regulation through the Wnt pathway itself may be present. In this chapter, in work in collaboration with Jaehyoung Cho and Thomas Hilzinger, I describe discovery of a second, distinct cis-element in the Wnt pathway in vertebrate systems, that is present with WREs and is also regulated by β-catenin and Tcf.

3.2 Results

We started with a previously reported discrepancy between the endogenous gene and TopFlash reporter expression (Figure 3.1B-D) [15]. Wnt signaling in early Xenopus blastulas activates dorsal regulators, including *siamois* and *Xnr3*. Treating the embryos for 5-10 minutes with 300 mM lithium is known to inhibit GSK3β [16], stabilize β-catenin [17], and dorsalize the embryos (Figure 3.1B-C) [18, 19]. We observed, however, that embryos treated with moderate doses of lithium (150 and 200 mM) largely retained a wild-type level of *siamois* and *Xnr3* expression (Figure 3.1C, see red arrows), and developed into wild-type tailbuds (Figure 3.1B). More strikingly, despite absence of marked phenotypic effects, the embryos showed increased β-catenin level (Figure 3.1D, see red arrows). Similar lack of embryo phenotypes despite increased β-catenin level was observed with other perturbations to the Wnt pathway, including injection of Axin1 and GBP mRNA. This led us to assay the TopFlash reporter, a commonly used reporter of Wnt signaling driven by a tandem repeat of WREs, to test whether the increased β-catenin was transcriptionally

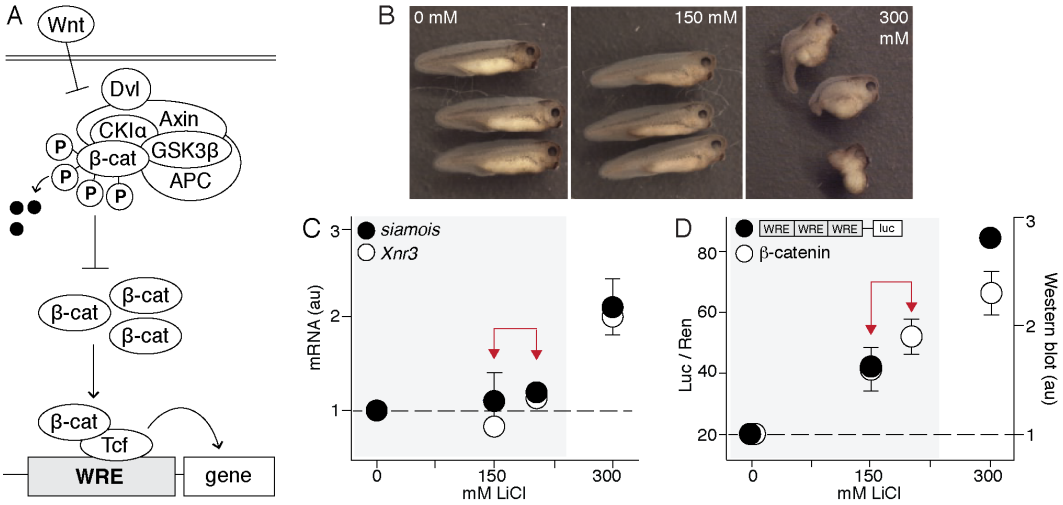


Figure 3.1: Endogenous genes show regulation not captured by WRE.
(A) In the canonical Wnt pathway, Wnt ligand stimulation inhibits the destruction complex, resulting in the accumulation of β-catenin. Together with Tcf/Lef proteins, β-catenin binds to WRE and activate or repress target genes.
(B) Xenopus embryos were treated with LiCl for 5 minutes at 32-cell stage, and harvested at stage 10 for qRT-PCR assay, and scored 3-4 days later (shown here).
(C) Expression of target genes, *siamois* (black circle) and *Xnr3* (white circle). Control embryos are untreated sibling embryos. Red arrows highlight how gene expression remains wild type despite perturbations.
(D) Black circle: luciferase/renilla signal from the TopFlash reporter injected at 4-cell stage. White circle: β-catenin level in the embryo measured using Western blot. Red arrows highlight how β-catenin level and TopFlash expression changes with moderate perturbations.
Figure B-D are reproduced from [15] with permission. Data are represented as mean ± SEM from 3-5 biological replicates. Error bars not visible have negligible SEM.

active. Indeed, the TopFlash reporter also showed increased activity at moderate doses of lithium (Figure 3.1D, red arrows). Therefore, even though the TopFlash reporter faithfully tracked the rise of β-catenin level, the contrasting wild-type expressions of the endogenous genes suggest a missing mechanism beyond WRE. It is notable that β-catenin itself, as a central regulator in the pathway, did not correlate with the target gene expression but tracked the extent of the perturbations, hinting at a possible role for β-catenin in the missing regulation.

An 11-bp negative regulatory element is necessary for dorsal specification of *siamois*

To locate the missing regulation, we focused on *siamois* [19–21]. Three WREs are found within 500 bp upstream of *siamois*, and are necessary for *siamois* activation (Figure 3.2A) [21, 22]. We built a luciferase reporter using a 3kb fragment of *siamois* promoter (pSia). We confirmed that 3kb pSia-luc mimics the temporal expression of *siamois*, beginning at mid-stage 8 and reaching steady state by stage 10. Reported here is the luciferase expression at early stage 10, reliably identified by the onset of dorsal lip formation.

We perturbed the GSK3β activity in a dose-response manner as described earlier. We found that the 3kb promoter recapitulated the endogenous *siamois* response, preserving wild-type expression at 150 mM LiCl (Figure 3.2B). By contrast, an 848bp fragment of *siamois* promoter that still contains the three WREs responded readily to perturbations (Figure 3.2C). To examine the spatial regulation of these promoter fragments, we tested lacZ reporters. Indeed, even though 3kb and 848bp-lacZ retained dorsal expression, the 848bp fragment showed expanded expression (Figure 3.2C-D inset). The missing regulation in the 848bp fragment was especially revealed when a moderate lithium dose was applied: the 3kb pSia-lacZ preserved dorsal expression at 150 mM LiCl, whereas the 848bp pSia-lacZ readily expanded.

Therefore, a suppressive, distal element located between 848bp and 3kb is necessary for the dorsal specification of pSia expression. Using the differential response at 150 mM lithium as a readout, we performed a promoter bashing screen to locate the element. Halving to 1.3kb preserved wild-type reporter expression, as did further truncations up until 963bp (Figure 3.2D, F). Loss of suppression at 150 mM LiCl was finally observed with truncation to 952bp, and with subsequent shorter constructs (Figure 3.2E, F). This analysis identified a suppressive element at 1kb upstream from *siamois*, with most activity centered on an 11-bp between 963 and 952bp (Figure 3.2G).

The 11-bp negative regulatory element (11-bp NRE) is AT-rich, 5'- CTG TTA TTT AA -3'. To further characterize the 11-bp NRE, we performed a mutagenesis analysis. With the 1.3kb promoter fragment, we mutated the 11-bp NRE one base at a time (i.e., purine into pyrimidine, or vice versa). We found that the majority of single-base alterations affected the response, with the largest effect produced by a single-base mutation on the 11th nucleotide, which recapitulated the effect of deletion of the 11-bp NRE (Figure 3.2H). Finally, the 11-bp NRE is sufficient to

recapitulate the suppression. Inserting the 11-bp NRE to 848bp promoter (Figure 3.2F) rescued the wild-type expression. The same effect was observed when the 11-bp NRE was pasted to even shorter promoters (Figure S3.1). These results suggest that regulation of *siamois* not only requires WREs, but also a suppressive 11-bp NRE.

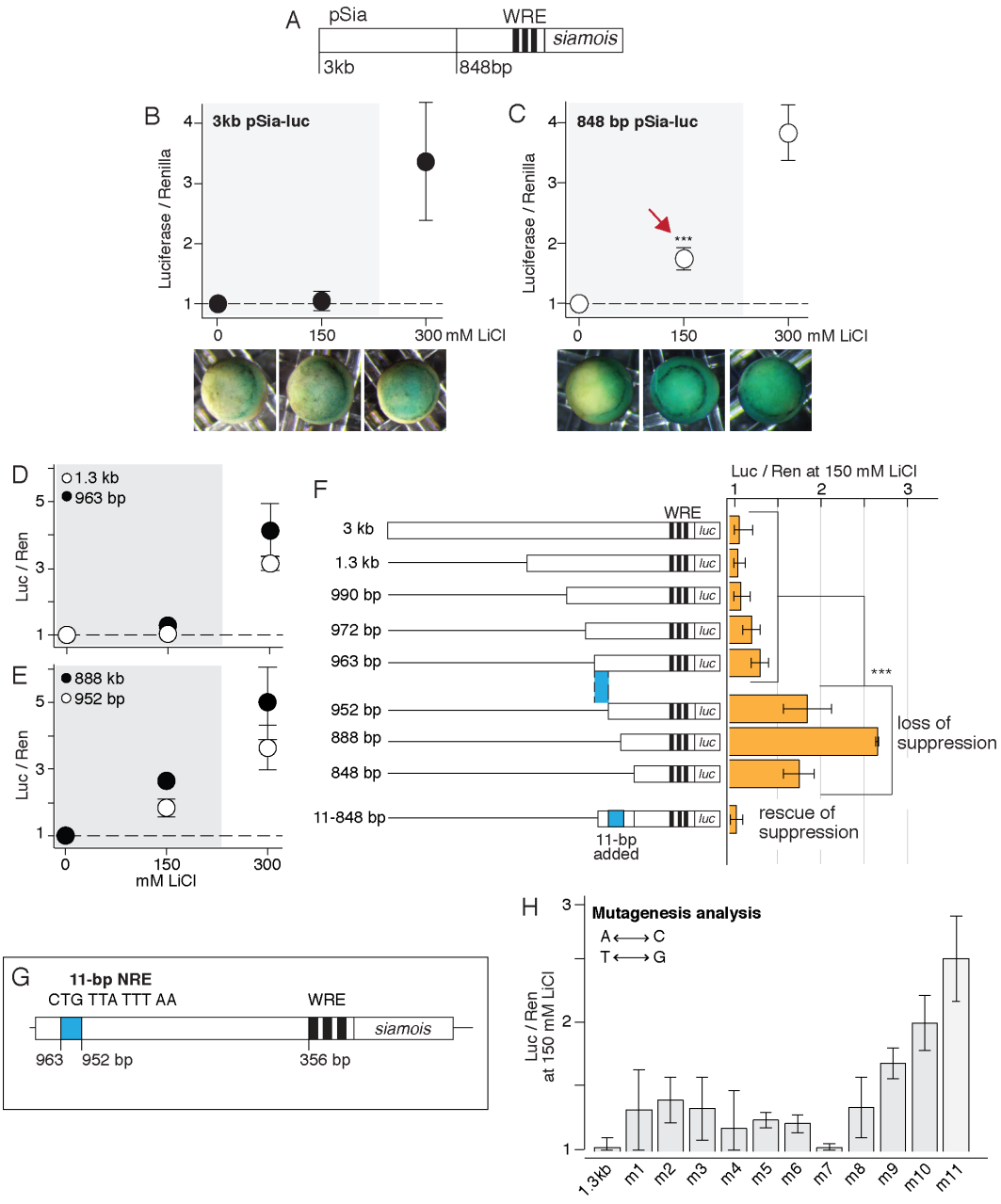


Figure 3.2: A suppressive 11-bp NRE is necessary for *siamois* regulation

(A) Three WREs (black) are located within 500 bp upstream in *siamois* promoter (pSia)

(B-F, H) We built luciferase reporters using 3kb and 848bp pSia. The luciferase reporters were injected into each cell at 4-cell stage. Injected embryos were treated with lithium for 5 minutes at 32-cell stage, and harvested for dual-luciferase assay at stage 10. As an injection control, pRL-TK constitutively expressing renilla luciferase was co-injected, and the pSia-driven firefly luciferase signal was measured relative to the renilla luciferase signal. In all the plots shown here, the luciferase/renilla signal is normalized to that in the control, untreated embryos. Data are presented as mean ± SEM from 3-5 biological replicates. Error bars not visible have negligible SEM.

(B) Expression of 3kb pSia-luc.

(C) Expression of 848bp pSia-luc. p-value = 8.4E-4 (Student's t-test). Insets below Figure B and C: *Xenopus* embryos were injected with 3kb or 848bp pSia-LacZ at 4-cell stage, treated with lithium for 5 minutes at 32-cell stage, and fixed at stage 10 for X-Gal staining. In all embryos, dorsal is to the right.

(D) Expression of 1.3kb pSia-luc (black) and 963bp pSia-luc (white).

(E) Expression of 952bp pSia-luc (white) and 888bp pSia-luc (black).

(F) Expression of pSia of various lengths in embryos treated with 150 mM LiCl. p-value = 1.5E-5 (Student's t-test). See also Figure S3.1.

(G) A suppressive 11-bp NRE (blue) is located between 963 and 952 bp upstream of *siamois*.

(H) Mutagenesis analysis of the 11-bp NRE. Data are mean luciferase/renilla signal ± SEM from 2-4 biological replicates.

The 11-bp NRE interacts with Tcf

Next, we set out to find the factors that bind to the 11-bp NRE using EMSA (Figure 3.3A, B). As the protein source, we used the high-speed supernatant of Xenopus egg extracts. We reasoned that the binding factor(s) must be maternal, as Wnt signaling begins in early blastula before zygotic transcription [22, 23]. As the DNA probe, we synthesized a 30-bp double stranded oligomer containing the 11-bp NRE, flanked by the endogenous sequence. As a control for the probe, we used the strongest mutant with an A->C switch at the 11th nucleotide (m11 mutant; Figure 3.2H). We observed specific binding to the 11-bp NRE in the EMSA (Figure 3.3C). The band was competed away when excess, unlabeled DNA probe was added (Figure 3.3C, lanes 3-4). The band was competed away much less effectively when excess, unlabeled m11 mutant probe was used (Figure 3.3C, compare lanes indicated by blue arrows). The specific band was reproducible across batches of extracts, despite variation in non-specific binding patterns and the amount of competitors needed (Figure 3.3C). Having the specific EMSA activity, we proceeded to identify the binding factor(s) by testing some known transcription factors. Lacking good antibodies to these factors in Xenopus, we tested if the DNA binding sites of the factors would compete with the EMSA band. Binding sites of various transcription factors, e.g., CREB, p53, NF-κB, AP-1, failed to compete significantly (Figure S3.2A). Unexpectedly, the one DNA fragment that competed away the EMSA band was the WRE itself (Figure 3.3D). The specific band was competed away when excess, unlabeled WRE probe was added (Figure 3.3D, lanes 5-6) – to a similar extent as it was competed away by excess, unlabeled wild-type probe (Figure 3.3C, lanes 3-4). By contrast, a negative control probe, carrying two mutations that destroy binding of the Tcf protein (i.e., FopFlash construct) [24], competed much less effectively (Figure 3.3D, lanes 7-8, see blue arrows). The competition by WRE suggests that Tcf proteins bind to the 11-bp NRE. Indeed, XTcf3 antibodies, raised against an N-terminal (XTcf3n) and C-terminal fragment (XTcf3c) [14], competed away the specific band (Figure 3.3E). These results suggest that a Tcf/Lef protein binds the 11-bp NRE, and that XTcf3 is the predominant binder in our in vitro assay.

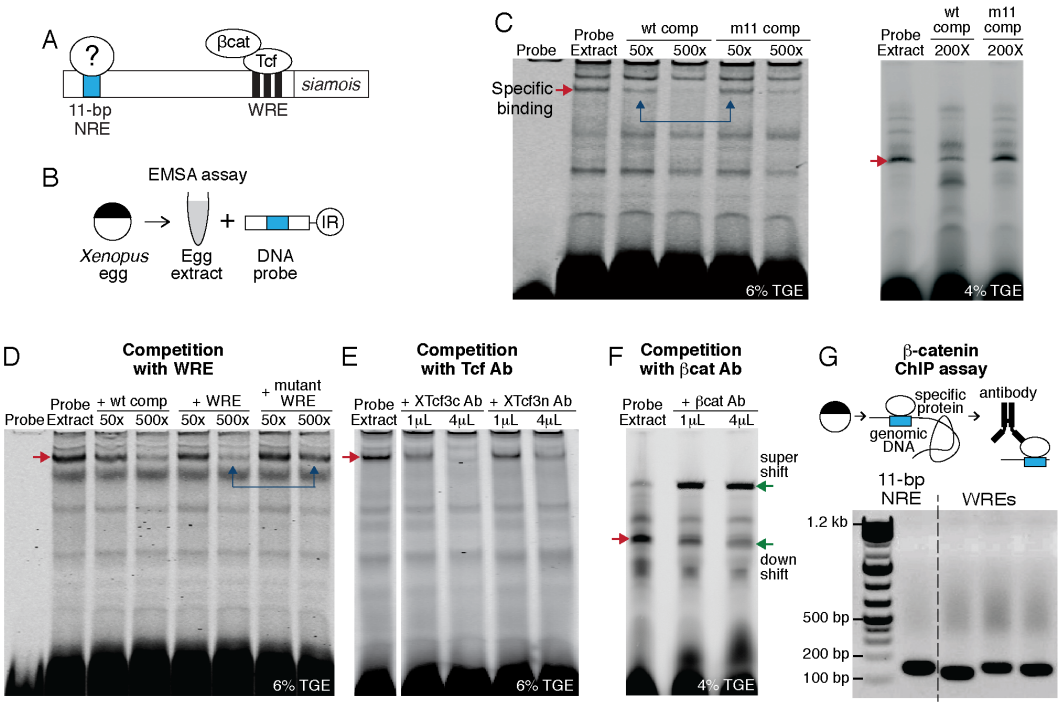


Figure 3.3: The 11-bp NRE binds to β-catenin and Tcf.

(A) We looked for the factor(s) that bind to the 11-bp NRE.

(B) We performed EMSA using *Xenopus* egg extract and a 30-bp probe containing the 11-bp NRE. IR denotes infrared dye used to tag the DNA probe.

(C-F) EMSA analysis. Every panel shown comes from a single gel. Each experiment was repeated 2-4 times. In all gels, red arrow indicates the specific EMSA band.

(C) Left gel: Competition with excess, unlabeled wild-type probe (lanes 3-4) and excess, unlabeled probe containing the m11 mutation (lanes 5-6). Right gel: EMSA using a different batch of *Xenopus* extracts, and resolved using a lower-percentage gel.

(D) Competition with excess, unlabeled wild-type probe (lanes 3-4), WRE probe (lanes 5-6), and mutant WRE probe (lanes 7-8). See also Figure S3.2A.

(E) Competition with polyclonal XTcf3 antibody against the C-terminal of XTcf3 (XTcf3c, lanes 2-3) and against the N-terminal of XTcf3 (XTcf3n, lanes 4-5).

(F) Competition with polyclonal antibody against β-catenin. Red arrow: the specific EMSA band. Green arrows: a supershift and a smear downshift. See also Figure S3.2B.

(G) Chromatin immunoprecipitation using β-catenin antibody. Genomic DNA was isolated from stage 10 *Xenopus* embryos, sonicated, and pulled-down with β-catenin antibody. Lane 1: PCR amplification from the 11-bp NRE region. Lanes 2-4: PCR amplification from regions containing WRE in the *siamois* promoter.

The 11-bp NRE interacts with β-catenin

The binding of Tcf protein to the 11-bp NRE raises questions on the binding partner, as Tcf does not usually act alone [4]. The binding of Tcf to the 11-bp NRE is also interesting in light of our earlier findings (Figure 3.1) that β-catenin intriguingly tracks the extent of perturbations, motivating us to test the roles of β-catenin. We found that adding polyclonal β-catenin antibody to the EMSA reaction produced a strong supershift, and a weak downshift smear (Figure 3.3F). To further confirm the binding of β-catenin, we performed binding assay using purified recombinant *Xenopus* β-catenin and Tcf3 (Figure S3.2B). As expected, Tcf3 alone produces a smear signal, suggesting an unstable binding. A strong, sharp band was observed in the presence of both Tcf3 and β-catenin.

To test if β-catenin acts on the 11-bp NRE in the *siamois* promoter in vivo, we performed chromatin immunoprecipitation in the embryos (Figure 3.3G). We collected embryos at stage 10, when *siamois* expression is at peak. We observed a strong signal indicating β-catenin binding in the 11-bp region. In the same experiment, as expected, β-catenin binding was also detected from the WRE region in the *siamois* promoter. As a negative control, negligible signal was detected with IgG antibody and from *siamois* or ODC coding regions. These results suggest that the 11-bp NRE, required for *siamois* regulation, interacts with β-catenin and Tcf.

The 11-bp NRE is present in the promoter of more target genes

Beyond *siamois*, might the 11-bp NRE regulate other Wnt target genes? At first, we found no results looking for the exact 11-bp sequence in other *Xenopus* targets. However, examining the *siamois* promoter closely, we identified two sites with a sequence pattern similar to that of the 11-bp NRE (Figure 3.4A-B), a G/C cap, followed by 8 A/T's – with an occasional C/G in the 5th position. All 11bp-like elements competed with the WRE (Figure S3.3A). Hence, three 11-bp NREs are present in the *siamois* promoter, although losing the distal one is sufficient to disrupt the suppression. Following this lead, we searched for a similar sequence pattern in other known direct Wnt targets in *Xenopus*. We found candidate 11-bp NREs in the promoter of *Xnr3* and *engrailed* (Figure 3.4B). The predicted 11-bp NREs in *Xnr3* and *engrailed* produced a specific EMSA band, which was shifted by polyclonal β-catenin antibody (Figure S3.3B).

The 11-bp NRE regulates *Brachyury* expression in mouse embryonic stem cells

With the predictive sequence pattern (Figure 3.4C), we investigated if 11-bp NRE is present in mammals. We examined promoters of known direct Wnt targets in mouse embryonic stem cells (mESC). We identified several candidates, including some in the promoters of well-characterized Wnt targets, such as *Brachyury* (*T*), *Axin2*, and *Cdx4* (Figure 3.4B). EMSA assay confirmed that these 11-bp NREs produced a specific band that was shifted by β-catenin antibody (Figure S3.3C).

To test whether the 11-bp NREs function in regulation of mammalian Wnt targets, we examined *T* regulation. *T* promoter has two WRE sites, at 191 and 273 bp upstream of the transcription start site (Figure 3.4D) [25]. *T* is an early marker of mesoderm differentiation. Basal expression of T, present in a fraction of stem cell population [26], is activated by endogenous secretion of Wnt proteins [27], and marks the early mesoderm-committed (EM) progenitors [26].

We identified two candidate 11-bp NREs, at 999 and 1613 bp upstream from the transcription start site (Figure 3.4D). To test whether the 11-bp NREs interact with β-catenin in vivo, we performed chromatin immunoprecipitation using β-catenin antibody. We observed signal from the proximal 11-bp NRE (Figure 3.4E). As positive control, strong signal was also observed from the -273bp WRE (Figure 3.4E). As negative controls, almost tenfold lower signal was observed using mouse IgG antibody, and negligible signal from the exon 2 region of *T*. We observed much lower signal from the distal NRE, suggesting the proximal one as the dominant NRE in the *T* promoter.

To test the roles of the 11-bp NREs in *T* regulation, we used Crispr/Cas9 to delete the 11-bp NREs from the endogenous promoter of *T*. Four sequence-verified TΔ11-bp clones were then assayed for *T* expression by qRT-PCR. We observed that TΔ11-bp clones showed significantly higher expression levels of *T* than the wild type cells (Figure 3.4F), suggesting increased EM progenitors. As a control, four wild-type clones that underwent blank transfection and clonal selection showed no significant change of T expression. These results suggest that balanced regulation of *T* in mESC require 11-bp NREs as well as the WREs.

The 11-bp NRE is prevalent in β-catenin Chip-Seq in human cells

Finally, equipped with the criterion of 11-bp NRE gathered across several Wnt direct targets (Figure 3.4C), and one confirmed with binding and functional assays (Figure 3.2,3.3, 3.4E-F), we asked whether 11-bp NREs are present more widely across Wnt target genes. To address this question, we analyzed β-catenin Chip-Seq

datasets from human cells [11, 28]. We found significant enrichment of 11-bp NREs in β-catenin peaks from HEK293T (p-value: 2.58 e-31) and HCT116 cells (p-value: 4.24e-12). Multiple 11-bp NREs were often predicted within the β-catenin-bound fragments. We found significant co-localization of 11-bp NREs with WRE: of the 2818 β-catenin peaks that contain WRE, 71% also contain 11-bp NREs in in HEK293T cells (Figure 3.4G), and 45% do so in HCT116 cells (Figure S3.4). Our analysis also indicates a significant fraction of β-catenin-bound fragments that contains 11-bp NREs only (Figure 3.4G), suggesting that the 11-bp NRE may have more functions beyond what we found here.

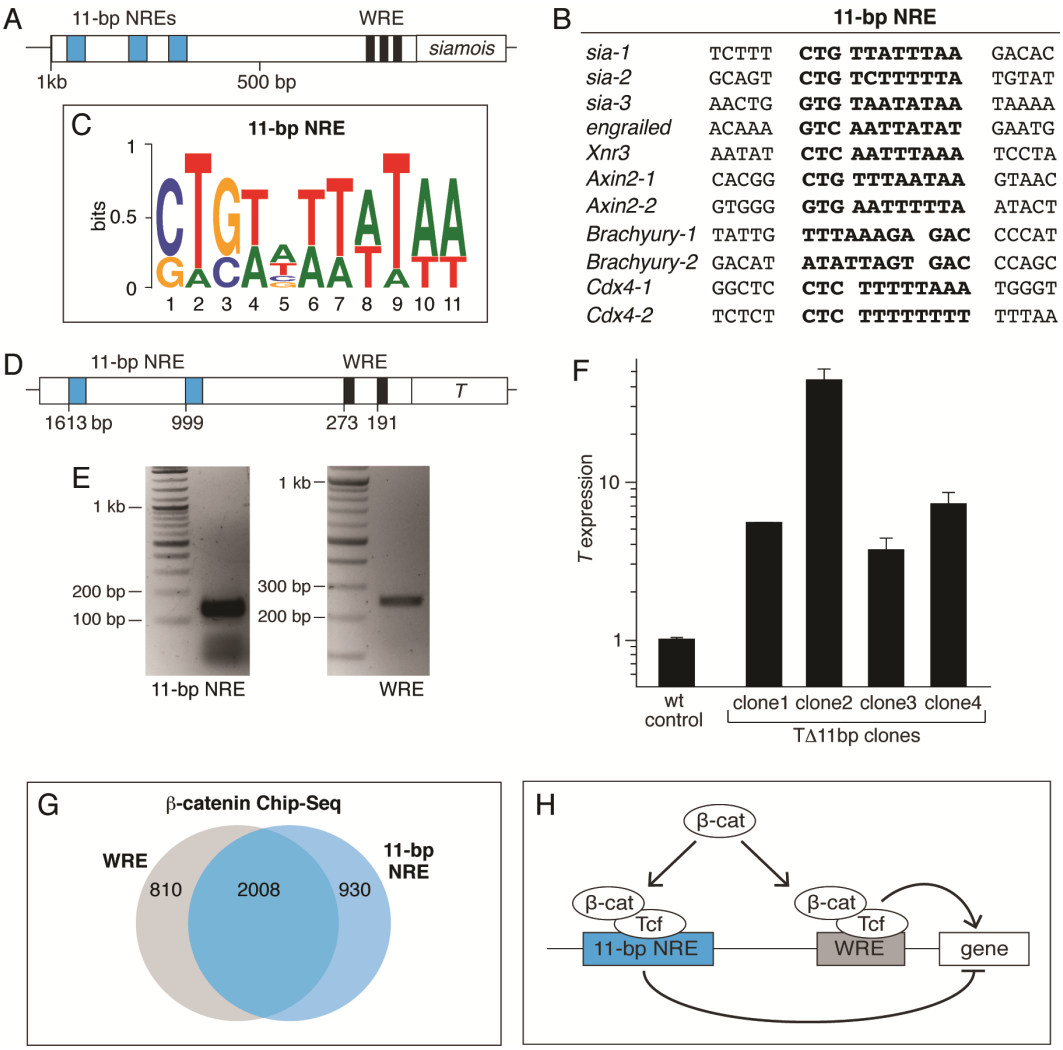


Figure 3.4: 11-bp NREs regulate *Brachyury (T)* in mouse embryonic stem cells.
(A) *siamois* promoter contains 3 11-bp NREs, including the distal element characterized so far (blue). See also Figure S3.3A.
(B) 11-bp NREs found in the promoters of *siamois, engrailed, Xnr3, Brachyury, Axin2, Cdx4*.
(C) Position frequency matrix built using the identified 11-bp NREs in (B). See also Figure S3B-C.
(D) *T* promoter in mESC contains 2 WREs and 2 predicted 11-bp NREs.
(E) ChIP using β-catenin antibody, followed by PCR amplification from 11-bp NRE (left) and WRE region (right) in the *T* promoter. The result was reproducible across two biological replicates. To ensure that immunoprecipitation of the 11-bp NRE fragments was not confounded by WRE, we sonicated the chromatin to 100-300 bp and performed PCR validation.
(F) Crisp/Cas9 was used to target genomic deletion of the 11-bp NREs in the *T* promoter. Four mESC clones carrying genomic deletion of 11-bp NREs (TΔ11-bp) were analyzed with qRT-PCR for *T* expression. Error bars indicate SD from 3 biological replicates.
(G) Analysis of β-catenin Chip-Seq on HEK293T cells. We examined 1kb β-catenin peak regions for the presence of the 11-bp NRE and WRE. Out of the 4484 total peaks from the β-catenin Chip-Seq, 3748 contain the 11-bp NRE and/or WRE motifs. Out of these 3748 peaks, 2008 contain both motifs. See also Figure S3.4.
(H) Our findings suggest that signal in the Wnt pathway does not only activate target genes through WRE, but also tunes expression of the gene through a suppressive 11-bp NRE. Further, our findings also suggest that β-catenin mediates this coupling.

3.3 Discussion

We found an 11-bp NRE that β-catenin and Tcf act on in the promoters of many Wnt target genes. The 11-bp NRE is necessary for *siamois* regulation in *Xenopus* embryos and Brachyury regulation in mouse embryonic stem cells. Our study suggests a new model of gene regulation in the Wnt pathway, where signal in the pathway not only activates target genes through WRE, but also in some contexts, tunes expression of the genes through a distinct 11-bp NRE. (Figure 4I). Interestingly, although multiple NREs are present in *siamois* and *T* promoter, there seems to be a dominant one that mediates suppression. The Chip data in *T* promoter suggests that the other NRE binds less strongly to β-catenin and therefore may play subtler roles. In Drosophila, it was found that Wnt signaling represses Ugt36Bc and TiG genes through a WGAWAW site that binds to β-catenin [14]. Our finding identifies a distinct repressive site in vertebrates, and one that implicates a different role for β-catenin, where it acts in opposite manner in the same promoter, as an activator through WRE and a suppressor through the 11-bp NRE. With regard to Tcf, our finding can be consistent with models in the field, where the same or distinct Tcf proteins may specialize as an activator or a repressor [4].

A circuit where the input activates the output, and at the same time suppressively tunes the output, is known in engineering as the incoherent feedforward loop (IFFL). Such a circuit has been found in transcriptional networks of bacteria, yeast, and human cells [22-25]. While these analysis focused on protein-protein interactions, an interesting aspect of the IFFL we found in the Wnt pathway is its implementation through multiple cis-elements. We know of one other instance where a regulator works through multiple cis-elements in an IFFL mode, in the regulation of porin OmpF in *E. coli* [26].

More generally, IFFL belongs to a class of recurring strategy in biological systems, where a biological molecule is used in a paradoxical manner (27). Paradoxical circuits, and the incoherent feedforward circuit specifically, moreover, are versatile circuits. By tuning the relative strengths and time scales of the activation and repression arm, an incoherent feedforward circuit can generate a sustained, net activation – but beyond that, also a net repression, a temporal pulse, response acceleration, band-pass filtering, and fold-change detection (see Methods S1, as well as refs. [28-34]). Moreover, inclusion of an IFFL downstream the Wnt pathway can explain how the endogenous gene response remains wild type despite perturbations, either by acting as an amplitude filter (if the timescales of activation and repression are similar) or fold-change detector (if the repression is slow and

strong) (see Methods S1). Although it is difficult presently to estimate the relative affinities of the two sites, an extract system combined with sophisticated methods such as SPR could be a next approach. It would also be interesting to investigate whether different co-factors or chromatin modifiers are recruited by β-catenin to the 11-bp NRE. Thus, the finding of the 11-bp NRE not only suggests a different model of gene regulation in the Wnt pathway, and one that implicates a function for β-catenin as a feedforward sensor, but also provides a plausible mechanism by which gene regulation by β-catenin and Tcf can generate more versatile dynamics across contexts.

3.4 Materials and Methods

***Xenopus* in vitro fertilization** was performed according to [29]. Testes were isolated from male X. laevis and stored in 1x Marc's Modified Ringer (MMR) with 50 ug/ml gentamicin for up to two weeks. Ovulation was induced in female X. laevis by injection of 600-800U human chorionic gonadotropin (HCG) 14-16 hours prior to egg collection. Eggs were fertilized by incubation in a sperm solution in 0.1xMMR for 30 minutes. Fertilized eggs were dejellied in 20 mg/ml cysteine, pH 8.0 in 0.1xMMR for 2-5 minutes. Embryos were reared in 0.1xMMR at 14C.

DNA constructs. 3kb fragment of the *siamois* promoter [20] was inserted into pGL4.12 (Promega, E6671) using KpnI and BglII. For promoter analysis, the desired length of a promoter fragment was amplified out, and cloned into pGL4.12 using KpnI and BglII. For the rescue experiments, the desired length of a promoter fragment was amplified out with the 11-bp sequence added on the 5' end, with 5bp of endogenous sequence 3' to the 11-bp NRE included as a spacer. For mutagenesis analysis, single-base mutation was performed on 1.3 kb pSia-luc using QuikChange II (Agilent Technologies, 200523).

Embryo microinjection. Needles for microinjection were made from capillary tubing (Borosil 1.2mm x 0.9mm ID, FHC Inc, 30-31-0) pulled using Flamming/Brown micropipette puller (Sutter Instrument, P-97). The needle was calibrated to produce a 10nl ejection volume. Calibrated needles were attached to a microinjector (Warner Instruments, Picoliter Microinjector PLI-100A) and filled with the DNA construct solution. Embryos were collected at 4-cell stage, and placed in 5% Ficoll/0.1xMMR. Each cell in the embryos was injected with 10nl solution on the equator for a total of 4 injections per embryo. Injected embryos were reared in 0.1x MMR.

Lithium treatment. Treatment with lithium chloride (Sigma, L4408) was performed when the embryos were at 32-cell stage, by incubating the embryos for 5 min in 0.1x MMR with 0 mM, 150 mM, or 300 mM LiCl. Embryos were subsequently thoroughly washed, let develop further in 0.1x MMR, and harvested at the appropriate stage for different assays.

Dual luciferase assay. Embryos were injected at 4-cell stage with a total of 100 pg/mL pSia-luc and 25 pg/mL modified pRL-TK (Promega, E2241). The

modified pRL-TK has the Rluc gene replaced with the hRlucCP from pGL4.84 (Promega, E7521), which is the destabilized version of Rluc and produces quicker and stronger expression than its stable counterpart. Dual luciferase assay was performed using the Promega system (E1960). Three to four pools of 10 embryos at stage 10 were collected from each sample and lysed in 50 ul of Passive Lysis Buffer. 5-10ul of cleared lysate was analyzed on a Victor X plate reader (PerkinElmer). For each biological replicate, 2-3 technical replicates were measured. The ratio of firefly to renilla luciferase activity (Luc/Ren ratio) was calculated for each data point, and averaged across the technical replicates.

β-galactosidase assay. Xenopus embryos were injected with 3kb or 848bp pSia-LacZ at 4-cell stage, treated with lithium for 5 minutes at 32- to 64-cell stage, and fixed at stage 10. To visualize the promoter activity, expression of β-galactosidase was detected by x-gal staining following the manufacturer's procedure (Thermo Fisher, K1465-01). 30 embryos/tube were incubated in the fixing buffer for 30 minutes on a nutator at room temperature, rinsed twice with 1x PBS + 2 mM MgCl2, and stained with 0.5 mL of staining buffer at 37C in the dark until the stain develops. Afterward, the embryos were washed twice with 1x PBS, twice with methanol, and photographed.

***Xenopus* extract preparation** was performed according to [30] with modifications. Female *X. laevis* frogs were injected with 50 units of PMSG 2 days prior and 600-800U HCG 16 hours prior to egg collections. Each frog was placed in a container with 3-4L of 1xMMR. The next day, eggs were rinsed, examined for quality, and dejellied in 10 mM DTT / 50mM Hepes pH 8.2. Dejellied eggs were rinsed in Wash Buffer (20 mM Hepes pH 7.9, 2 mM MgCl2, 0.1 mM EDTA pH 8, 100 mM KCl), transferred to Extraction Buffer (Wash Buffer, 1mM DTT, 10 ug/mL each of leupeptin, chymostatin and pepstatin A). Eggs were packed by quick centrifugation, the excess liquid removed, and crushed by centrifugation at 10,000g for 15 minutes. The crude supernatant was diluted to 10mg/mL, and then cleared further with centrifugation at 100,000g for 1 hour. The high speed supernatant was used for EMSA analysis.

Electric Mobility Shift Assay (EMSA). Double stranded DNA probes were synthesized from IDT and labeled at the 5' end of one strand with IR680 or IR800 dye. 4 ul of Xenopus egg extract was added to a 20 ul total reaction

mixture containing 20 mM HEPES pH 7.9, 50 mM KCl, 10 mM MgCl2, 1 ug/ml poly(dI-dC), 4% Ficoll 400, and incubated for 30 min on ice. 1 ul of labelled probe DNA (50 fmol) was then added, and the reaction was incubated for another 2 hours to overnight on ice. The whole binding reaction was loaded onto a 4% or 6% native DNA retardation gel with 1x TGE buffer (Novex EC6058BOX, Invitrogen). Gels were run at 115 V at 4C in 1x TGE buffer for 1.75 hr (before the actual run, the gels were pre-run for an hour). For the probe competition assays, the reaction mixture was pre-incubated with unlabeled dsDNA competitors for 30 minutes, before adding the labeled probe. For the antibody competition assay, specific antibodies were pre-incubated with the reaction mixture for 1.5 hrs before addition of the labeled probe. The gel was imaged directly with the LiCOR Odyssey Imager.

Purification of recombinant Xβ-catenin and XTcf3. Xβ-catenin-His9 or GST-XTCF3-His9 in pETKatN10 were expressed and purified from E. coli BL21 (DE3) (EMD, 69450). For protease inhibitors, we used 1x cOmplete EDTA-free protease inhibitor (Roche, 11873580001). Transformed cells were cultured with shaking at 37C until OD600 is 0.4, and then chilled on ice for 20 min. Protein expression was induced with 0.4 mM IPTG for 3hrs at 25C. Cell pellets were suspended in lysis buffer (1x PBS, 500 mM NaCl, 0.5% Triton X-100, 0.1% BSA, 1% glycerol, 5 mM DTT, protease inhibitors) and sonicated 5x10 seconds at 30% output. Crude cell extracts were centrifuged for 20 min. GST affinity purification: clarified extract was incubated with glutathione sepharose beads (GE Healthcare, 17-0756-01) for 2.5 hrs at 4C. Beads were washed with wash buffer (1x PBS, 500 mM NaCl, 0.5% Triton X-100,1% glycerol, 5 mM DTT, protease inhibitors) for 30 min at 4C. GST-XTCF3-His9 was eluted with elution buffer (20 mM reduced glutathione, 50 mM Tris HCl pH 9.5, 150 mM NaCl, 1 mM DTT, 0.1 % Triton X-100), dialysed against dialysis buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM DTT, 0.01% TritonX-100) using Slide-A-Lyzer 10,000 MWCO (Thermo Scientific), and concentrated with Vivaspin2 (VIVAproducts). His-tag affinity purification: the partially purified fraction from GST-affinity purification for GST-XTCF3-His9 or bacterial lysate expressing Xβ-catenin-His9 was applied to Ni-NTA column, and incubated with gentle rotation for 3 hrs at 4C. Unbound proteins were removed by gravity-flow and washed 3 times with wash buffer (1x PBS, 20 mM Imidazole, protease inhibitors). Target proteins were eluted with elution buffer (1x PBS, 250 mM Imidazole, pH 7.4) with gentle rotation at 4C, 30 minutes, concentrated and buffer changed to 10 mM HEPES-KOH (pH 7.9) with

Amicon Ultra-15 10,000 MWCO (EMD Millipore, UFC901008).

Chromatin immuniprecipitation. *Xenopus* embryos. Embryos at stage 10 were fixed and crosslinked with 1% formaldehyde/PBS for 1 hr. Crosslinking was stopped with 125 mM glycine/PBS. Fixed embryos were transferred to 600 ul of cold RIPA buffer (50 mM Tris-HCl, pH 7.4, 1% NP-40, 0.25% Na-Deoxycholate, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.5 mM DTT, Roche protease inhibitor cocktail), homogenized with a pestle, and then let sit on ice for 15 min. The chromatin was pelleted at 14,000g for 10 min at 4C. The pellet was resuspended in 650 ul of cold RIPA buffer, sonicated for 10 rounds of 20 second bursts with 30% output and 45 second intervals for each round. The supernatant was cleared with 14,000g for 10 min at 4C, and incubated with BSA-blocked protein G agarose beads (50 ul) for 2 hours to reduce non-specific binding. The pre-cleared supernatant was incubated with 1 ug of Xβ-catenin antibody or control IgG pre-coated protein G beads (50 ul per each condition) overnight at 4C with mixing. The beads were washed three times each with wash buffer A (20 mM Tris-HCl, pH 8.0, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl) and wash buffer B (10 mM Tris, pH 8.0, 0.25M LiCl, 1% NP-40, 1% Na-Deoxycholate, 1 mM EDTA). The washed beads were treated with DNase-free RNase A in 10 mM Tris-HCl, pH 8.0 for 1 hr, and then washed twice with 1 ml of TE (10 mM Tris, p8.0, 1 mM EDTA). To elute the immune-complexes, the beads were incubated in 250 ul of TES buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65C for 10 min with vigorous vortexing every 2 min. Crosslinking was reversed by heating at 65C overnight with 0.5 mg/ml Proteinase K, 0.1% SDS in PBS. The released DNA was purified and analyzed by PCR.

mESC. Cells were to grown to 80% confluency on 10 cm plates, crosslinked with 1% formaldehyde/PBS for 10 minutes at 21C. Crosslinking was stopped with 125 mM glycine/PBS and washed twice with PBS. Fixed cells were scraped and collected in 1 mL PBS with 10uL of 100x protease/phosphatase inhibitor cocktail (Thermo Scientific, 1861281). The cells were then centrifuged at 10,000 g for 10 minutes at 4C and resulting pellet was frozen at -80C. Nuclear fraction of the cells was isolated using a hypotonic lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.3% NP-40, 10% glycerol) [31]. We then proceeded with Chip-IT kit (Active Motif, 53040), with the given protocol except that we used Dynabeads Protein G (Thermo Fisher, 10003D). Sonication performed at 40 cycles of 0.7s on / 3.3s off at 20% amplitude using Branson SFX 250 Sonicator to break

down chromosomes into 0.1 – 0.3 kb fragments. For pulldowns, we used mouse β-catenin antibody (BD Transduction Laboratories, 610153) and normal mouse IgG (Millipore, 12-371B) as a control.

Crispr/Cas9 transfection in mESC. We obtained PX330 generously donated to Addgene by the Feng Zhang lab [32]. Into the PX330 plasmid, we inserted gRNA sequence targeting the T promoter, removed the Cas9 coding cassette, and inserted neomycin resistance cassette. To synthesize the donor plasmid for TD11-bp clones in mESCs, we used genomic prep of the cells to clone out 5' and 3' homologous arms in the T promoter. Another insert fragment was cloned from the genomic prep that included deleted 11-bp regions. The three fragments were integrated into a generic plasmid backbone using Gibson assembly (New England Biolabs, E2611S). For transfection, $2.5x10^5$ mESCs were plated on 24 wells and grown overnight. We used 400 ng of the modified PX330 plasmid, 150 ng of donor plasmid, 12 pmol Cas9 recombinant protein (IDT, 1074182) in Lipofectamine LTX with PLUS (Thermo Fisher, 15338100). The cells were transfected for 24 hours, transiently selected with neomycin for 5 days, dissociated into single cells, and grown as single clones.

RT-PCR. Total RNA from mESC was isolated with RNeasy extraction kit (Qiagen 74104). cDNA was synthesized using Quantitect RT (Qiagen 205311). Real-time RT-PCR was performed using Quantitect SYBR-Green kit (Qiagen 204143) using StepOnePlus RT-PCR System (Applied Biosystems).

3.5 Supplementary Figures

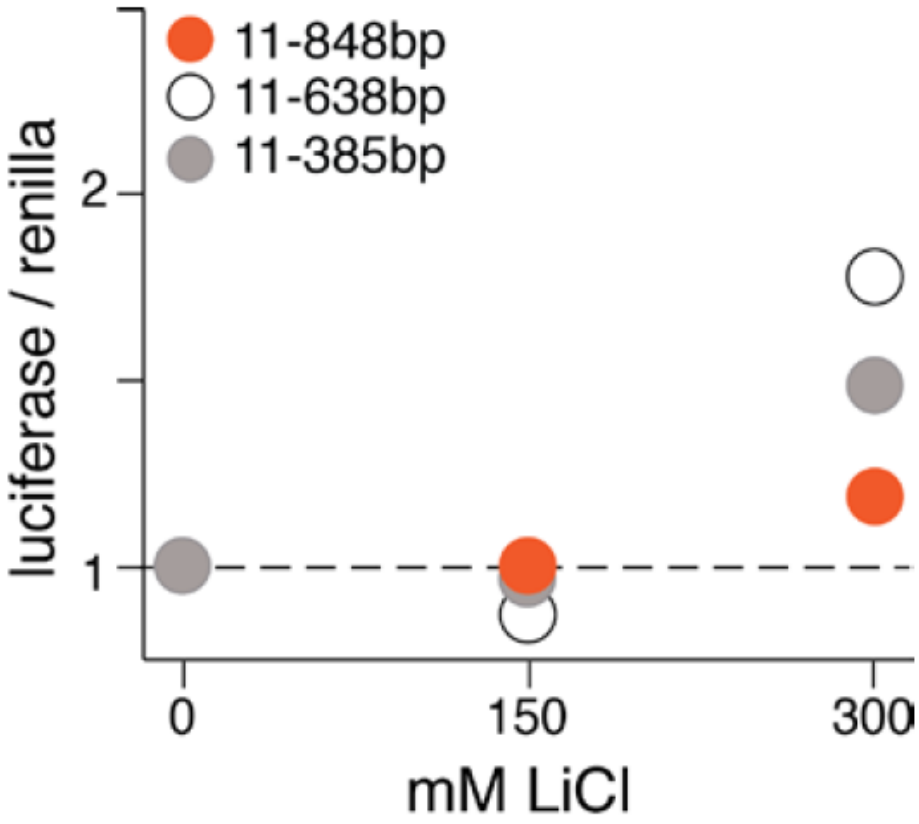


Figure S3.1: Rescue of suppression by 11-bp NRE was observed with even shorter promoters, related to Figure 3.2.
A 30-bp region containing the 11-bp NRE was added to the 5'-end of 848bp (orange), 638bp (white), and 385bp (grey) of *siamois* promoter. *Xenopus* embryos were injected with the constructs at 4-cell stage, treated with lithium for 5 minutes at 32- to 64-cell stage, and harvested at stage 10 for luciferase analysis.

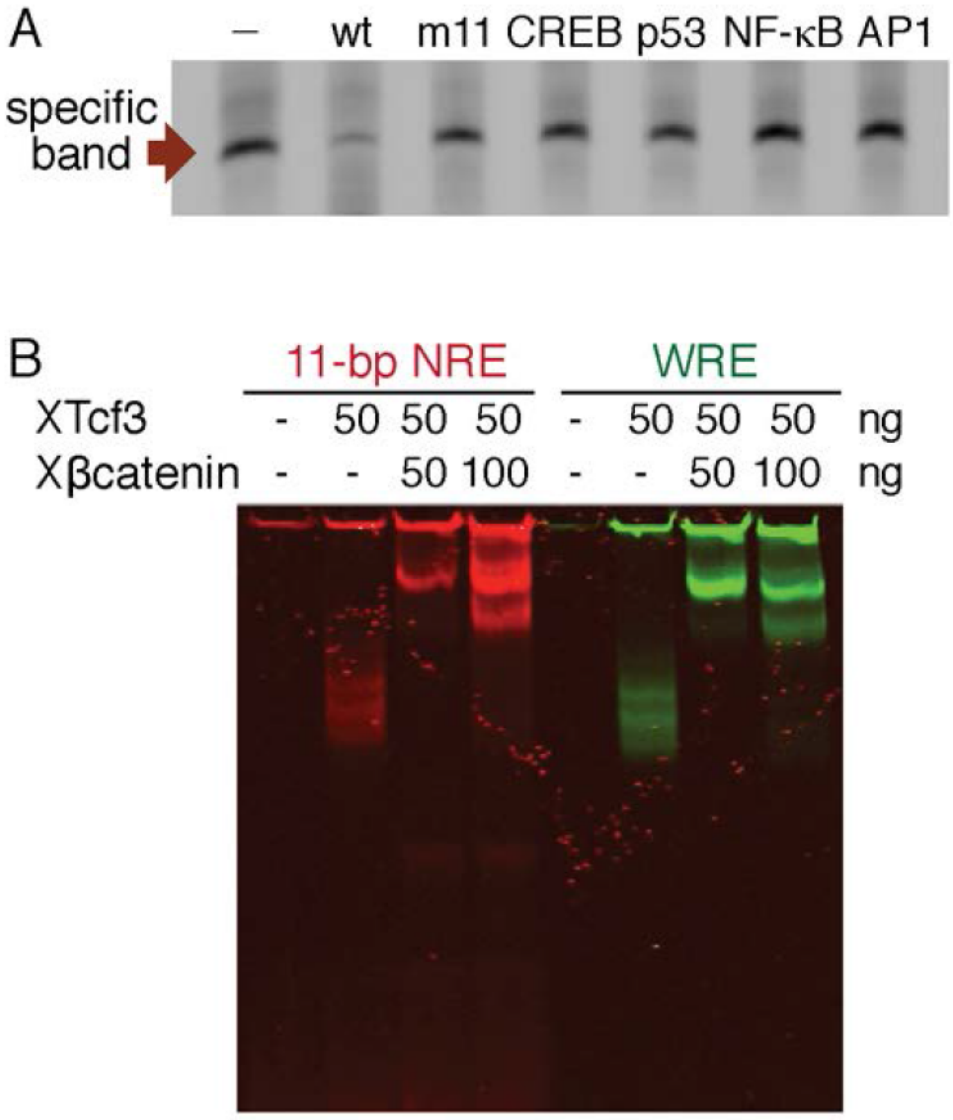


Figure S3.2: 11-bp NREs interact with β-catenin and Tcf3, but not other transcription factors, related to Figure 3.3.
(A) Known DNA binding sites of transcription factors failed to compete with the specific gel shift band. WRE probe without competition (lane 1), with 200x unlabeled 11-bp NRE (lane 2), m11 mutation competition (lane 3), and binding sites of transcription factors (lanes 4-7).
(B) Purified Xenopus β-catenin and Tcf3 proteins bound to the 11-bp NRE and WRE in dose dependent manner.

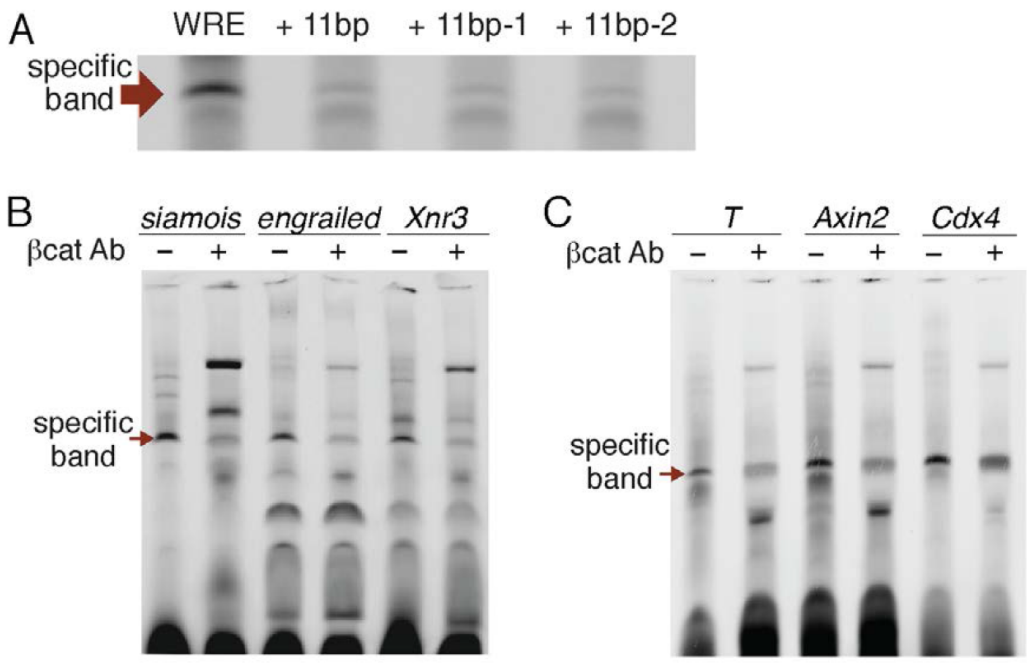


Figure S3.3: Predicted 11-bp NREs interact with TCFs and β-catenin, related to Figure 3.4.
(A) The three 11-bp NREs in siamois promoter compete with WRE. WRE probe without competition (lane 1) and with 200x unlabeled 11-bp NRE (lane 2) and the 2 elements similar to the 11-bp NRE (lanes 3 and 4).
(B-C) EMSA assays of predicted 11-bp NREs in promoters of siamois, engrailed, Xnr3
(B) T, Axin2, and Cdx4 (C). Red arrow: specific binding to the 11-bp NRE.

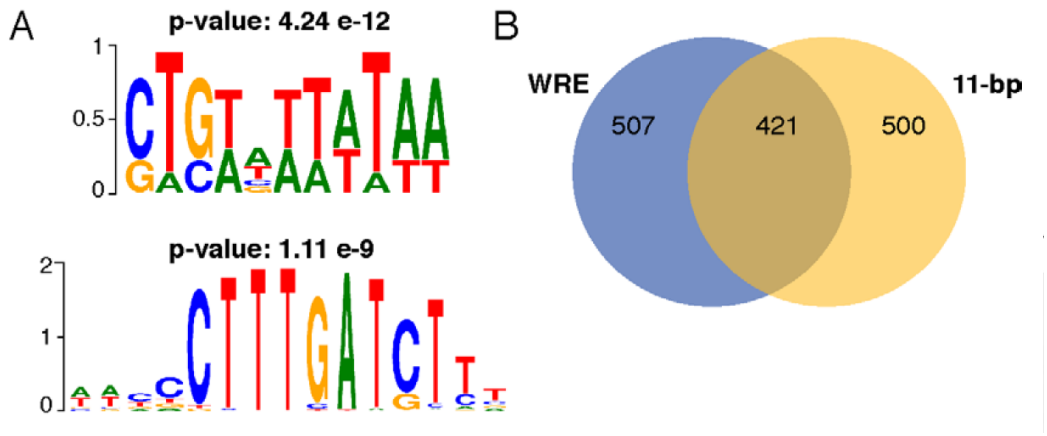


Figure S3.4: Analysis of β-catenin Chip-Seq on HCT-116 cells, related to Figure 3.4.

(A) The 11-bp NRE motif is significantly enriched in 600-bp peaks of β-catenin Chip-seq on HCT-116 cells. As a positive control, significant enrichment was also found for the WRE motif.

(B) Co-enrichment analysis. Out of the 2624 total 600-bp peaks from the β-catenin Chip-seq, 1428 contain the 11-bp NRE and/or WRE motifs. Out of these 1428 peaks, 421 contain both motifs.

3.6 Supplementary Information: Modeling analysis of incoherent feedforward loop

In this analysis, in collaboration with Harry Nunns, we illustrate how the 11-bp NRE mediates an incoherent feedforward loop (IFFL) that can be tuned to produce a variety of gene response dynamics. To convey the general concepts, we modeled IFFL using a simple, phenomenological description previously derived in refs [33, 34]. We considered two possible implementations of the IFFL, one involving single Tcf protein (Figure S3.5A), and another involving multiple Tcf proteins (Figure S3.5B). This is motivated by the observations that sometimes multiple Tcf proteins are involved in gene regulation [35].

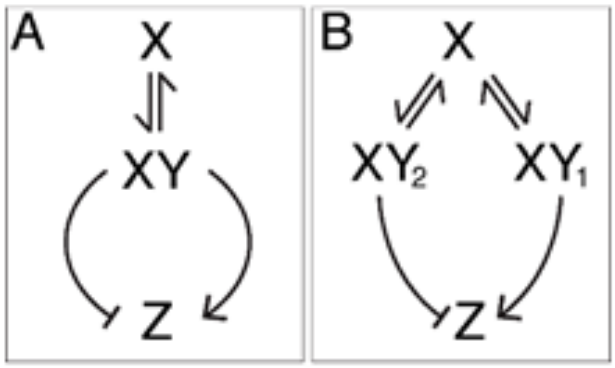


Figure S3.5: Model IFFL circuits in the Wnt pathway
(A) IFFL with a single Tcf protein. The input X binds to Y to form the complex XY. XY can activate or repress gene Z, depending on where it binds on the promoter.
(B) IFFL with multiple Tcf proteins. The input X binds to Y1 or Y2. XY1 activates Z, whereas XY2 represses Z.

X denotes concentration of the input β-catenin, Y's denote the concentrations of Tcf's, and Z denotes concentration of gene transcript output. Despite the different physical implementations, both models are mathematically similar, and produce similar outputs. Therefore, we show here only the analysis of a model with multiple Tcf's.

In the model, the input X can form a complex with Y1 or Y2 (Figure S3.5A). The dynamics of the complex formation are described by mass-action binding kinetics,

$$\frac{dXY_1}{dt} = \beta_1 \cdot X - \alpha_1 XY_1 \tag{1}$$

$$\frac{dXY_2}{dt} = \beta_2 \cdot X - \alpha_2 XY_2 \tag{2}$$

where α_1 and α_2 are the respective dissociation rate constants (k_{off}) of XY_1 and XY_2, and β_1 and β_2 are the product of the binding rate constants (k_{on}) of X to Y and concentration of the respective Y. During Wnt signaling, Tcf concentration is not thought to change considerably; therefore Y1 and Y2 are assumed as constants for simplicity in this analysis. Including time-varying Tcf would only facilitate more complex response dynamics. XY_1 binds to WRE and activates gene Z, whereas XY_2 binds to 11-bp NRE and represses gene Z. The dynamics of gene transcription is

$$\frac{dZ}{dt} = \beta_3 \cdot P_a - \alpha_3 Z \tag{3}$$

where β_3 is the rate of transcript production per unit time, α_3 is the degradation rate constant of the transcript Z, and P_a is the probability that transcription is active. In siamois and T promoter, WRE and 11-bp NRE are >500 bp apart. We therefore present here the case where the activator XY_1 and the repressor XY_2 bind to the DNA independently. The general conclusions hold for exclusive and cooperative binding, as was derived in [34]. With independent binding of the activator and repressor, the probability function Pa is, as derived in [33, 36],

$$P_a = (P_1)(P_{\text{N2}}) \tag{4}$$

$$P_1 = \frac{XY_1/K_1}{1 + XY_1/K_1} \tag{5}$$

$$P_{\text{N2}} = \frac{1}{1 + XY_2/K_2} \tag{6}$$

where P_1 is the probability that XY_1 is bound to WRE and P_{N2} is the probability that XY_2 is not bound to 11-bp NRE. K_1 is the dissociation constant of XY_1 to WRE and K_2 is the dissociation constant of XY_2 to 11-bp NRE. The full dynamic equation for Z is,

$$\frac{dZ}{dt} = \beta_3 \cdot \frac{XY_1/K_1}{1 + XY_1/K_1 + XY_2/K_2 + (XY_1 \cdot XY_2)/(K_1 \cdot K_2)} - a_3 Z \tag{7}$$

Figure S3.6 shows the solutions of Equation 7 in different regimes of parameters. For instance, when repression is slow, the activator gets a head start, and transcription commences, until repression begins and turns it off. In this case, the net effect is a pulse of gene activation. In the reverse case, when repression is fast, IFFL produces instead a pulse of gene repression. In the case where repression is slow and weak, IFFL can produce a net, sustained activation, where the repression does not completely turn off gene transcription, but merely modulates the steady-state level. Vice versa, when activation is slow and weak, IFFL produces a net, sustained

repression. Further, various feedbacks are present in the Wnt pathway (e.g., Dkk, βTRCP) [34, 37]. When coupled with a positive or negative feedback, an IFFL can produce bistable or oscillatory responses. Beyond producing the various response dynamics, IFFL can facilitate more complex computations. For instance, when the activation and repression arms are of a similar time scale, IFFL can act as an amplitude filter [38]. When repression is strong, IFFL can act as a fold-change detection [34]. A more detailed write-up on the analytics of these different solutions is available upon request.

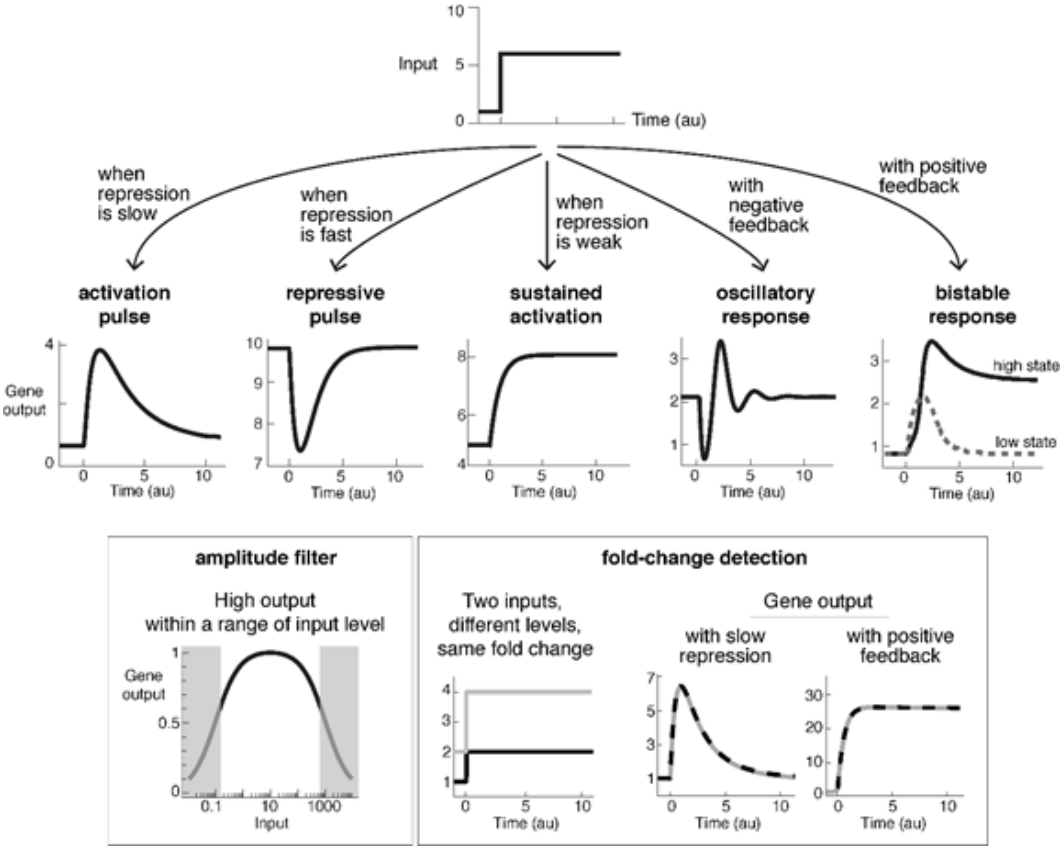


Figure S3.6: Simulations of IFFL in the Wnt pathway
(A) Model of the canonical Wnt pathway, redrawn from [39]. We implemented an IFFL downstream of this model, and simulated the model with varying GSK3β activity (k_9).
(B) Simulation of gene transcription in the presence and absence of 11-bp NRE. The IFFL parameter values used in these plots are: $\beta_1 = 3.0 \cdot 10^{(} - 4)\text{min}^{-1}$;$\alpha_1 = 0.30\text{min}^{-1}$; $\beta_2 = 2.0 \cdot 10^{-5}\text{min}^{-1}$; $\alpha_2 = 2.0 \cdot 10^{-4}\text{min}^{-1}$; $\beta_3 = 4.1\text{nM} \cdot \text{min}^{-1}$;$K_1 = 100\text{nM}$; $K_2 = .1\text{nM}$; $\alpha_3 = 4.1 \cdot 10^{-4}\text{min}^{-1}$

3.7 Work Contributions

Results from earlier work [15] contributed in directly motivating this work. K.K participated in design and execution of experiments in figure 3.4, figure development, and writing of the manuscript. Jaehyoung Cho participated in design and execution of experiments in Figures 3.2 and 3.3. Thomas Hilzinger participated in design and execution of experiments in 3.3 and 3.4 and writing initial manuscript. Harry Nunns performed modeling analysis of IFFL. Andrew Liu participated in designa and execution of experiments in figure 3.4.

3.8 Bibliography

1. Nusse, R. & Varmus, H. Three decades of Wnts: a personal perspective on how a scientific field developed. *The EMBO journal* **31,** 2670–2684 (2012).
2. Anastas, J. N. & Moon, R. T. WNT signalling pathways as therapeutic targets in cancer. *Nature Reviews Cancer* **13,** 11 (2013).
3. Loh, K. M., van Amerongen, R. & Nusse, R. Generating cellular diversity and spatial form: Wnt signaling and the evolution of multicellular animals. *Developmental cell* **38,** 643–655 (2016).
4. Cadigan, K. M. & Waterman, M. L. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harbor perspectives in biology* **4,** a007906 (2012).
5. Van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88,** 789–799 (1997).
6. Van Beest, M. *et al.* Sequence-specific high mobility group box factors recognize 10–12-base pair minor groove motifs. *Journal of Biological Chemistry* **275,** 27266–27273 (2000).
7. Hallikas, O. & Taipale, J. High-throughput assay for determining specificity and affinity of protein-DNA binding interactions. *Nature protocols* **1,** 215 (2006).
8. Atcha, F. A. *et al.* A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Molecular and cellular biology* **27,** 8352–8363 (2007).
9. Hatzis, P. *et al.* Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Molecular and cellular biology* **28,** 2732–2744 (2008).
10. Blahnik, K. R. *et al.* Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic acids research* **38,** e13–e13 (2009).
11. Bottomly, D., Kyler, S. L., McWeeney, S. K. & Yochum, G. S. Identification of β-catenin binding regions in colon cancer cells using ChIP-Seq. *Nucleic acids research* **38,** 5735–5745 (2010).
12. Norton, L. *et al.* Chromatin occupancy of transcription factor 7-like 2 (TCF7L2) and its role in hepatic glucose metabolism. *Diabetologia* **54,** 3132–3142 (2011).
13. Hikasa, H. *et al.* Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification. *Developmental cell* **19,** 521–532 (2010).

14. Zhang, C. U., Blauwkamp, T. A., Burby, P. E. & Cadigan, K. M. Wnt-mediated repression via bipartite DNA recognition by TCF in the Drosophila hematopoietic system. *PLoS genetics* **10,** e1004509 (2014).

15. Goentoro, L. & Kirschner, M. W. Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling. *Molecular cell* **36,** 872–884 (2009).

16. Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proceedings of the National Academy of Sciences* **93,** 8455–8459 (1996).

17. Schohl, A. & Fagotto, F. β-catenin, MAPK and Smad signaling during early Xenopus development. *Development* **129,** 37–52 (2002).

18. Kao, K. R. & Elinson, R. P. Dorsalization of mesoderm induction by lithium. *Developmental biology* **132,** 81–90 (1989).

19. Lemaire, P., Garrett, N. & Gurdon, J. Expression cloning of Siamois, a Xenopus homeobox gene expressed in dorsal-vegetal cells of blastulae and able to induce a complete secondary axis. *Cell* **81,** 85–94 (1995).

20. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A β-catenin/XTcf-3 complex binds to thesiamois promoter to regulate dorsal axis specification in Xenopus. *Genes & development* **11,** 2359–2370 (1997).

21. Fan, M. J., Grüning, W., Walz, G. & Sokol, S. Y. Wnt signaling and transcriptional control of Siamois in Xenopus embryos. *Proceedings of the National Academy of Sciences* **95,** 5626–5631 (1998).

22. Larabell, C. A. *et al.* Establishment of the dorso-ventral axis in Xenopus embryos is presaged by early asymmetries in β-catenin that are modulated by the Wnt signaling pathway. *The Journal of cell biology* **136,** 1123–1136 (1997).

23. Tao, Q. *et al.* Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos. *Cell* **120,** 857–871 (2005).

24. Korinek, V. *et al.* Constitutive transcriptional activation by a β-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275,** 1784–1787 (1997).

25. Arnold, S. J. *et al.* Brachyury is a target gene of the Wnt/β-catenin signaling pathway. *Mechanisms of development* **91,** 249–258 (2000).

26. Suzuki, A. *et al.* Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. *Nature Reviews Cardiology* **3,** S114 (2006).

27. Ten Berge, D. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nature cell biology* **13,** 1070 (2011).

28. Schuijers, J., Mokry, M., Hatzis, P., Cuppen, E. & Clevers, H. Wnt-induced transcriptional activation is exclusively mediated by TCF/LEF. *The EMBO journal* **33,** 146–156 (2014).

29. Sive, H. L., Grainger, R. M. & Harland, R. M. *Early development of Xenopus laevis: a laboratory manual* (CSHL Press, 2000).

30. Murray, A. W. in *Methods in cell biology* 581–605 (Elsevier, 1991).

31. Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A. & Corey, D. R. RNAi factors are present and active in human cell nuclei. *Cell reports* **6,** 211–221 (2014).

32. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (Feb. 2013).

33. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15,** 116–124 (Apr. 2005).

34. Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feed-forward loop can provide fold-change detection in gene regulation. *Molecular cell* **36,** 894–899 (2009).

35. Cadigan, K. M. & Waterman, M. L. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb Perspect Biol* **4** (Nov. 2012).

36. Alon, U. *Introduction to System Biology: design principles of biological circuits, second edition* (CRC PRESS, 2019).

37. Niida, A. *et al.* DKK1, a negative regulator of Wnt signaling, is a target of the β-catenin/TCF pathway. *Oncogene* **23,** 8520 (2004).

38. Kaplan, S., Bren, A., Dekel, E. & Alon, U. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular systems biology* **4,** 203 (2008).

39. Lee, E., Salic, A., Krüger, R., Heinrich, R. & Kirschner, M. W. The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS biology* **1,** e10 (2003).

Chapter 4

CONCLUSIONS

This work contributes to the growing understanding of how signaling information is encoded by dynamics in second messengers, and decoded during transcription mediated by the second messengers. Chapter 2 characterizes the property of memory in β-catenin and Smad3, investigating its timescales and degrees of tunability at a single cell level. Chapter 3 uncovers a novel transcriptional regulation in the Wnt pathway, enabled by DNA motif element which recruit second messenger β-catenin for repressive effect to the target gene, which allow complexity of outputs in decoding β-catenin dynamics. The findings laid out in this work reflect how signaling pathways have evolved in complexity to accommodate cellular decision making in an astounding number of contexts.

Advancements in methods and understanding in biology allow us to investigate cellular processes in finer resolution, higher fidelity, and larger scale. In this work, we've utilized Crispr-cas9 genome editing to homozygously integrate fluorescent proteins to β-catenin and Smad3 in clonal cell lines, allowing us to track endogenous levels of the second messengers at a single cell level. Previously, observing variability of second messengers were limited to snapshots at fixed timepoints, or at a limited capacity - using either exogenously expressed second messengers or heterozygous reporters. Capturing the total dynamics of second messengers gave us a finer scope to investigate these regulatory proteins- in this study investigating the temporal aspects of variability. Utility in development of the cell lines such as ours is not limited to this work: there is potential for assays to directly correlate second messenger dynamics to gene output- exemplified by work of Chris Frick, ligand-receptor combinations to second messenger dynamics, or effects of pathway crosstalk.

Memory of second messengers β-catenin and Smad3, in mESC and C2C12 respectively, were demonstrated to be at a shorter timescale, with rank autocorrelation decaying to half ('mixing time') within a cell cycle or just beyond. In comparison with those of human housekeeping genes, the memory of these second messengers is at the lowest extreme. Moreover, variability of signaling proteins including β-catenin and Smad3 is very high, ranging from 26% to 52% of the mean. Given

previous findings demonstrating that variability correlates with mixing time, the variability and memory of β-catenin and Smad3 are quite exceptional. For signaling second messengers, short memory and high variability seem to be at odds with their roles in transducing external signal through their dynamics.

Although we have yet to elucidate how the short memory in second messenger dynamics impacts transcriptional output and cellular response, we can make educated guesses as to how short memory could confer utility in signaling. There is ample evidence and theory demonstrating that noise and cell-to-cell variability confers utility in cell biology. One such example, relevant to one of the cell lines we investigated, is in transient state switching in stem cell niche. Pluripotency factor Nanog is known to fluctuate and exhibit heterogeneity in both mouth embryo and mESc culture, affecting individual cell's propensity for differentiation [1, 2]. The fluctuations and cell-to-cell variability of Nanog can be modulated [3], and one of the factors modulating Nanog is β-catenin in canonical Wnt signaling [4, 5]. Indeed, in human ES cells, variability in the level of Wnt activity is shown to affect propensity for differentiation into three germ layers [6]. A remarkable characteristic in this heterogeneity is the transience states: up until the point of cell fate commitment, the sub-states of cells generated by the variability is reversible [3, 6]. We hypothesize that memory of β-catenin could be intimately linked with this dynamic heterogeneity in the stem cell niche, allowing transitions in variable cell states in defined, short timeframe.

In addition to utility in transient state switching, short memory in second messengers could allow short refractory periods of signaling states at the population level. Use of signaling pathways is constant and ubiquitous in the lifetime of a multicellular organism, and especially during development. In this light, the capacity to be continuously reused could be a nontrivial trait for signaling pathways. Refractory periods in signaling pathways have previously been characterized in desensitization of signal receptors after ligand stimulation in pathways such as Tgf-β and NF-κB pathways [7, 8]. Memory in second messengers could further dictate how quickly signaling pathways could rebound in their capacity to transduce necessary information for decision making. For example, in cases where newly differentiated subpopulation of cells after signaling event require a successive signal for further branching of cell fates, the short memory could enable second messengers to 'reset', and population of cells could regain capacity for dynamic response.

Our findings also demonstrate that memory of β-catenin and Smad3 are highly tun-

able, changing with signaling events and cellular contexts. Between Wnt and Tgf-β pathways, how memory is modulated have significant differences, reflecting the disparate mechanisms of regulation of the second messengers. Further understanding the tunability of memory in second messengers could serve well for therapeutic purposes. Specifically for cancer, it is known that hetrogeneity in second messengers in tumor cells confer varying degrees of resistance to treatment [9, 10]. Characterization of memory could allow insights on timescale of treatment, but tuning the memory to shorten periods of resistance could also be part of the treatment.

In chapter 3, we uncovered a novel 11-bp DNA motif which recruit β-catenin/Tcf complex for transcriptional regulation. Unlike the canonical Wnt-responsive element (WRE), recruitment of β-catenin and Tcf to this novel motif lead to suppression of the target gene. The finding reveals that Tcfs are more promiscuous in DNA binding, and is consistent with previous studies in Drosophila that have identified alternative negative regulatory motifs [11]. As the motif is enriched in β-catenin binding sites in human genome as well as target genes in mice and *Xenopus*, it presents a new consideration in how we interpret promoter context in decoding β-catenin dynamics.

Perhaps the most interesting aspect in the discovery of the 11-bp NRE is the fact that it works in conjunction, but antagonistically, with the canonical WRE on the target gene, sharing the same input as β-catenin. This generates an incoherent feedforward loop (IFFL) and we demonstrate this interaction in the regulation of Siamois in Xenopus and Brachyury in mouse ES cells, and find colocalization of the WRE and NRE in many other targets. The IFFL enables complexity in decoding of second messenger dynamics in generating complex outputs from a monotonic input. Previous studies have shown several useful output functions that are possible with the IFFL: pulse response [12, 13], response acceleration [12, 14], and fold-change detection [15]. Our own mathematical simulations confirm the aforementioned functions, but also demonstrate possibilities for net repression, oscillatory, and bistable output, as well as band pass filtering. In particular, fold change detection and band pass filtering suggest potential mechanism to observations of how target genes achieve robust patterns of activation despite pathway perturbations [16].

While simulations show decoding complexities possible with the IFFL, future work is necessary to gain understanding of how the circuit may operate in different contexts. First, we have not yet elucidated mechanism of how the recruitment β-catenin to the 11bp NRE element leads to suppression of the target gene. General

mechanism of β-catenin mediated repression is not well understood, but case studies illuminate diverse strategies that could shed light on the mechanism of 11-bp NRE. β-catenin/TCFs could recruit Repressive factors such as Reptins/Pontins or Fhit could interact with β-catenin/TCFs to antagonize transcriptional activation [17, 18], mediate repression by Snail [19], or physically block sites of TF recruitment [20]. Elucidation of the β-catenin dependent suppression at 11-bp NRE could bring important insights such as timescale or strength of the suppression, and elucidate output functions possible with the IFFL. In a related context, we have observed that the 11-bp NREs do not have stringent sequence requirements, but rather follow a pattern of G/C cap followed by A/Ts. The loose requirement, and results from EMSA of candidate 11-bp NREs, strongly suggest that the element may have varying degrees of affinities for β-catenin/TCFs depending on the sequence. Since oUtput functions of the IFFL could change drastically by relative binding affinities between 11-bp NRE and the WRE, this variability could be an important factor in determining how the IFFL operates to decode second messenger dynamics.

The work from this thesis reveal new insights in encoding and decoding of information in signaling pathways. Second messengers β-catenin and Smad3 are found to have short timescales of memory, which could be useful in maintaining transient states of heterogeneity and regaining capacity for transduction after signaling events. Despite high variability and short memory of the second messengers, the pathways could still produce robust signaling outcomes with linear transduction of signal [21]. In transcription regulation of Wnt target genes that display signal response robust to perturbations, we find an incoherent feedforward loop utilizing novel 11-bp NRE. The pervasive presence of the 11-bp NRE and colocalization with canonical WRE suggest broad use of the IFFL in decoding β-catenin dynamics and generating complex output that include fold change detection and band pass filtering. These findings motivate continued endeavors to elucidate complexities at each layer of information transfer in signaling pathways, for more comprehensive understanding that may prove useful in how we approach health and therapy.

4.1 Bibliography

1. Dietrich, J.-E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134,** 4219–4231 (2007).
2. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230 (2007).
3. Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Molecular cell* **55,** 319–331 (2014).
4. Takao, Y., Yokota, T. & Koide, H. β-Catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. *Biochemical and biophysical research communications* **353,** 699–705 (2007).
5. Yi, F. *et al.* Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nature cell biology* **13,** 762 (2011).
6. Blauwkamp, T. A., Nigam, S., Ardehali, R., Weissman, I. L. & Nusse, R. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nature communications* **3,** 1070 (2012).
7. Vizán, P. *et al.* Controlling long-term signaling: receptor dynamics determine attenuation and refractory behavior of the TGF-β pathway. *Sci. Signal.* **6,** ra106–ra106 (2013).
8. Moss, B. L., Gross, S., Gammon, S. T., Vinjamoori, A. & Piwnica-Worms, D. Identification of a ligand-induced transient refractory period in nuclear factor-κB signaling. *Journal of Biological Chemistry* **283,** 8687–8698 (2008).
9. Singh, D. K. *et al.* Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. *Molecular systems biology* **6,** 369 (2010).
10. Oshimori, N., Oristian, D. & Fuchs, E. TGF-β promotes heterogeneity and drug resistance in squamous cell carcinoma. *Cell* **160,** 963–976 (2015).
11. Zhang, C. U., Blauwkamp, T. A., Burby, P. E. & Cadigan, K. M. Wnt-mediated repression via bipartite DNA recognition by TCF in the Drosophila hematopoietic system. *PLoS genetics* **10,** e1004509 (2014).
12. Mangan, S., Itzkovitz, S., Zaslaver, A. & Alon, U. The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli. *Journal of molecular biology* **356,** 1073–1081 (2006).
13. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular systems biology* **7,** 519 (2011).
14. Alon, U. *Introduction to System Biology: design principles of biological circuits, second edition* (CRC PRESS, 2019).

15. Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feed-forward loop can provide fold-change detection in gene regulation. *Molecular cell* **36,** 894–899 (2009).

16. Goentoro, L. & Kirschner, M. W. Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling. *Molecular cell* **36,** 872–884 (2009).

17. Bauer, A. *et al.* Pontin52 and Reptin52 function as antagonistic regulators of β-catenin signalling activity. *The EMBO journal* **19,** 6121–6130 (2000).

18. Weiske, J., Albring, K. F. & Huber, O. The tumor suppressor Fhit acts as a repressor of β-catenin transcriptional activity. *Proceedings of the National Academy of Sciences* **104,** 20344–20349 (2007).

19. Jamora, C., DasGupta, R., Kocieniewski, P. & Fuchs, E. Links between signal transduction, transcription and adhesion in epithelial bud development. *Nature* **422,** 317 (2003).

20. Piepenburg, O., Vorbrüggen, G. & Jäckle, H. Drosophila segment borders result from unilateral repression of hedgehog activity by wingless signaling. *Molecular cell* **6,** 203–209 (2000).

21. Nunns, H. & Goentoro, L. Signaling pathways as linear transmitters. *eLife* **7,** e33617 (2018).