

Quantitative single cell dynamics of signaling and transcriptional response in mammalian cells

THÈSE N° 8629 (2018)

PRÉSENTÉE LE 16 NOVEMBRE 2018

À LA FACULTÉ DES SCIENCES DE LA VIE

UNITÉ DU PROF. SUTER

PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Onur TIDIN

acceptée sur proposition du jury:

Prof. M. Dal Peraro, président du jury
Prof. D. M. Suter, Prof. F. Naef, directeurs de thèse
Prof. O. Pertz, rapporteur
Dr Z. Zi, rapporteur
Prof. D. Constan, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

Acknowledgements

I firstly would like to thank David for the great opportunity to work in his lab, his patient supervision and guidance throughout the past four years of my research. I am equally thankful to my co-advisor Felix for his trust in my work and his constructive and patient supervision during my PhD. I am also indebted to them for granting me the scholarship from the program Swiss Initiative in Systems Biology (SystemsX). I feel quite privileged to have their simultaneous feedback, without which this dissertation would not have been possible.

I would like to thank my committee members, Prof. Matteo Dal Peraro, Prof. Daniel Constam, Prof. Olivier Pertz and Dr. Zhike Zi whose feedback guided me to improve the work presented here.

I am also thankful for all co-workers in both wet and dry lab for making the past years less difficult and more fun, including Elias Friman, Daniel Strebinger, Aleksandra Mandic, Andrea Alber, Cédric Deluz, Mahé Raccaud, Subashika Govindan, Damien Nicholas, Nicholas Phillips, Saeed Omid, Daniel Mauvoisin, Eric Pacquet, Cédric Gobet, Jake Yeung, Clémence Hurni, Eric Durandau, Benjamin Weger. I would also like to express my gratitude to past members of the lab, especially Jonathan Bieler, Simon Blanchoud, Benjamin Zoller, Rosamaria Cannavo, Jérôme Mermet, Kyle Gustafson, Jonathan Sobel, Jingkui Wang, Julia Cajan, Johannes Becker who made my integration into the lab easier. I have learnt much from these amazing scientists and they became very good friends of me. Thank you as well to our laboratory secretaries Laura Bischoff and Sophie Barret for providing their assistance in different administrative procedures.

In addition, I would like to thank all scientific and administrative members of SystemsX and Swiss Institute of Bioinformatics (SIB) communities who provided very useful support with a variety of educational and scientific events including workshops, PhD retreats and conferences.

Last but not least, a great thanks to my parents, my brother Ozan and his wife Turkey for their constant support throughout my life including these PhD years.

Abstract

Cells live in ever-changing environments, thereby facing a variety of dynamic environmental signals. Environmental stimuli elicit intracellular responses through signaling pathways, which converge on transcriptional activation or repression of target genes. Despite intensive research, dissecting the complex interactions between pathway components that modulate mRNA and protein production still remains a difficult task. To this end, single-cell approaches provide unique insights into intracellular processes and cell responses to environmental stimuli, otherwise inaccessible with traditional bulk studies.

Single-cell measurements have revealed that isogenic cells sharing the same environment display substantial heterogeneity in both transcript and protein level. As the initial step in gene expression, transcription is an important source of such variability in eukaryotic cells due to low number of molecules involved, transcription factor dynamics and the discrete nature of biochemical reactions. Notably, production of mRNA during transcription occurs in short periods of activity, termed as transcriptional bursts, followed by longer periods of inactivity. Despite widespread observations of transcriptional dynamics in mammals, upstream molecular mechanisms shaping the eukaryotic transcription remained elusive. In this study, we quantitatively linked upstream factors and transcriptional kinetics by developing an experimental system to temporally monitor, simultaneously, inputs (transcription factors) and outputs (target gene expression) in mammalian cells, in response to stimulus.

We established two Tet-On inducible stable cell lines each expressing a fusion protein of a luminescence (Nanoluciferase) reporter to either SMAD4 or SMAD2, the two main transcription factors downstream of the TGF- β pathway. These cell lines that also contain a short-lived firefly luciferase reporter for the expression of the target endogenous connective tissue growth factor (*ctgf*) gene allowed us to quantitatively link nuclear accumulation of SMADs upon TGF- β stimulation to the target gene expression in real-time single cell measurements. Time-lapse luminescence microscopy and image analysis with the custom developed CAST (Cell Automated Segmentation and Tracking platform) platform provided quantitative single-cell data to link the upstream transcription factor profile to its target gene activity.

The data revealed weak single-cell correlations between translocation level and the target gene expression response which suggests a mechanism in which the translocation of SMADs initiates the transcriptional response but affects the response amplitude minimally. However, SMAD4 expression levels influenced the target gene response dynamics such that cells with high SMAD4 abundance favored to respond in a more sustained and even oscillatory manner. This suggests a mechanism consisting of a dynamic interplay between the transcriptional

Abstract

activators and feedback mechanisms in TGF- β signaling. We explored further the effect of different factors such as ligand concentration, ligand type on both signaling and target gene response. Taken together, this study proposes an experimental and quantitative framework that dissect mechanisms underlying transcriptional response to stimulus.

Key words: Quantitative live-cell imaging, luminescence microscopy, fluorescence microscopy, single cell dynamics, cellular heterogeneity, signal transduction, transcription factors, transcriptional bursting

Résumé

Les cellules vivent dans des environnements en changement permanent et font face à de multiples signaux environnementaux. Ces stimuli provoquent des réponses intracellulaires au travers de voies de signalisations, qui permettent l'activation ou la répression de gènes cibles en aval. Cependant, la complexité des interactions entre les voies de signalisation et production d'ARNm ou de protéines reste peu comprise. Dans ce but, les approches sur cellules individuelles permettent de comprendre des processus cellulaires en réponses à des stimuli qui resteraient inaccessibles au niveau de la population.

Les mesures effectuées sur cellules individuelles ont révélé une importante hétérogénéité dans les niveaux de transcrits et de protéines dans des cellules isogéniques dans un même environnement. En tant que première étape dans l'expression des gènes, la transcription participe activement à cette variabilité chez les cellules eucaryotes, notamment en raison du faible nombre de molécules impliquées, de la dynamique des facteurs de transcription et de la nature discrète des réactions biochimiques. Dans cette étude, nous avons lié quantitativement les facteurs en aval et la cinétique de transcription en développant un système permettant de mesurer simultanément une cause (facteurs de transcription) et une conséquence (expression d'un gène cible) dans des cellules mammifères et en réponse à des stimuli.

Premièrement, nous avons établis un système de cellules stables contenant le système d'induction Tet-On exprimant une protéine fusion entre un reporteur luminescent (Nanoluciférase) et SMAD4 ou SMAD2, deux facteurs de transactions en aval de la voie de signalisation de TGF- β . Ces cellules contiennent également un reporter luciférase à demi-vie réduite pour le gène cible endogène "connective tissue growth factor" (*ctgf*), qui permet un lien entre l'accumulation nucléaire des SMADs suite aux stimulations de TGF- β et l'expression du gène cible en temps réel et dans des cellules individuelles. La mesure de signal luminescent par microscopie et l'analyse des images par la plateforme CAST fournissent des données quantitatives pour chaque cellule, et permettant de comprendre le lien entre la présence de facteurs de transcription et l'activité d'un gène cible.

Nous avons obtenu des mesures temporelles montrant un niveau d'hétérogénéité substantiel autant dans la translocation de SMAD que dans la réponse du gène cible *ctgf*. Dans chaque cellule, le niveau de corrélation entre la magnitude de la translocation et la réponse du gène cible suggère que la translocation de SMAD initie la réponse transcriptionnelle mais n'affecte que marginalement son amplitude. Cependant, les niveaux d'expression de SMAD4 influencent la réponse du gène cible de telle sorte que des niveaux élevés de SMAD4 favorisent une réponse soutenue ou même oscillatoire. Ceci suggère l'existence de mécanismes d'interac-

Abstract

tion dynamique entre les activateurs de la transcription, et des rétroactions au sein de la voie de signalisation de TGF- β . Nous avons également investigué l'impact de différents facteurs tels que la concentration ou les types de ligand sur la voie de signalisation et la réponse du gène cible. Nous proposons donc un cadre expérimental quantitatif permettant de disséquer les mécanismes impliqués dans la réponse transcriptionnelle à un stimulus.

Mots clefs : Imagerie quantitative de cellules vivantes, microscopie à luminescence, dynamique de cellules individuelles, hétérogénéité cellulaire, transduction de signal, facteurs de transcription, transcription en rafales

Contents

Acknowledgements	iii
Abstract (Français)	v
List of figures	xi
1 Introduction	1
1.1 Transcription-centered gene expression	1
1.1.1 Transcription in eukaryotes	3
1.1.2 Multifactorial regulation of transcription	4
1.1.3 Single-cell measurements of transcription and gene expression to investigate cellular variability	5
1.1.4 Noise in gene expression	10
1.1.5 Stochastic modeling of gene expression	12
1.1.6 Molecular mechanisms regulating transcriptional bursting	15
1.1.7 Stimulus-induced modulation of transcriptional bursting	17
1.2 TGF- β Signal Transduction	18
1.2.1 Mechanisms of regulation at receptors	18
1.2.2 Signaling through <i>SMADs</i>	20
1.2.3 Transcriptional control by the <i>SMAD</i> -mediated TGF- β signaling	23
1.2.4 Negative regulation of <i>SMAD</i> mediated TGF- β signaling	25
1.2.5 TGF- β signaling in context	26
1.2.6 Connective tissue growth factor functions and mechanisms of regulation	28
1.2.7 Single-cell dynamics of <i>SMAD</i> -mediated TGF- β response	31
2 Materials and Methods	33
2.1 Cell line generation and cell culture	33
2.1.1 Construction of lentiviral plasmid constructs	33
2.1.2 Lentiviral vector production and generation of stable cell lines	34
2.2 Time-lapse measurements	34
2.2.1 Population real-time luminescence recordings	34
2.2.2 Single-cell time lapse luminescence microscopy	35
2.3 Extraction of single cell traces	35
2.3.1 Pre-processing images	36

Contents

2.3.2	Image segmentation	36
2.3.3	Cell-tracking	40
2.3.4	Assessment of the cross-talk between Fluc and Nluc channels	41
2.4	Clustering of single cell trajectories	41
2.5	Western Blot	42
2.6	Immunofluorescence	43
3	Results	45
3.1	Simultaneous monitoring of transcription factor signaling and target gene activation in single living cells	45
3.1.1	Generation of stable cell lines for dual-luminescence imaging	45
3.1.2	Validation and characterization of dual-luminescence reporters	48
3.1.3	Single-cell imaging measurements of SMAD2/SMAD4 translocation and <i>ctgf</i> expression response in TGF- β stimulation experiments	50
3.2	Cellular heterogeneity in transcription factor translocation and the target gene response dynamics	55
3.2.1	<i>ctgf</i> dynamics decompose into two distinct classes: transient and sustained	59
3.2.2	Long-term dynamics of <i>ctgf</i> response is influenced by SMAD4, but not SMAD2, levels	62
3.3	TGF- β dose modulation on both SMAD signaling and <i>ctgf</i> response	67
3.4	Ligand-dependent translocation and the target gene response	73
3.5	Single-cell correlations	75
3.5.1	Temporal correlation of signaling activity and the expression response .	76
3.5.2	The expression response amplitude scales with the expression response duration	78
3.6	Improvements of the experimental pipeline	79
4	Discussion and Outlook	83
4.1	Dual-luminescence imaging system allows highly sensitive monitoring of protein levels over extended time periods	83
4.2	Simultaneous monitoring of SMAD mediated signaling and target gene activation in response to TGF- β revealed high variability in single cells	85
4.2.1	Higher SMAD4, but not SMAD2, levels favor sustained and oscillatory <i>ctgf</i> responses	88
4.2.2	Analog encoding of TGF- β dose information in both signaling and transcriptional response	89
4.2.3	Integrative processing of independent inputs in signaling and transcriptional response	91
4.3	Negative feedback and oscillations in TGF- β signaling	92
4.4	SMAD modulation of transcriptional regulation and bursting parameters	94
4.5	Implications for <i>ctgf</i> function	95
4.6	Conclusion and Outlook	96

A Appendix	99
A.1 Supplementary Figures	99
A.2 CAST: An automated segmentation and tracking tool for the analysis of trans- scriptional kinetics from single-cell time-lapse recordings	102
A.3 Revisions	112
Bibliography	116
Curriculum Vitae	139

List of Figures

1.1	A traditional and unified view of gene expression	2
1.2	mRNA detection in single cells using single molecule fluorescence in situ hybridization (sm-FISH).	7
1.3	MS2-GFP system for detection of RNA in living cells	8
1.4	Continuous long-term quantification of single cells is crucial to detect heterogeneous and dynamic cellular responses to stimuli.	9
1.5	Schematic for a gene switching between “on” and “off” states and expressing short-lived mRNA and proteins.	10
1.6	Random telegraph model describing the stochastic gene expression is applied to short-lived luciferase reporter system to infer underlying transcriptional bursting parameters.	13
1.7	Molecular mechanisms regulating transcriptional bursting.	16
1.8	Simplified SMAD mediated TGF- β pathway.	22
1.9	Signaling pathways involved in the regulation of <i>ctgf</i> expression.	29
2.1	Graphical representation of the pLVTRE3G-NLuc-SMAD4 vector	33
2.2	Graphical representation of the pLVTRE3G-NLuc-SMAD2 vector	34
2.3	Preprocessing images.	37
2.4	Cell segmentation overview	37
2.5	Automated object splitting process detects and corrects merged cells.	38
2.6	Segmentation validation tool.	38
2.7	Robustness assessment of nuclear segmentation	39
2.8	Robustness assessment of cytoplasmic segmentation	40
2.9	Crosstalk analysis of dual-luminescence imaging system	41
2.10	Quantification of immunofluorescence samples with the automated CellProfiler pipeline.	44
3.1	Schematic of the lentiviral gene-trap vector to generate the gene-trap cell line	45
3.2	Comparison of luminescence systems.	46
3.3	Schematic illustration of the dual-luminescent reporter system	46
3.4	Dual-luminescence detection of SMAD4 and <i>ctgf</i> in single cells	47
3.5	Schematic description of Tet-On 3G inducible system.	47

List of Figures

3.6	Doxycycline induced SMAD reporter levels and its effect on <i>ctgf</i> response upon TGF- β stimulation in population-level measurements.	48
3.7	Western blot analysis of endogenous and Nluc-tagged SMAD2/4 in the stable clonal reporter cell lines	49
3.8	Dual-luminescence reporter system control experiments	51
3.9	Single-cell quantification of dual-luminescence imaging movies in TGF- β stimulation experiments.	52
3.10	Nuclear/cytoplasmic SMAD ratio is a more precise response to ligand stimulation.	53
3.11	Validation of the reporter system using immunofluorescence and western blot.	54
3.12	Dual-luminescence system performance in single-cell TGF- β stimulation experiments	54
3.13	Signal-to-noise ratio assessment of SMAD translocation reporter measurements	55
3.14	Single cell temporal variability of SMAD4 translocation and <i>ctgf</i> expression dynamics	56
3.15	Single cell variability of SMAD4 translocation and <i>ctgf</i> expression levels	57
3.16	Single cell temporal variability of SMAD2 translocation and <i>ctgf</i> expression dynamics.	58
3.17	Single cell variability of SMAD2 translocation and the <i>ctgf</i> expression levels	58
3.18	Inspection of individual <i>ctgf</i> responses revealed two distinct profiles	60
3.19	<i>Ctgf</i> response dynamics decompose into two distinct response classes.	61
3.20	Decomposition of cell-specific <i>ctgf</i> responses into transient and sustained classes	62
3.21	Not SMAD shuttling dynamics but SMAD4 levels differ between transient and sustained <i>ctgf</i> response TGF- β stimulation.	63
3.22	Long-term dynamics of <i>ctgf</i> response is influenced by SMAD4, but not SMAD2, levels	64
3.23	Decomposition of cell-specific <i>ctgf</i> responses into transient and sustained classes	65
3.24	Distribution of <i>ctgf</i> response clusters depending on SMAD levels.	66
3.25	Comparing SMAD levels in the transient and the sustained classes in samples treated with 2 ng/ml and 10 ng/ml doxycycline.	66
3.26	Determination of the concentration dynamic range for TGF- β in population-level measurements.	68
3.27	Consecutive stimulation with high doses of TGF- β revealed refractoriness in the response.	69
3.28	Ligand dose response of TGF- β -induced SMAD4 signalling and <i>ctgf</i> response in single cells.	70
3.29	Population averaged ligand dose–response of TGF- β -induced SMAD4 signalling and the <i>ctgf</i> expression	71
3.30	Population averaged ligand dose–response of TGF- β -induced SMAD2 signalling and the <i>ctgf</i> expression	71
3.31	Heat maps of stimulation dose response of TGF- β -induced SMAD4 signalling and <i>ctgf</i> response in samples treated with a range of ligand dose.	72

3.32	Analog encoding of TGF- dose information by dose-dependent SMAD and <i>ctgf</i> responses.	73
3.33	Encoding of TGF- β concentration information in temporal profiles of SMAD signaling and <i>ctgf</i> responses.	73
3.34	Experimental observation of integrative processing of simultaneous TGF- β and BMP4 signals	74
3.35	Consecutive stimulation experiment with TGF- β (first) and BMP4 (second) stimulation	75
3.36	Single-cell features of SMAD translocation and <i>ctgf</i> response dynamics in response to TGF- β stimulation	76
3.37	Schematic representation for temporal correlation of signaling activity and the expression response	76
3.38	Temporal correlation of signaling activity and the expression response based on <i>ctgf</i> peaks	77
3.39	Temporal correlation of signaling activity and the expression response based on SMAD4 translocation peaks	77
3.40	Temporal correlation of signaling activity and the expression response based on SMAD2 translocation peaks	78
3.41	The response duration effect on <i>ctgf</i> expression response amplitude	79
3.42	Nluc-Fluc cell line for signal correction	80
3.43	Initial implementations of microfluidic chambers	81
4.1	Target gene expression response showed a higher cell-to-cell variability compared to SMAD signaling.	87
4.2	Schematics of the two different scenarios for single-cell responses to stimuli.	89
4.3	Negative feedback and oscillations in TGF- β signaling.	93
4.4	Potential scenarios for SMAD mediated modulation of transcriptional bursting.	94
A.1	Single-cell <i>ctgf</i> expression levels in samples treated with TGF- β -receptor agonist SB-431542	99
A.2	Single-cell quantification of Nluc-SMAD2 cell line in luminescence imaging.	99
A.3	<i>Ctgf</i> response dynamics decompose into two distinct classes in Nluc-SMAD2 cell line	100
A.4	Decomposition of cell-specific <i>ctgf</i> responses into transient and sustained classes in Nluc-SMAD2 cell line	100
A.5	Control experiments with non-inducible doxycycline control cell line	101
A.6	<i>ctgf</i> response dynamics decompose into distinct classes in experiments with varying doses of doxycycline stimulation.	101
A.7	TGF- β -induced SMAD signalling and <i>ctgf</i> response in Nluc-SMAD4 and Nluc-SMAD2 cell lines treated with different doses of Dox.	102
A.8	Decomposition of cell-specific <i>ctgf</i> responses into transient and sustained classes in case of high SMAD4 abundance.	102

1 Introduction

1.1 Transcription-centered gene expression

Cells live in dynamic environments in which they face a variety of environmental cues that change with time. All cells respond dynamically and differently to ever-changing environmental stimuli by modulating their gene expression through a series of events [Crick, 1970]. Cellular responses to stimuli encompass several consecutive steps. Figure 1.1 depicts gene expression from the ligand induction of transcriptional activators to the formation of a mature protein. Signal propagation from cell membrane to nucleus is initiated by the activation of transcription factors specific to the ligand. Upon activation, the transcription factor translocates into the nucleus where it binds to gene regulatory elements and, through interactions with other components of the transcriptional machinery, promote access to DNA and facilitate the recruitment of the RNA polymerase enzymes to the transcriptional start site. Subsequently, a gene is transcribed to RNA which needs to be exported from the nucleus to the cytoplasm where RNAs are translated into proteins. Therefore, together with the dynamic interplay between chromatin and several other mechanisms, transcription factors shape the response of cells to stimulus.

In the classical view, gene expression response levels are predominantly regulated by transcription factors [Spitz and Furlong, 2012]. The end point of many signal transduction pathways is the activation of transcription factors that bind to short sequence motifs found in the promoter and enhancer regions of genes. Higher eukaryotes have evolved complex mechanisms for controlling the rate of gene transcription through these proteins. Considering the diversity of physiological signals that regulate gene expression, the activities of transcription factors are subject to multiple modes of regulation. The activities of many transcription factors are context dependent and can be modulated by other regulators. One particular transcription factor can activate transcription of one gene while repressing that of another. Moreover, the regulatory sequences of most eukaryotic genes contain binding sites for multiple transcription factors, allowing each gene to respond to multiple signaling pathways and to fine-tune transcript levels. Such combinatorial and context-dependent regulation of transcription allows

Chapter 1. Introduction

higher eukaryotes to respond to a broad array of stimuli using the same factors. In addition, to respond to changes in stimulus, a cell must be able to inactivate a transcriptional activator as quickly as it is induced which further requires formation of other mechanisms to ensure stable responses. While transcriptional control is a simpler task in prokaryotes, higher eukaryotes have more complex mechanisms for transcriptional regulation. Transcription is controlled by a variety of transcription factors, cofactors and chromatin regulators [Voss and Hager, 2014, Reiter et al., 2017]. Information flow from DNA to RNA (transcription) and from RNA to protein (translation) are also spatially separated in eukaryotes by a nuclear membrane that encompasses the nucleus: transcription occurs in the nucleus.

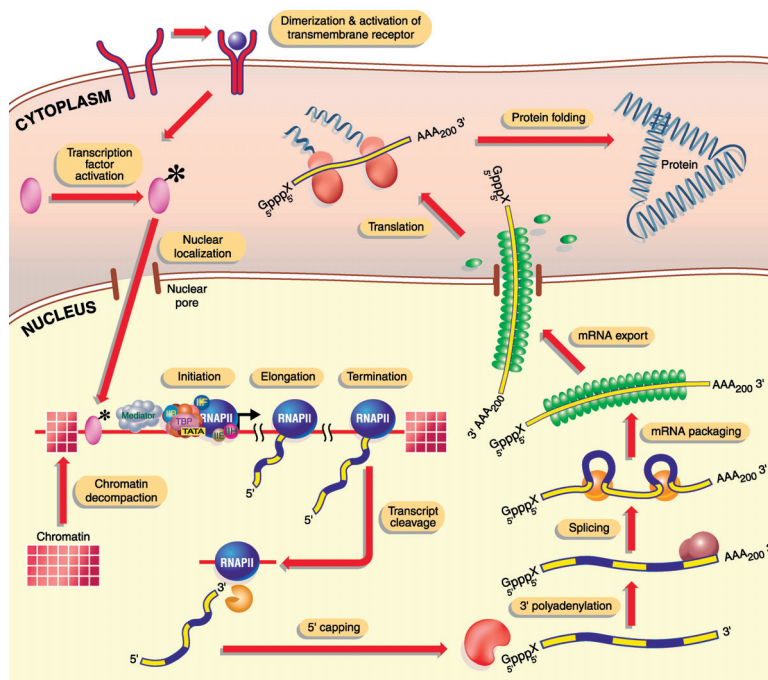


Figure 1.1 – A traditional and unified view of gene expression. Modified from [Orphanides and Reinberg, 2002].

The complexity of each of the steps in the pathway from activation to protein has required that they be studied initially in isolation using classical biochemistry. From these earlier studies, we know that consecutive steps in the pathway are interdependent or are influenced by one another. Over the past two decades, emerging single-cell methods also revealed that gene expression exhibits substantial cell-to-cell variability depending on the gene and the organism. Individual cells even in isogenic populations exhibit a unique response due to inherent stochasticity of biochemical processes, the variability of each component in the signal transduction pathway, the phenotypic variability of cells [Elowitz et al., 2002, Raj and van Oudenaarden, 2008], or cell-cycle phase [Zopf et al., 2013, McDavid et al., 2014]. Moreover, the dynamic interactions within a pathway in which two or more transcription factors co-regulate the same target gene can tune the cellular response further [Spitz and Furlong, 2012]. Such variations among cells and the dynamic pathway interactions carry valuable information regarding the mechanisms underlying transcriptional response. Therefore, it is essential

to decipher how these variations and upstream interactions determine the transcriptional response of a gene. The quantitative link between transcription factor dynamic profile and the activity of the regulated gene is still not clear and under investigation in a variety of systems.

1.1.1 Transcription in eukaryotes

DNA carries the hereditary information needed for the existence and development of living organisms. It contains sparse fragments of functional DNA sequences, termed as genes, that form the basic unit of genetic information. The encoded information in genes is transcribed into RNA which further serves as the template for protein synthesis. Therefore, transcription is the main regulatory stage of gene expression determining which genes are expressed and when. Moreover, mRNAs exist in much smaller quantities (typically fewer than 100 copies) than its expressed protein counterparts (10^3 to 10^6 copies) due to the sporadic nature of transcription and amplification from the translation of many protein copies from a single mRNA [Schwanhäusser et al., 2011]. Therefore, temporal fluctuations in transcribed mRNAs adds an important source of variability in gene expression. Thus, heterogeneity in protein expression levels is mainly attributed to variability in transcriptional dynamics in individual cells. Transcription itself is also highly sophisticated and tightly regulated process especially in eukaryotes. The organization of the genome into chromatin structure puts the first layer of complexity and regulation by preventing DNA from initiating spurious transcription cycles. Combinatorial molecular interactions at multiple stages of the transcription cycle adds another layer of regulation [Fuda et al., 2009].

Transcriptional initiation

Eukaryotic transcription requires sequence-specific transcription factors to bind to regulatory DNA sequences either in close proximity or distal regions of the promoter. These sequence-specific binding of the activator results in sequential recruitment of general transcription factors (GTFs) and RNA polymerase II (RNAPII) to assemble the pre-initiation complex (PIC) on the core promoter which leads to depletion of nucleosomes [Juven-Gershon et al., 2008]. PIC encompasses GTFs such as TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH which contributes to the formation of a stable PIC complex and recruitment of RNAPII at the transcription start site (TSS) [Grünberg et al., 2012]. Concurrently, activators recruit the Mediator complex which further stabilizes the complex. The formation of a stable PIC at the gene promoter may result in the escape of RNAPII from the promoter and advancing the complex to the next stage in the transcription, the elongation. Some GTFs appear to remain associated with the promoter enabling efficient reinitiation for the next transcription cycles [Sainsbury et al., 2015].

Transcriptional elongation and termination

Initiation step is followed by the promoter escape in which RNAPII loses its contacts with core promoter elements and concurrently tightens its grip on the nascent RNA [Saunders et al.]. The initial promoter escape does not guarantee productive transcription since RNAPII may transiently stop in early transcriptional elongations due to some pause inducing factors blocking its progression, termed as promoter-proximal pausing. The paused RNAPII still holds together with nascent RNA but needs further signals to progress elongation. This pause whose range vary from several minutes up to hours in some cases adds another layer of transcriptional regulation which may contribute to a synchronized activation in certain signal response genes [Adelman and Lis, 2012]. After proximal pausing release, the elongation transcription complex continues to incorporate nucleotides to extend the nascent mRNA along with co-transcriptional splicing, the mechanism of removing introns from the nascent pre-mRNA [Martin et al., 2013]. The elongation rate differs for different genes. It also changes throughout the elongation depending on some features of the regions transcribed such as GC content, histone organization and splicing. Dissociation of RNAPII from DNA followed by cleavage and polyadenylation of the 3-end of the nascent transcript marks the transcriptional termination [Kuehner et al., 2011]. Upon its fall off from the DNA template, RNAPII can still be recycled for reinitiation of a new transcription cycle either on the same or another promoter.

1.1.2 Multifactorial regulation of transcription

Transcription orchestration depends on various regulatory components and dynamic interactions of those in each stage of the transcription cycle. Firstly, the chromatin organization in the vicinity of the gene of interest plays a key role in transcription regulation. A nucleosome is a basic unit of genome organization in eukaryotes, encompassing a 147 base pairs (bp) sequence of DNA wound around eight histone protein cores, 2 copies from each of the core histones H2A, H2B, H3, and H4. The linker histone H1, forming the variable length linker regions of nucleosomes, stabilizes the structure further [Kouzarides, 2007]. In principle, this compact structure of nucleosomes in the nucleus restricts DNA accessibility preventing transcription. The compact structure hinders the binding of certain factors to DNA and PolII elongation along the gene. However, N-terminal regions, termed as histone tails, can be altered post-translationally through various histone modifications which alters the chromatin compaction [Andrews et al., 2016]. These histone modification are essential for the maintenance of a repressed or activated status of a gene. Different combinations of histone modifications and the dynamic interplay between them determine the density of the chromatin environment and its permissiveness to transcription. Methods such as ChIP-seq shed light on the influence of the histone code on the transcriptional regulation [Ho et al., 2014]. Histone modifications include acetylation, methylation, ubiquitination of lysine; methylation of arginine; and phosphorylation of serine to name a few among many others. For example, acetylation of lysine residues at histone tails neutralizes the positive charge which leads to chromatin destabilization, thereby marking transcriptional activation. Therefore, histone acetylation marks are almost always associated

1.1. Transcription-centered gene expression

with active transcription. For example, H3K27ac is found at both the active promoter and enhancer regions [Creyghton et al., 2010]. H3K9ac also contributes to the decompaction of chromatin by nucleosome depletion in the close vicinity of the transcription start site [Nishida et al., 2006]. In addition, H4 histone acetylation also contributes to the activation disrupting the linking between neighbouring nucleosomes [Shogren-Knaak et al.]. However, methylation of the lysine residues can be associated with either active or repressive states. Among them, H3K9me3 and H3K27me3 are marks of heterochromatin, referring to a transcriptionally inactive state of chromatin. H3K9me3 binds heterochromatin protein 1 (HP1) which is responsible for the formation and maintenance of heterochromatin leading to transcriptional repression [Hiragami-Hamada et al., 2016]. H3K27me3 interacts with H3K27ac in an antagonistic manner: it functions via recruiting a polycomb repressive complex (PRC) which contributes to the compaction of the chromatin [Di Croce and Helin, 2013]. H3K27me3 and H3K4me3 play key roles in embryonic stem cells by regulating proper cell differentiation. However, histone methylation can also be associated with an active state of transcription. Among them, H3K36me3 and H3K4me3 are frequently encountered in coding regions of active genes. H3K36me3 contributes to deacetylation of histones along PolII elongation to suppress intragenic transcription initiation [Carrozza et al., 2005]. H3K4me3 is frequently located either in the promoter or enhancer regions of genes. The combinatorial diversity of these modifications leads to precise adjustment of compaction and consequently transcription [Filion et al., 2010]. This complex histone code also determine accessible genomic regions for transcription factor binding. DNA regulatory sequences add another layer of regulation for transcription. These sequences can be subdivided into three classes depending on their proximity to the transcription start site: 1) the core promoter that encompasses short sequences enclosing the transcription start site. 2) the promoter-proximal region where sequence-specific transcription factors bind, either activating or repressing the transcription. 3) enhancer and silencer sequences that activate or suppress their target genes over a greater separation [Sanyal et al., 2012]. The core promoter region is responsible for the recruitment of transcription preinitiation complex (PIC) while it further needs proximal-promoter elements to initiate the transcription located in 100-200 bp proximity of the transcription start site. Enhancers also establish proximity with their target genes through DNA looping and activate transcription [Plank and Dean, 2014]. The interplay between these regulatory elements, their target promoters and the chromatin state fine-tune the transcription and consequently protein expression.

1.1.3 Single-cell measurements of transcription and gene expression to investigate cellular variability

Over the past decades, numerous techniques have been developed allowing quantitative analysis of gene expression both at mRNA and protein levels including western blots, northern blots, RNA-seq, flow cytometry and mass spectrometry. Most of these methods are performed at the population level and provide snapshots of the cellular content at a certain time point. Therefore, these population-averaged methods obscure details of cellular dynamics and variability since they require measurements on a large number of cells [Larson et al., 2009]. Yet,

the copy numbers of mRNAs and proteins change over time and vary from cell to cell even in isogenic cell populations. Single-cell approaches have recently provided valuable quantitative information to investigate cellular variability at the single-cell level with high temporal and spatial resolution [Larson, 2011]. This quantitative view of single-cell gene expression is based on counting mRNAs in single cells, monitoring transcription in real-time, and monitoring gene expression with fluorescent protein-based technologies. Here I will review these experimental approaches that have become essential in understanding the rules governing gene expression and transcriptional kinetics.

Fluorescence in situ hybridization (FISH) is a method developed to detect and localize the presence or absence of specific DNA sequences on chromosomes. It uses fluorescently labeled nucleic acid probes that bind to specific parts of the chromosome. The technique can also be utilized to detect and localize specific RNA targets. The RNA-FISH method originally used limited number of probes harboring multiple fluorophores [Femino et al., 1998, Levsky and Singer, 2003]. With recent advances, the method typically uses 20 to 50 probes of 18-20 nucleotide length for the same transcript. The use of higher number of colocalized probes, each harboring a single fluorophore, resulted in a diffraction-limited detectable spot, which prevented the detection of false positives and thereby allowing detection of single molecule RNA (sm-FISH, Raj et al. [2006, 2008]), Figure 1.2). Therefore, the method is used for the precise and direct visualization of active transcription sites in the nucleus and of endogenous transcripts allowing to distinguish between mRNA and nascent pre-mRNA in cells. The method was initially limited to measure only a few genes in parallel [Levsky and Singer, 2003]. Further improvements based on combinatorial labeling extended this approach to the simultaneous measurement of a larger number of RNA species [Jakt et al., 2013, Levsky et al., 2002, Lubeck and Cai, 2012]. A modified approach called iceFISH (intron chromosomal expression FISH), generated transcriptional profiles of 20 genes simultaneously by using distinguishable combinations of fluorophores, along a single chromosome in single cells [Levesque and Raj, 2013]. Similarly, bDNA (branched DNA) approach further enabled large-scale multivariate profiling RNA transcript abundance using tree-like structure of probes [Battich et al., 2013]. More recent methods ultimately achieved transcriptome scale multiplexing using barcoding approach through successive rounds of hybridization and imaging on the same sample [Chen et al., 2015, Moffitt and Zhuang, 2016, Eng et al., 2017]. smRNA-FISH has been applied to different organisms including E.coli [Taniguchi et al., 2010, So et al., 2011, Chong et al., 2014], yeast [Mao et al., 2010, Zopf et al., 2013, Neuert et al., 2013], and mammals [Raj et al., 2006, Brody et al., 2011, Senecal et al., 2014, Singer et al., 2014, Battich et al., 2015]. In addition to cultured cells, the method was successfully applied to tissue slices [Itzkovitz et al., 2011, Bahar Halpern et al., 2015] and thin organisms [Lécuyer et al., 2008, Raj et al., 2008]. Overall, smRNA-FISH allows the quantification of both the copy number of individual transcripts and the spatial distribution of these molecules within single cells. The transcriptional bursting parameters of the telegraph model can be estimated by fitting measurements of mRNA copy number at a single-cell level [Dey et al., 2015]. This approach, providing rapid estimates of transcriptional kinetics, contributes to our understanding of transcription. Yet, these dynamic

1.1. Transcription-centered gene expression

properties are usually inferred from a population distribution at a single time point, assuming each cell is part of a homogeneous population. smFISH is still a static method since it is performed on fixed samples and it does not allow real-time measurements of transcriptional fluctuations. Thus, it is essential to utilize dynamic methods to monitor and quantify how transcription evolves in time.

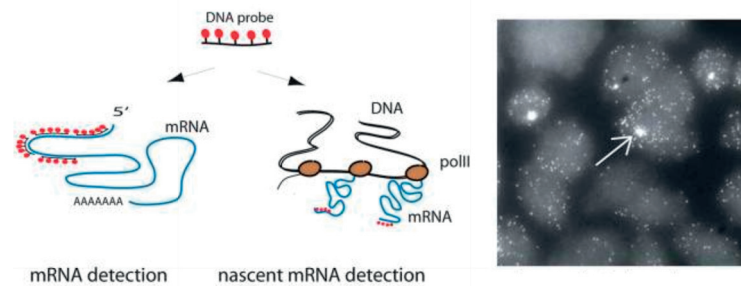


Figure 1.2 – mRNA detection in single cells using single molecule fluorescence in situ hybridization (*sm-FISH*). Schematic representation of *smFISH* method (left side). The binding of fluorescently labeled synthetic oligonucleotides allows visualization of mRNAs including nascent mRNAs at the transcription site. Active transcription sites, shown by the arrow, create brighter spots due to synthesis of multiple RNAs (right). Modified from [Larson et al., 2009].

MS2/PP7 techniques allowed visualization of mRNA in real time in living cells for the first time. The method is based on the high binding affinity of MS2/PP7 capsid proteins to RNA stem loop structure (Figure 1.3). The system has two parts: The first part constitutively expresses MS2 or PP7 capsid protein labeled with a fluorescent protein. The second part includes a DNA cassette of sequence repeats coding for RNA stem loop in the untranslated regions of the reporter mRNA. As soon as the reporter mRNA is transcribed, MS2/PP7-fluorescent fusion protein binds to these stem-loop sequences [Bertrand et al., 1998]. This system allows visualization of the transcription spot directly and consequently monitoring of transcription kinetics in living cells. Real-time imaging of transcriptional activity in single cells provide the possibility to develop models describing the stochastic transcription. The method has been applied to many organisms such as *E. coli* [Golding et al., 2005], yeast [Larson et al., 2011], *dictyostelium* [Chubb et al., 2006] and mammals [Darzacq et al., 2007, Brody et al., 2011, Martin et al., 2013, Yunger et al., 2010, Ochiai et al., 2015]. The technique has also been applied to live imaging of transcription in *Drosophila* embryos [Bothma et al., 2014]. Moreover, these systems have been shown to allow simultaneous measurements of two gene loci in living cells [Hocine et al., 2013]. The method is suitable for time-lapse recording with high-temporal resolution (5-60 sec); however, the duration of the recording is limited by photobleaching and phototoxicity. The system has also the disadvantage that mRNA stability and trafficking may deviate from its normal profile due to insertion of stem-loops while modifications of the system improved this issue [Tantale et al., 2016]. Moreover, the establishment of the system requires an initial titration of each expression construct to achieve an optimal ratio of MS2-GFP stem loop to RNA expression [Boireau et al., 2007, Lionnet et al., 2011, Yunger et al., 2013, Suter et al., 2011, Yunger et al., 2010]. Precise detection of RNA signal localization is also a challenging task due to strong fluorescent background sourced from accumulated

Chapter 1. Introduction

fluorescent-coat protein.

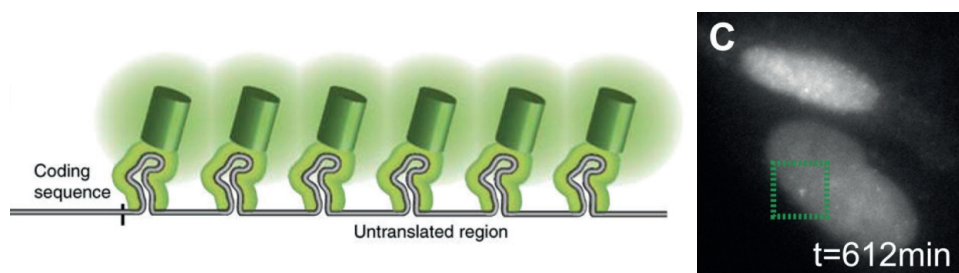


Figure 1.3 – MS2-GFP system for detection of RNA in living cells. Schematic depiction of MS2-GFP system (left), MS2 coat protein binds to MS2 stem loops inserted into 3-untranslated region of mRNA of interest. MS2 system long-term imaging allowing the visualization of transcription site as the bright spot. Modified from [Tyagi, 2009, Larson et al., 2013].

Single-cell real-time imaging of gene expression reporters provides an alternative tool to quantify the temporal evolution of protein expression in a given cell. Live-cell microscopy combined with fluorescent protein based technologies has been used to quantify gene expression and its variability in single cells [Warmflash et al., 2012, Aymoz et al., 2016, Strasen et al., 2018]. Fusing the fluorescent protein to a target protein or inserting the fluorescent protein under the control of a promoter of interest allowed monitoring of expression dynamics and variability over time. Despite its widespread use in live-cell imaging, fluorescent based technologies have some limitations. One limitation is that fluorophores lose their capacity to fluoresce over time due to photobleaching. Moreover, cells are prone to phototoxic effect of generated reactive chemical species due to repeated exposure of fluorescently labeled cells to illumination from lasers and high-intensity lamps. In addition, a fluorescent protein has to go through several stages to become functional, termed as maturation. This maturation time, varying from several minutes to more than an hour, restrains use of fluorescent imaging in some signalling pathways where rapid changes are investigated [Miyawaki et al., 2003, Shaner et al., 2005]. Finally, autofluorescence sets a limit for sensitivity which would not allow detection of fewer than several thousands of proteins per cell [Newman et al., 2011]. It is possible to overcome most of aforementioned limitation by using luminescence based imaging methods which utilize an enzymatic reaction of luciferases to produce light reaching much higher sensitivity without any phototoxicity effect for extended periods of time [Suter et al., 2011, Mazo-Vargas et al., 2014]. Furthermore, luminescence imaging shows the potential of dual-imaging by combining different luminescence systems [Hall et al., 2012, England et al., 2016]. One drawback of luminescence imaging is that it needs long exposure times up to a few minutes due to lower amounts of light produced, which may limit the temporal and spatial resolution depending on expression levels.

Deciphering the molecular mechanisms of cellular decisions, how they are dynamically regulated based on environmental cues and how they manifest phenotypically, is a central topic in biology research and is important for clinical applications. Yet, heterogeneity is

1.1. Transcription-centered gene expression

pronounced, even in clonal and closely related cell populations regarding cellular dynamics. Especially, in the pathway signaling research, such heterogeneity has crucial importance to understand why and how cells respond differently. Population level measurements, such as western blots provide the discrete measurements of many cells which may hide the true variability of the response. Single-cell snapshot measurements such as sm-FISH and flow cytometry, on the other hand, reveal the heterogeneity of cell populations displaying different activation states in single cells. Yet, these methods lack the temporal resolution which may result in the misinterpretation with different conclusions from the same experimental measures (Figure 1.4). Only dynamic single-cell measurements reveal the true dynamics and the underlying heterogeneity of the response to stimuli. Therefore, single-cell measurements are crucial to gain insight into complex interconnections of signaling pathways consisting of a network of positive and negative feedback or feedforward loops [Skylaki et al., 2016, Wosika and Pelet, 2017]

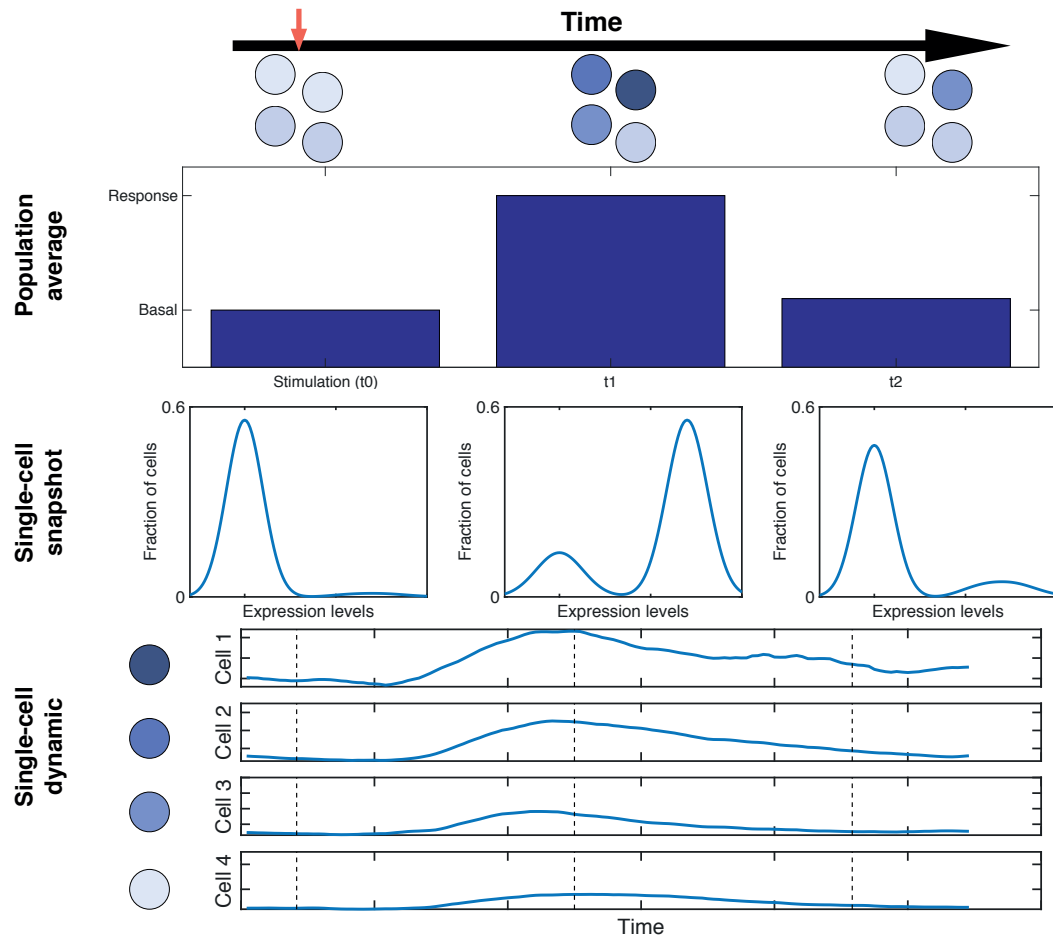


Figure 1.4 – Continuous long-term quantification of single cells is crucial to detect heterogeneous and dynamic cellular responses to stimuli. Population average (top) or single-cell snapshot (middle) analyses lack the required resolution for reliable inference of cellular and molecular dynamics. Only dynamic single-cell measurements reveal the true dynamics and the underlying heterogeneity of the response to stimuli (bottom).

Chapter 1. Introduction

One important modification that can be utilized in single-cell live microscopy based on either fluorescent or luminescent proteins is to destabilize reporters in an attempt to decipher transcriptional activity. This would allow more precise assessment of transcriptional dynamics since longer half-lives significantly buffer the noise in transcriptional kinetics. Indeed, destabilized fluorescent [Bean et al., 2006, Zhang et al., 2013] and luminescent reporters provides a powerful technique to investigate transcriptional kinetic parameters [Suter et al., 2011, Molina et al., 2013]. In such systems, promoter activation leads to an intermittent signal increase in the reporter signal. When combined with stochastic modeling, the transcriptional state of the gene can be inferred from measurements using the telegraph model of stochastic transcription (Figure 1.5). The transcriptional bursting parameters can be inferred by calculating the likelihood of moving between successive time points in the time series given the telegraph parameters [Bronstein et al., 2015]. Live-cell imaging methods allowed us to gain more mechanistic insights about transcriptional dynamics which would not be possible to quantify in static methods performed on fixed cells.

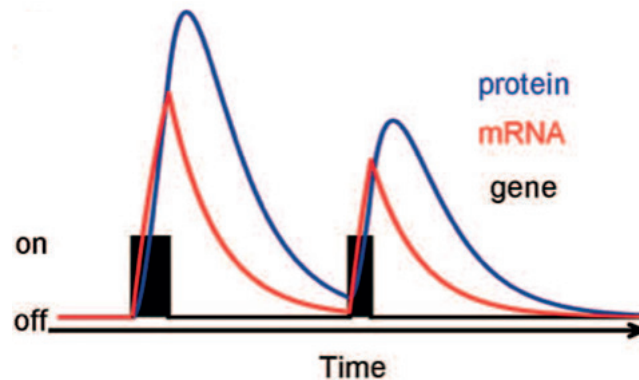


Figure 1.5 – Schematic for a gene switching between “on” and “off” states and expressing short-lived mRNA and proteins. The gene switches between the active “on” and inactive “off” state (black). mRNA is produced during “on” state (red) followed by short-lived protein production (blue). Using protein level measurements, the underlying transcriptional bursting parameters and the state of the gene can be inferred. Adapted from [Suter et al., 2011].

1.1.4 Noise in gene expression

Over the last two decades, emerging single-cell methods improved our understanding of gene expression in individual cells and revealed that protein expression is a dynamic process that varies in time and between individual cells. Cells even in isogenic populations exhibit substantial heterogeneity in protein expression and factors contributing to this variability are still not clear and under investigation. In principle, this heterogeneity may arise partly from changes in the cellular state due to fluctuations in cellular components while ever-changing cellular microenvironment adds another layer of variability.

To decipher sources of cellular component fluctuations, initial studies used two fluorescent protein reporters expressed separately under two alleles of the identical promoters

1.1. Transcription-centered gene expression

[Elowitz et al., 2002]. Owing to this design, the extrinsic noise influenced both expression concurrently while intrinsic noise influenced each copy individually. This system showed the contribution of both intrinsic and extrinsic sources of variability on gene expression and revealed that both contributed substantially to overall variation. Extrinsic factors include cell-to-cell differences in the cellular state including cell-cycle phase, cell size, stress from intrinsic sources, or fluctuations in cellular components. On the other hand, intrinsic noise is the result of discrete nature of basic cellular processes, the low number of components involved, the random partitioning of molecules at cellular division and the bursty nature of eukaryotic transcription [Suter et al., 2011, Singh and Soltani, 2013]. Finally, microenvironment parameters including cell density, number of cell-to-cell contact or the existence of environmental cues vary substantially between individual cells. Therefore, it is essential to decipher how intrinsic stochastic fluctuations, cellular states and microenvironment determines heterogeneity in single cell responses.

Transcription emerges as the central dynamic process contributing to noise in gene expression. Temporal fluctuations in lowly abundant mRNAs resulting from sporadic transcriptional activities, termed as bursts, form an important source of noise in gene expression [Golding et al., 2005, Chubb et al., 2006, Raj et al., 2006, Suter et al., 2011]. These bursts represent short and intense periods of transcriptional activity followed by longer periods of gene inactivity. The discrete nature of biochemical reactions involving low abundance of molecules, random diffusion processes of TFs to find their targets and even fluctuations of transcriptional reactants lead to stochasticity in transcription. These factors combined with slow transitions between promoter states result in mRNA heterogeneity measured in mammalian systems. This sporadic mRNA production is subsequently further amplified by translation. Therefore, in individual cells, mRNAs usually exist in low amounts (< 100 copies) while their protein counterparts is much more abundant (10^3 - 10^6 copies) due to the amplification effect of translation [Schwanhäusser et al., 2011]. Yet, this amplification effect on protein levels and long protein lifetimes may average out fluctuations in protein levels. In a system with a short mRNA half-life coupled to a long protein half-life, the noise in mRNA profile does not reflect into protein levels which may prevent observations of corresponding protein bursts. Destabilized reporters of short-lived mRNA and protein allowed measurements of transcriptional bursting on the protein level. Although some recent studies revealed stochastic events occurring during the ribosomal scanning process and posttranscriptional noise generation in *E.coli* [Dacheux et al., 2017], the noise of gene expression in mammalian systems predominantly depends on promoter activity and transcription.

Noise in gene expression influences all other processes including genes, RNAs and proteins which leads to propagation and summation of such noise in one gene to other downstream genes in complex networks [Pedraza and van Oudenaarden, 2005]. Fluctuations in the expression of transcription factors may contribute to the noise further due to their involvement in transcription. Therefore, even a network whose components have low noise, may have substantial noise depending on gene interactions. Notably, cells develop certain mechanisms such as negative feedback loops which potentially limit the propagation of noise

[Lestas et al., 2010]. In addition, the formation of multimeric complexes [Konkoli, 2010] or sequestration of promoters of the downstream genes within compact chromatin has also been shown to prevent transmission of the transcriptional noise further [Shah and Tyagi, 2013]. While noise in gene expression has been associated with undesirable phenotype in certain contexts, it may have evolved to have certain benefits providing better survival strategy than synchronized cellular response. On one hand, synchronized cellular responses provided by control mechanisms is beneficial in multicellular organisms to assure the proper functioning in a uniform manner. Yet, some studies revealed that heterogeneity in cell populations can be beneficial depending on the pathway, the input type and the tissue [Loewer and Lahav, 2011]. In the NF κ B pathway, for instance, a dual-negative feedback system contributes to the heterogeneity in NF κ B oscillations which in return minimizes paracrine signalling fluctuations of the tissue [Paszek et al., 2010]. Similarly in E.coli, the persistence of individuals to stress such as antibiotic treatment was associated with preexisting variability based on a phenotypic switch [Balaban et al., 2004]. Also in the p53 response pathway, the variability in sensitivity of individuals to intrinsic stress in the basal state has been shown to select for healthy strong cells for survival [Lahav et al., 2004]. Thus, in some cases high noise creating a broad variety of phenotypes can be an evolutionary survival strategy while minimal noise is favorable in some others.

1.1.5 Stochastic modeling of gene expression

Transcription is an important source of noise in gene expression due to its irregular pulsatile pattern leading to fluctuations of mRNAs over time. The simplistic model of such stochastic nature of transcription is the birth-death process. In this model, the number of mRNA molecules can change either due to stochastic production or degradation, according to a Poisson process. This model involves transcript production with a constant rate (k_m) and a constant degradation rate per molecule (γ_m). The model leads to a Poisson distribution for the number of mRNA molecules at steady state. Yet, the noise in transcription measurements, characterized by dispersion of mRNA numbers in measurements, deviates from Poissonian statistics showing a more widespread distribution especially in higher eukaryotes. Therefore, the simple birth-death process is not able to reflect and explain the high levels of fluctuations in mRNA levels. A modified model, termed as telegraph model, hypothesizes that high levels of fluctuations are due to sporadic activations of the gene. In this model, the promoter switches stochastically between an active (on-state) and inactive state (off-state) with constant switching rates. Transcripts are produced only in the active state while degradation occurs continuously. In the telegraph model, the degradation process occurs with constant degradation rate per molecule (γ_m) while mRNA production is modeled using three parameters: two switching rate from off-state to on-state or the opposite) and the average transcription rate during on-state. This modified version of the simple birth-death model can fit to other distributions observed during measurements. If on states are very short, the mRNA distribution exhibits a higher super-Poissonian variance [Shahrezaei and Swain, 2008]. When promoter transition rates are low meaning that cells stay longer in one particular state, then mRNA

1.1. Transcription-centered gene expression

distribution exhibits bimodality. Thus, the telegraph model provides sufficient plasticity to model possible measured RNA distributions corresponding to different dynamic scenarios of promoter activation. In the regime assuming short off-states and mRNA lifetime while large burst size (transcripts per burst), measurements of mRNA copy numbers can be used to estimate bursting parameters such as normalized burst frequency (events per transcript life time) and burst size. This provides a powerful framework to obtain rapid estimates of bursting parameters using smRNA-FISH measurements.

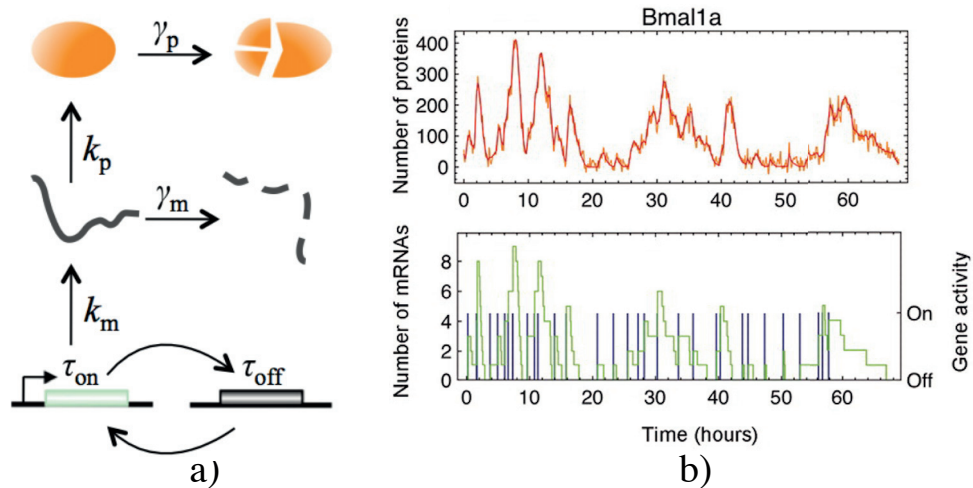


Figure 1.6 – Random telegraph model describing the stochastic gene expression is applied to short-lived luciferase reporter system to infer underlying transcriptional bursting parameters. a) The gene promoter switches between active (on) and inactive (off) states stochastically with rates k_{on} and k_{off} . Transcription and translation occurs with rates k_m and k_p , respectively. mRNA and protein degradation are modeled as Poissonian processes with rates γ_m and γ_p , respectively. b) Real-time monitoring of the short-lived luciferase reporter expressed under Bmal1 promoter. Because the reporter is destabilized both at mRNA and protein level, the underlying mRNA profile and activation state of the gene can be inferred using telegraph model. Modified from [Suter et al., 2011]

In addition, the telegraph model can be utilized to infer bursting parameters from live-cell imaging measurements [Suter et al., 2011], Figure 1.6. In this case, a second birth-death process can be inserted to represent the translation event (Figure 1.6a). Proteins are stochastically produced with a rate proportionally to the amount of existing mRNAs (mk_p) and degraded with a rate proportionally to the amount of existing proteins ($p\gamma_p$) according to a Poisson process. Overall, k_m and mk_p represents the transcription and translation rate, respectively while γ_m and γ_p describe constant degradation rates for transcription and translation. k_m represents the effective transcription rate including processes such as synthesis, splicing and nuclear export while k_p describes the effective translation rate per mRNA encompassing events such as translation initiation, peptide elongation and termination. This minimal model is sufficient to represent measurements on both mRNA and protein levels despite its simplistic approach combining several processes in effective rate constants. The telegraph model is utilized to estimate bursting parameters, otherwise not easy to measure directly, using data collected by several experimental methods. For example, these parameters

can be estimated in smRNA-FISH measurements of mRNA copy number for each single cell, by explicitly calculating the likelihood of the observed data for given parameters [Raj et al., 2006, Senecal et al., 2014, Bahar Halpern et al., 2015]. Instead of the transcripts, the telegraph gene expression model can be applied to short-lived fluorescent and luminescent reporter measurements. Suter et al. established various gene trap cell lines and transgenic cell lines expressing a short-lived luciferase protein from an unstable mRNA which allowed the telegraph model to infer underlying transcriptional parameters using real-time bioluminescence data. The analysis revealed highly gene-specific bursting kinetics and also refractory periods representing inactive gene states before turned on again.

Short-lived luciferase offers a solution to eliminate the buffering effect of long protein half-lives [Suter et al., 2011]. With long protein lifetime, protein levels represent transcriptional activity integrated over certain amount of time which is undesirable to resolve transcriptional bursts. In this system, the degradation rate of protein (γ_p) and mRNA (γ_m) were estimated using deterministic model defined by Eq 1.1. Inhibiting transcription ($k_{eff}=0$, k_{eff} : effective transcription rate) and translation ($k_p=0$) in separate population level experiments allowed to estimate degradation rates. In addition, calibrations of photon emission to different luciferase concentrations allowed the quantification of number of luciferase molecules per cell in single-cell imaging experiments. The calibration of mRNA molecules is also conducted using real-time quantitative polymerase chain reaction. These calibrations were incorporated into the model to precisely describe the data. The other four parameters of the telegraph model (k_{on} , k_{off} , k_m , k_p) were estimated by computing the probability of the time traces using maximum likelihood approach.

$$\begin{aligned}\dot{p} &= k_p m(t) - \gamma_p p(t) \\ \dot{m} &= k_{eff} - \gamma_m m(t)\end{aligned}\tag{1.1}$$

In the telegraph model described, the state of the system at a certain point is determined by three integers p, m, g which represent the number of proteins p (any integer), the number of mRNA molecules (any integer), and the state of the gene g (0 or 1). The model is defined as Markovian. Therefore, the model has *memoryless* property meaning that at any time point t , the state of the system depends only on the state at the previous time point. Eq 1.2 describes the transition probability of the state p, m, g at time t . Eq 1.2 represents an intractable differential equation which can be solved under different approximations [Shahrezaei and

Swain, 2008, Iyer-Biswas et al., 2009].

$$\begin{aligned}
 \frac{P_{p,m,g}}{dt} = & k_p m P_{p-1,m,g} + \gamma_p (p+1) P_{p+1,m,g} \\
 & + k_m g P_{p,m-1,g} + \gamma_m (m+1) P_{p,m+1,g} \\
 & - k_p m P_{p,m,g} - \gamma_p p P_{p,m,g} \\
 & - k_m g P_{p,m,g} - \gamma_m m P_{p,m,g} \\
 & + k_g P_{p,m,\bar{g}} - k_{\bar{g}} P_{p,m,g}
 \end{aligned} \tag{1.2}$$

1.1.6 Molecular mechanisms regulating transcriptional bursting

Aforementioned emerging methods and analysis approaches have together revealed the bursty nature of transcription in different organisms from prokaryotes to yeasts and higher eukaryotes . Despite its rare occurrence in bacteria due to formation of positive supercoiled DNA following the passaging of RNA polymerase [Ma and Wang, 2016, Chong et al., 2014], higher level of organisms exhibit transcriptional bursting as a quasi-universal behaviour. These studies revealed the substantial difference of bursting kinetics among different genes which can frequently be characterized by using a combination of two parameters: 1) the burst frequency referring the number of bursts per unit time 2) mean number of transcripts produced along one burst period. In mammalian systems, burst frequency typically varies from a burst every 30 minutes to 10 hours while the burst size from one to several hundreds of transcripts [Lionnet and Singer, 2012]. Such diversity in bursting profiles among genes can be attributed to the their complex regulation and mechanisms involved in shaping the gene expression on the transcription level[Nicolas et al., 2017]. Despite remarkable progress in experimental techniques to monitor transcription at the single-cell level and the analytical tools such as new variants of the telegraph model, molecular mechanisms involved in regulation and their influence on transcriptional bursting is still elusive and under investigation [Molina et al., 2013, Lenstra et al., 2016].

Nucleosome occupancy and histone mark have been shown to shape the gene expression in yeast [Brown and Boeger, 2014, Tirosh and Barkai, 2008]. PH05 gene expression fluctuations and the promoter nucleosome dynamics observed by electron microscopy revealed the coupling between nucleosome configurations and stochastic gene activity [Brown et al., 2013]. This interaction seemed to modulate bursting dynamics through changes in burst frequency. Another study used DNase hypersensitivity assays in mammalian cells to assess the effect of chromatin accessibility on the gene expression noise [Dey et al., 2015]. Denser nucleosome occupancy leads to noisier gene expression while lower nucleosome occupancy around the transcription start site exhibited higher burst frequency and lower gene expression noise. Combinations of histone marks also influence the genomic organization, thereby transcriptional bursting parameters. In Dictyostelium, H3K4 methylation states inherited to

daughter cells exhibited their characteristic transcriptional profiles regarding burst frequency and duration [Muramoto et al., 2010]. The effect of acetylation on transcriptional bursting was also investigated by blocking histone deacetylation with Trichostatin A (TSA). Yet, this perturbation yield in various transcriptional bursting modulations in a highly gene specific manner [Harper et al., 2011, Dar et al., 2012, Suter et al., 2011]. Overall, nucleosome occupancy exerted its influence on transcriptional burst frequency while histone marks exhibited gene-specific alterations of transcriptional bursting parameters [Nicolas et al., 2018].

Core promoters and other DNA regulatory sequences also influence the bursting profile. Similar reporters expressed under different promoters at the same genomic locations lead to varying burst frequency profiles [Yunger et al., 2010]. In some studies, the effect of regulatory sequences is investigated by observing gene expression dynamics in altered DNA contexts [Suter et al., 2011, Bartman et al., 2016, Fukaya et al., 2016] while other studies revealed associations of certain expression profiles with the existence of specific regulatory elements [Zoller et al., 2015, Newman et al., 2006, Bengtsson et al., 2005, Stewart-Ornstein et al., 2012]. Certain binding motifs such as TATA-boxes typically exhibited high-expression noise genome-wide and correlated with the absence of refractory period [Newman et al., 2006, Blake et al., 2003, Hornung et al., 2012]. The core promoter architecture also shaped the transcriptional bursting profile. Increased number of transcription factor binding sites in inducible synthetic promoters resulted in increased burst size [Raj et al., 2006]. Similarly additional binding sites of TALE transcription activators around the TSS resulted in prolonged burst episodes [Senecal et al., 2014]. In addition to core promoter and its proximity, distal regulatory sequences such as enhancers have recently been investigated recently. Perturbation in loop formation between β -globin gene and its LCR enhancer resulted in lowered burst frequency observed with smRNA-FISH [Bartman et al., 2016]. Similar results in *Drosophila* embryos using MS2-GFP system revealed that distal genomic regions are likely to modulate burst frequency while the proximal element alterations mainly influenced the burst size [Fukaya et al., 2016].

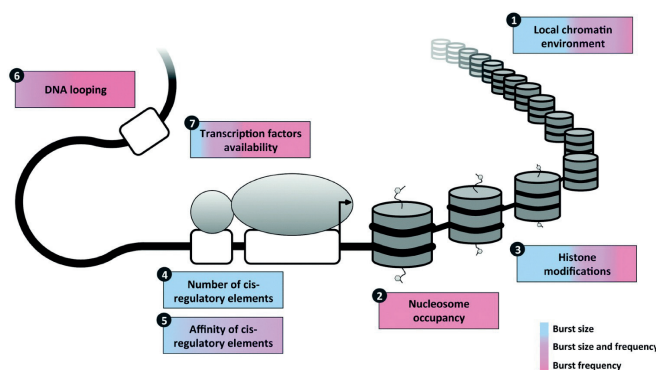


Figure 1.7 – Molecular mechanisms regulating transcriptional bursting. Schematic represents an active gene and other components participating in the regulation of bursting parameters. Color code represents the involvement likelihood of shown entities in different bursting parameter modulations (DNA-black line, the core promoter-large white boxes, TSS-arrow, a specific transcription factor-small grey sphere, the pre-initiation complex-large grey oval, nucleosomes-grey cylinders). Modified from [Nicolas et al., 2017].

1.1.7 Stimulus-induced modulation of transcriptional bursting

Cells operate in dynamic environments and respond to external changes modulating their gene expression through a series of events. External stimuli activate regulatory pathways through specific transcription factors which possibly alter transcriptional kinetics of their target genes in certain ways. Since bursting modulation has been attributed predominantly to chromatin accessibility and modifications, only a few studies attempted to link transcriptional activators to downstream bursting profile. smFISH measurements of the endogenous *c-Fos* gene upon stimulation through MAPK pathway revealed a correlation between transcriptional activator p38 and the number of active transcription sites suggesting the transcriptional burst frequency modulation [Senecal et al., 2014]. Another study showed steroid driven mRNA synthesis through dose-dependent frequency modulation of transcription [Larson et al., 2013]. In blood differentiation process, the effect of TFs on each other's transcription measured by qPCR revealed the burst frequency as the main regulated parameter [Ezer et al., 2016]. In mouse liver, upregulated genes in response to fasting conditions predominantly revealed increased burst fraction—the fraction of time a promoter is actively transcribing [Bahar Halpern et al., 2015]. These results suggest that stimulation altered mainly the bursting frequency while some other studies showed combined modulation of burst frequency and burst size in other systems. The transcriptional response of β -actin to extracellular stimuli showed alterations both in burst frequency and size in perturbed levels of key signaling factors [Kalo et al., 2015]. In Wnt pathway, MS2-GFP measurements revealed increases in burst duration, burst intensity and on-rate of transcription on a cyclinD1 reporter correlating with the accumulation rate of its activator β -catenin [Kafri et al., 2013]. The controversy that whether stimuli modulate burst frequency alone or burst size and burst frequency together is further elucidated with additional studies. Using a large panel of reporters obtained with HIV based random lentiviral insertions at 8000 loci concluded that both the frequency and size of the transcriptional bursts is modulated in response to TNF- α stimulation. Interestingly, weakly expressing loci alter only burst size while loci expressing stronger modulate also the burst size [Singh et al., 2010, Dar et al., 2012, 2016]. This observation suggests an expression threshold level determining the mode of bursting modulation which may explain different suggestions in other works. Indeed, this mechanism holds true for the case in which the endogenous mouse gene *ctgf* seemed to respond strongly to TGF- β and serum stimulation with a rapid and acute increase in burst sizes [Molina et al., 2013]. Similarly, another highly expressing doxycycline synthetic promoter exhibits modulation of average size of the bursts rather than their frequency in accordance with increased transcription factor levels [Raj et al., 2006]. In these cases, cells may have surpassed the expression threshold over which cells starts to use burst size modulation rather than burst frequency. In summary, approaches that activate these broad array of inducible systems and investigation of transcription factor modulation on target genes will contribute to our understanding of mechanisms shaping the transcriptional bursting.

1.2 TGF- β Signal Transduction

As multicellular organisms emerged, it became essential to develop a tight control of individual cells to move, divide, differentiate and organize in synchrony. This was achieved by certain intercellular communication systems. Among them, secretory polypeptides that signal through membrane receptors coupled to transcriptional regulatory networks provided an efficient way of communication. The transforming growth factor (TGF) family is one of the best examples of such communication molecules. The TGF- β superfamily of ligands include TGF- β s, bone morphogenetic proteins (BMPs), growth and differentiation factors (GDFs), activin and nodal playing roles in many biological processes such as cell proliferation, embryogenesis, lineage determination, motility, adhesion and tissue homeostasis [Massagué, 1998], [Attisano and Wrana, 2002]. Each member, expressed in certain temporal and tissue-specific patterns, has been discovered and investigated in separate lines of research due to their complex interactions with each other [Massagué, 2012]. However, the superfamily can be categorized into two groups based on sequence comparisons between the bioactive domains and sequence similarity and the specific signaling pathways that they induce. The first group includes TGF- β , Activin, Nodal while the second group includes Bone Morphogenetic Protein(BMP), Growth and Differentiation Factors(GDF), MIS (Muellerian inhibiting substance). TGF- β is the prototypical ligand of the TGF- β superfamily which we will focus more in this study.

1.2.1 Mechanisms of regulation at receptors

The mature TGF- β is a secreted cytokine that dimerizes to produce a 25 KDa dimer covalently linked by a disulphide bond between cysteine residues of two monomers. However, it is secreted as part of a latent complex (large latent complex, LLC) that localizes on the extracellular matrix. This complex includes latency associated proteins (LAP) and latent TGF- β binding protein (LTBP) [Lobe et al., 2000]. After TGF- β is synthesized, it interacts with LAP forming an initial complex which further binds to LTBP, a large glycoprotein disulfide-linked to LAP. LTBP is not essential for the latency of the TGF- β complex, however, it attaches tightly to extracellular matrix preventing the complex from binding to TGF- β receptors. Upon certain perturbations of ECM associated with processes such as angiogenesis, wound repair, inflammation, cell growth, this latent complex is released from ECM and activated by different molecules including proteases, integrins and thrombospondin (TSP). Such storage of latent TGF- β in the ECM and the activation through a diversity of activators allow for differential responses of TGF- β in vivo [Constam, 2014]. In addition to this mechanisms, some accessory receptor molecules such as betaglycan, biglycan, decorin, endoglin also regulate the bioavailability of TGF- β for receptors [Annes et al., 2003, Shi and Massagué, 2003, Heldin and Moustakas, 2016].

Liberated TGF- β and related factors signal through a family of transmembrane protein serine/threonine kinases referred to as the TGF- β receptor family. Based on their structural and functional properties, the TGF- β receptor family is divided into two subfamilies: type

I receptors and type II receptors ($T\beta$ RI and $T\beta$ RII in case of TGF- β). Both receptor types have an N-terminal extracellular ligand binding domain, a transmembrane region, and a C-terminal serine/threonine kinase domain. The type I, different from type II, receptors contain a characteristic GS domain where type II exerts its effect on. In basal state, $T\beta$ RII can form ligand-independent homo-oligomers which can prime the formation of heteromeric $T\beta$ RI/ $T\beta$ RII receptor complex upon ligand binding. Each ligand initiates signaling by binding to a specific combination of type I and type II receptors forming tetramers consisting of two type II and two type I receptors. In case of no stimulation, both type I and type II receptors are found as either dynamic monomers or dimers on the membrane. TGF- β binding to the receptor stabilizes receptor dimerization of $T\beta$ RII followed by $T\beta$ RII autophosphorylation. Consequently, $T\beta$ RII dimer transphosphorylates serine residues of $T\beta$ RI in their GS domain which induces conformational changes that activate $T\beta$ RI kinase activity [Heldin and Moustakas, 2016]. Upon this ligand-induced oligomerization, these dual specificity kinase receptors phosphorylate receptor regulated *SMADs*, *R-SMADs*.

The type of activated *SMADs* depends on the ligand since each TGF- β super-family member signals through specific sets of type I and type II receptor complexes. As stated earlier, TGF- β signals via TGF- β type II receptor ($T\beta$ RII) and $T\beta$ RI (also known as ALK5). Similarly, Activin and NODAL signals via type I receptors ACVR1B (ALK4) and ACVR1C (ALK7) coupled to the type II receptors ACTR2A and ACTR2B. In case of BMP, the receptor binding affinity depends on the type of the ligand. BMP6 and BMP7 interacts with the type II receptors ACTRII recruiting type I receptors ALK2, ALK3 and ALK6. Therefore, activins and BMPs share the type II receptor ACTRII and compete with each other when binding. In contrast, BMP2 and BMP4 prefer to bind to type I receptors ALK3 and ALK6 and only then recruit type II receptors while BMP9 and BMP10 prefers ACVRL1 (ALK1) as the type I receptor. ACVR1 (ALK2) and AMHR2 are the type I and type II receptors for AMH [Massagué, 2012, Wakefield and Hill, 2013]. This complex combinatorial pairing of receptors and overlapping receptor usage increases the diversity of downstream responses in different cellular contexts. [Shi and Massagué, 2003]. However, two main distinct branches described is an oversimplification, since other combinations of ligands and receptor interaction may happen which contributes to the plasticity of the pathway. TGF- β , for example, has been shown to signal additionally through SMAD1/5/9 via mixed ligand-receptor combinations [Goumans et al., 2003, Wu and Hill, 2009, Grönroos et al., 2012]. Finally, in addition to canonical SMAD-mediated signaling, other non-canonical pathways including MAPK and PI3K have been reported [Heldin and Moustakas, 2016]. However, my thesis is mainly based on the canonical SMAD pathway mediated TGF- β responses which will be further investigated.

The dynamic nature of TGF- β receptor activity requires additional tight control through posttranslational modifications. The kinase activity of $T\beta$ RI upon ligand-induced phosphorylation is reversible and targeted by phosphatases. The protein phosphatase PP1c dephosphorylates $T\beta$ RI with SARA, a scaffolding protein that stabilizes the interaction with the receptor. Another phosphatase molecule PP2A was shown similarly to interact with $T\beta$ RI upon stimulation [Bennett and Alpey, 2002, Shi et al., 2004, Batut et al., 2008]. Similarly, ubiquitin

modification of T β RI plays a key role in defining receptor stability and turnover. T β RI receptors can be ubiquitinated and degraded through the ubiquitin system. In general, ubiquitination is defined by sequential actions of a ubiquitin activating enzyme(E1), a ubiquitin conjugating enzyme(E2) and a ubiquitin ligase(E3) that determine the specificity of the process. An inhibitory *SMAD*, *SMAD7*, recruits E3 ubiquitin ligases such as Smurf1, Smurf2, WWP1, and NEDD4-2 which target receptors for proteasomal degradation [Lönn et al., 2009, Lo and Massagué, 1999, Zhang et al., 2001, Boeck and Dijke, 2012, Lin et al., 2000]. Conversely, deubiquitinating enzymes (DUBs) such as USP4, USP11 and USP15 interact with T β RI through *SMAD7* and remove poly-ubiquitin chains resulting in the stabilization of the receptor. On the other hand, no dephosphorylation or ubiquitination enzymes of T β RII have been demonstrated. Taken together, the balance of the aforementioned reversible posttranslational modifications shapes the response on the receptor level [Xu et al., 2012, Zhang et al., 2014].

Internalization of activated surface receptors provides a further regulation mechanism for TGF- β signal transduction [Chen, 2009, Meyer et al., 2011]. This receptor-mediated endocytosis can modulate the signaling positively or negatively. It can suppress the signalling by reducing the number of receptor molecules on the surface or by mediating the receptor degradation. On the other hand, it can recycle receptors back to the cell surface for reuse. It has been shown that activated receptor complex internalization can occur in two ways: clathrin-mediated and non-clathrin mediated (caveola-mediated) [Penheiter et al., 2002, Le Roy and Wrana, 2005]. In the absence of ligand, TGF- β receptors are constitutively endocytosed and recycled back to the cellular membrane. In case of clathrin-mediated endocytosis, upon ligand binding, the activated receptor-*RSMAD* complex enters early endosomes where they are brought into contact with signaling-promoting molecules such as SARA, Hrs, endofin which allow for phosphorylation of *RSMADs* and their association with *SMAD4* [Runyan et al., 2005, Hayes et al., 2002]. In this respect, early endosomes may operate as a TGF- β signal promoting organelle. Moreover, some of ligand-bound complexes in early endosomes are sorted to late endosomes for either degradation together with TGF- β or for separation from bound TGF- β and recycling to the cell surface. The caveola/lipid raft mechanism, on the other hand, was shown to promote receptor turnover and signal termination since the inhibitory *SMAD7* together with Smurf proteins targets activated receptor complex for ubiquitination and eventually for proteasomal degradation [Di Guglielmo et al., 2003]. Therefore, the distribution of activated receptors in between two compartment-specific mechanism adds another layer of regulation in TGF- β signaling [Zhang et al., 2005].

1.2.2 Signaling through *SMADs*

SMADs include a family of structurally similar proteins that transduce signals coming through receptors of TGF- β superfamily ligands to target genes in the nucleus. Located downstream of the receptor activation mechanism and just upstream of the transcriptional activity, they can provide valuable information regarding transcriptional regulatory mechanisms. They play key roles in many different cellular contexts such as cell proliferation, embryogenesis, lineage

determination, motility, adhesion and tissue homeostasis. The first identified member of the *SMAD* family is the Mad (mothers against dpp) protein in *Drosophila*. Based on structural and functional similarities, *SMADs* can be categorized into three classes: **1)** *SMADs* that are involved in direct signaling from the TGF- β receptor (receptor regulated *SMADs*, *R-SMADs*, *SMAD1/2/3/5/8* **2)** *SMAD* that has the role of partnering with *R-SMADs* to recruit co-regulators to the complex (co-*SMAD*, *SMAD4*), **3)** *SMADs* that suppress the signalling activity of the other two groups (inhibitory *SMADs*, *SMAD6* and *SMAD7*. Among *R-SMADs*, *SMAD1*, *SMAD5* and *SMAD8* mediates BMP signaling through the receptor BMPRI while *SMAD2* and *SMAD3* are substrates of T β RI transducing TGF- β and activin signals. However, the common mediating partner for all *R-SMADs* is *SMAD4* [Massagué et al., 2005, Hill, 2009, Warmflash et al., 2012].

SMADs are ~500 aminoacids long proteins that have similar structural characteristics [Macias et al., 2015]. They consists of two relatively well-conserved globular domains (N-terminal MH1 and C-terminal MH2) separated by a quite divergent linker region. MH1 domain, conserved in all *R-SMADs* and *SMAD4* but not in inhibitory *SMADs*, has DNA-binding activity established by β -hairpin structure. Therefore, MH1 domain determines the DNA-binding affinity and specificity of *SMAD* transcriptional complexes to the particular target gene. MH2 domain, a highly-conserved domain as well, functions as a very versatile region for many protein interactions. It allows for homo-oligomeric complex formation of all three subgroups of *SMADs*. *R-SMADs* interact in their MH2 domain with activated receptors through a Ser-X-Ser motif which gets phosphorylated and associates with *SMAD4* to form further transcriptional complexes. MH2 domain also mediates interactions with cytoplasmic retention proteins, nuclear pore complex (nucleoporins), transcriptional activators and repressors. The linker region contains MAPK, CDK and other protein kinase phosphorylation sites in *R-SMADs* allowing interactions with other pathways and an NES (nuclear export signal) in *SMAD4* which regulates the nuclear translocation [Massagué, 1998, Attisano and Wrana, 2002, Massagué et al., 2005, Derynck and Zhang, 2003, Zhang, 2009].

In the basal state, *SMADs* exist as homo-oligomers mainly residing in the cytoplasm. *SMAD4*, for example, is constitutively imported into the nucleus via nuclear localization signal (NLS) in its MH1 domain and also exported back to cytoplasm via CRM1-mediated nuclear export mechanism associated with its linker region. Therefore, the distribution between nucleus and cytoplasm is determined by the relative strengths of the nuclear import and export signals. When the nuclear export is inhibited with CRM-1 export mechanism inhibitor LeptomycinB treatment, *SMAD4* strongly localizes into nucleus which points to an active export mechanism [Pierreux et al., 2000]. However, it has been shown that this nuclear *SMAD4* in unstimulated cells is not transcriptionally active. Moreover, *SMAD4* and also *SMAD3* have been shown to translocate into nucleus via direct interactions with the nuclear pore complex. Similarly, *R-SMADs* (*SMAD2* and *SMAD3* are also predominantly localized in the cytoplasm in the basal state while this localization is based on some anchor and adaptor molecules that contributes to their retention in the cytoplasm. The protein SARA (*SMAD* anchor for receptor activation) binds to *SMAD2/3* in a region on the MH2 domain which is also involved in the interaction with nucleoporins. This configuration prevents *SMAD2/3* from translocating into

nucleus in the absence of ligand [Massagué et al., 2005, Clarke et al., 2006].

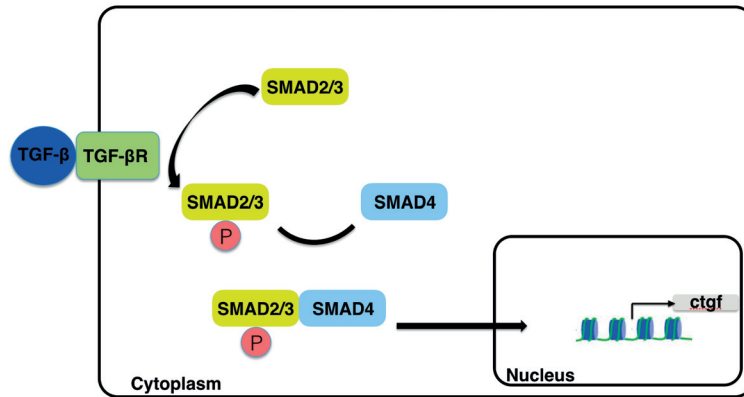


Figure 1.8 – Simplified SMAD mediated TGF- β activation pathway.

Upon ligand binding, the association of *SMAD4* with the receptor-phosphorylated *SMAD2* and *SMAD3* masks the NES in *SMAD4* resulting in the nuclear accumulation of *R-SMAD-SMAD4* complexes. In addition, receptor-activation of *SMAD2/3* decreases the binding affinity of *SMAD2* for SARA which frees *SMAD2/3* from cytoplasmic retention resulting in their interaction with nucleoporins and constant shuttling, providing a dynamic pool that is competitively drawn by cytoplasmic and nuclear signal transduction partners including *SMAD4* [Xu et al., 2002]. *SMAD4* and *R-SMAD* complex formation can happen both in cytoplasm or nucleus while the latter causes nuclear trapping of activated *SMAD4-SMAD2/3* complexes [Nicolás et al., 2004]. Activated SMAD oligomers of diverse stoichiometric compositions, both heterodimer *R-SMAD-SMAD4* and *SMAD2-SMAD3-SMAD4* trimer forms, were proposed and observed to accumulate into nucleus, in response to stimulation [Zieba et al., 2012].

In principle, the nuclear accumulation of *SMADs* despite constant shuttling can be explained by two mechanisms, either an increase in import rate or a decrease in export rate. Using photoactivatable fluorescent proteins (PAFPs) and fluorescence recovery after photobleaching (FRAP) approach, it has been shown that ligand-induced increased affinity for DNA-binding is the main mechanism resulting in a drop in export rates without affecting import rates [Schmierer and Hill, 2005]. TGF- β treatment decreases the nuclear *SMAD* mobility most likely due to increased DNA interaction with activated *SMAD* complexes. However, the difficulty of *SMAD* complexes being recognized by nuclear export mechanisms as explained in case of *SMAD4* can be another explanation for altered nuclear retention. Moreover, the incapability of complexed *SMAD2* to interact with cytoplasmic anchor proteins contributes to the nuclear localization. Therefore, subsequent dephosphorylation of *R-SMADs* can drive their return to the cytoplasm for another cycle of receptor-mediated phosphorylation and nuclear accumulation. The phosphatase activity in the nucleus causes the dissociation of *SMAD2/SMAD4* complexes resulting their monomers to be exported from the nucleus independently as explained previously [Lin et al., 2006]. In summary, the predominant cytoplasmic localization of *SMAD4* and *R-SMADs* without ligand treatment can be explained by continuous

nuclear import and export mechanisms, having faster export rates than the import. Upon ligand binding, nucleus functions as a trap for activated *SMAD* complexes filtering out the inactivated portion and contributing to the formation of increased DNA-bound transcriptionally active complexes [Massagué et al., 2005, Clarke et al., 2006].

1.2.3 Transcriptional control by the SMAD-mediated TGF- β signaling

Activated *SMAD* complexes interact directly with defined DNA sequences via their MH1 domain or indirectly via other transcription factors to regulate gene expression negatively or positively. It has been shown that *SMAD1*, *SMAD3* and *SMAD4* specifically recognize a DNA sequence motif 5'-CAGAC-3' known as *SMAD*-binding element (SBE) [Dennler et al., 1998, Ross and Hill, 2008b]. However, instead of a single SBE providing minimal affinity, direct or inverse palindromic repeats of SBEs are recognized in the regulatory regions of TGF- β target genes. In contrast to other *R-SMADs*, the full length *SMAD2* does not bind DNA due to an insert encoded by exon 3 upstream of the DNA-binding β -hairpin. The *SMAD2* isoform that lacks exon3, on the other hand, shows binding affinity equivalent to other *SMADs* [Yagi et al., 1999]. The MH1 domain can also interact with GC-rich palindrome motifs (GGCGCC) as efficiently as SBE [BabuRajendran et al., 2010]. Despite such consensus sequences, the binding affinity of *SMADs* for DNA is not strong [Shi et al., 1998]. These short DNA-binding sites of *SMADs* necessitate the need for cooperation with additional molecules for required versatility of the pathway.

Different from the majority of transcription factors, the activated *SMADs* need the interaction with chromatin to recruit the general transcription machinery. In vitro studies revealed that *SMAD* complexes only activate transcription from chromatin templates, and not at all from naked DNA [Ross et al., 2006]. Therefore, *SMAD* transcriptional activity relies on the interaction with other factors together with epigenetic landscapes and chromatin-modifying enzymes. It is now more evident that different programmes of gene expression are based on the context and cell type dependent interplay between these interactions. In this section, I will summarize *SMAD* transcription regulation mechanisms reported so far and emphasize questions still unanswered.

Due to limited DNA binding affinity and specificity of *SMADs*, the interaction with partner TFs plays a key role in shaping the transcriptional responses to TGF- β signaling. Genome-wide studies of *SMAD* binding across different cell types revealed *SMAD* co-occupation of the genome at different target genes with lineage-specific master TFs [Chan and Kyba, 2013]. High-affinity and high-specificity binding of *SMADs* to DNA is achieved by recruitment of additional transcription factors since *SMAD* affinity alone for DNA is weak. *SMADs* interact with a vast diversity of other transcriptional factors via their MH1 and MH2 domains [Feng and Derynck, 2005, Ross and Hill, 2008a]. These interacting transcriptional factors, in most cases, are involved in other signaling pathways which contributes to the cell context dependence of the target gene transcriptional activities. They are essential to recruit *SMAD* complexes to DNA

in some cases. For example, in response to activin, receptor activated *SMAD2-SMAD4* complex interacts with FoxH1 (a forkhead transcription factor), the first transcriptional cofactor reported to cooperate with *SMADs* [Chen et al., 1996]. In this complex, FoxH1 associate with *SMAD2* via its MH2 domain while *SMAD4* binds DNA. *SMAD2*-FoxH1 interaction, using the same domain as SARA, excludes *SMAD2*-SARA association which contributes nuclear accumulation of the complex. In this case, FoxH1 is essentially required to recruit *SMAD2-SMAD4* complexes to its own promoters [Germain et al., 2000, Randall et al., 2004]. Similarly, some other transcription factors have been shown recruit *SMAD3-SMAD4* complexes to DNA. In other cases, *SMADs* also co-operate with other transcription factors at TGF- β dependent promoters. For that, transcription factors bind DNA in regions next to SBE repeats and co-operate with *SMADs* to activate transcription. For example, FoxO3 protein has been shown to bind the distal region of the p21 promoter at a conserved forkhead-binding element next to SBE repeats which concurrently associate with *SMAD3-SMAD4* complexes [Seoane et al., 2004]. These diverse set of transcription factors and mechanisms guide *SMADs* to specific subsets of target genes and provide pleiotropic capacity of TGF- β signaling.

SMADs also interact with certain coactivators (chromatin readers, modifiers and remodelers) and corepressors resulting in chromatin remodeling. Both *R-SMADs* and *SMAD4* can bind molecules such as p300 and CBP with histone acetyltransferase (HAT) activity. Upon TGF- β stimulation, *SMADs* associate with CBP/p300. These proteins have been shown necessary for *SMAD* mediated transcription and its specificity [Ross et al., 2006, Janknecht et al., 1998, Feng et al., 1998]. In addition to HATs, *SMAD2* and *SMAD3* binds BRG1 (also known as SMARCA4) which is a subunit of SWI/SNF chromatin remodeling complexes. SWI/SNF complexes modulate gene expression by using ATP energy to mediate nucleosome sliding. In a TGF- β dependent manner, *SMADs* interact with SWI/SNF complexes involved differentially in the transcriptional response of many target genes of this cytokine [Ross et al., 2006, Xi et al., 2008]. Some studies also demonstrated signal induced DNA demethylation mechanism in which *SMAD* transcriptional complex recruits removes DNA methylation of a TGF- β target gene via recruiting a base excision repair complex [Thillainadesan et al., 2012, Laugesen and Helin, 2014]. On the other hand, *SMAD* complexes can also achieve transcriptional repression by triggering repressive histone marks and reversing histone activating modifications. In certain target genes, *SMADs* recruit corepressors such as SnoN, Ski, TGIF and histone deacetylases (HDACs) into the transcription machinery which removes activating histone acetylations in a TGF- β dependent manner [Liu et al., 2001, Wotton et al., 1999, Kang et al., 2005a]. Moreover, some genes may form synexpression groups to achieve a complex and balanced response to a TGF family stimulus [Gomis et al., 2006]. These observations reveal mechanisms of how TGF signaling through the *SMAD* pathway translates complex inputs to a variety of concrete responses.

1.2.4 Negative regulation of SMAD mediated TGF- β signaling

Due to its key importance in diverse biological processes, TGF- β signaling is spatiotemporally fine-tuned. Activated complexes are subject to ubiquitin-mediated proteasomal degradation. Inhibitory SMADs (*SMAD6/7*) play a key role in this respect. TGF- β signaling ensures a stable response by targeting also inhibitory SMAD genes (*SMAD7* in case of TGF- β , *SMAD6* in case of BMP) [Xu et al., 2012, Itoh and ten Dijke, 2007]. I-SMADs lack the MH1 domain preventing their binding to the DNA [Aragón et al., 2012].

These inhibitory SMADs, in turn, repress the response at multiple stages of the pathway [Wegner et al., 2012]. *SMAD7*, as the key negative regulator, exerts its inhibitory effects through different mechanisms both at the receptor level and in the nucleus. It has been first demonstrated that *SMAD7* inhibits signaling by binding to TGF- β type I receptors (T β RI) and preventing R-SMAD activation [Kavsak et al., 2000]. It was also shown that *SMAD7* regulates the T β RI activity by recruiting different molecules to the receptor. *SMAD7* associates with both Smurf1 and Smurf2, E3 ubiquitin-protein ligases, which in turn induces the cytoplasmic localization of the complex. This inhibitory complex is recruited to TGF- β type I receptors (T β RI) and targets receptors for ubiquitin-mediated proteolysis [Ebisawa et al., 2001]. In addition, *SMAD7* recruits a phosphatase complex to phosphorylated TGF- β type I receptors resulting in their dephosphorylation and inactivation [Shi et al., 2004]. Inhibitory SMADs also interfere with the transcriptional by disrupting SMAD-DNA complex via recruiting repressors [Itoh and ten Dijke, 2007, Moustakas and Heldin, 2009]. Contrarily, *SMAD7* also associates with deubiquitinase enzymes to reverse the ubiquitination of receptors potentiating the signaling [Wicks et al., 2000].

TGF- β signaling is subject to tight regulation due to negative feedback also through R-SMADs and co-SMAD, *SMAD4*. Upon signaling, Smurf1/2, being transcriptional targets of TGF- β cytokines, show enhanced association with phosphorylated *SMAD2/3* targeting it for ubiquitination and subsequent degradation [Lin et al., 2000]. Proto-oncoprotein Ski and SnoN are also important negative regulators of the pathway since they interact with *SMAD2/3/4* preventing activated R-SMAD-*SMAD4* formation [Stroschein et al., 1999, Wegner et al., 2012]. Similarly, some negative feedback mechanisms provide control on *SMAD4* activity upon ligand induction. Transcriptionally induced *SMAD7* binds phosphorylated *SMAD2/3* preventing *SMAD2/3-SMAD4* association [Yan et al., 2009]. *SMAD4* association with *SMAD2* and I-SMADs recruits different complexes causing mono- and polyubiquitylation via different molecules and subsequent degradation [Dupont et al., 2009]. In addition, negative regulators Ski and SnoN can participate in formation of histone deacetylation complexes which shows that some negative feedback mechanisms target the pathway in the nucleus at the chromatin level [Deheuninck and Luo, 2009]. Another E3 ubiquitin ligase Arkadia has also been shown to interact with *SMAD2/3* and *SMAD7* inducing their turnover. Arkadia perform this only on inactivated *SMAD2/3* in the nucleus which couples degradation of *SMAD2/3* with the target gene transcription [Mavrakis et al., 2007, Levy et al., 2007]. Moreover, it prevents the inhibitory activity of *SMAD7* via ubiquitin-dependent degradation of *SMAD7* [Koinuma et al., 2003].

Therefore, the feedback provided by Arkadia counterintuitively amplifies the transcriptional response induced by TGF- β superfamily proteins.

Different studies identifying a vast variety of feedback mechanisms changed our understanding of the pathway as a linear arrangement of components. The complex interplay between activation and negative feedback contributes to the versatility of the pathway. Moreover, the balance between activation and negative regulation causes heterogeneous responses in single cells such that some cells respond to the stimulus in a transient way while some respond either in a sustained or oscillatory fashion [Strasen et al., 2018]. Therefore, it needs further investigation to determine factors that contribute the response variability. Single-cell approaches may reveal underlying mechanisms [Warmflash et al., 2012]. However, it remains a big challenge considering the versatile dynamics of the pathway and the contextual nature of TGF- β /SMAD pathway.

1.2.5 TGF- β signaling in context

Recent progress in the TGF- β field has contributed to our understanding of molecular mechanisms in the pathway to a first approximation. However, the influence of TGF- β depends heavily on the cellular context. TGF- β signaling mediates a broad array of medically important cellular processes such as immunity, inflammation, cancer, fibrosis and also muscle, bone, adipose, vascular, and hematopoietic homeostasis [Leask and Abraham, 2004]. The influence of TGF- β signaling, depending on the cell type and conditions, can vary to a great extent and can even be opposite. It has been shown that TGF- β can suppress cell proliferation but also induce cell growth, contribute to stem cell pluripotency but also differentiation, promote metastatic cells and also suppress pre-malignant cells. These contradictory functions point to the contextual functionality of the pathway [Massagué, 2012].

ES cell self-renewal and differentiation: Pluripotency in embryonic stem (ES) cells is maintained by the core pluripotency complex including the core transcriptional regulators OCT4, SOX2 and NANOG [Young, 2011]. The core transcription factors together positively regulate their own promoters while repressing lineage-specific transcription factors. The core pluripotency triad maintain repressive marks on differentiation genes via chromatin-modifying complexes preventing the escape from the pluripotent state. Some studies showed that SMAD proteins play also a key role in this state. BMP-induced SMAD1 co-occupy genomic sites marked active by H3K4me3 together with OCT4, SOX2, NANOG, and STAT3 [Ying et al.]. In principle, Nodal signaling drives mesendodermal differentiation. However, in the pluripotent state, OCT4 drives Nodal induced SMAD3 to some Nodal negative feedback sites which keeps cells in a non-responsive state to differentiation signals. Moreover, differentiation genes targeted by Nodal signaling are marked by H3K9me3 which keeps them in a repressed but a poised state via HP1 γ (heterochromatin protein 1 γ) [Azzaz et al., 2014]. However, in ES cells lacking self-renewal signals, Nodal signaling results in TRIM33-SMAD2/3 and SMAD4-SMAD2/3 associations. TRIM33 recognizes H3K9me3 histone marks, displaces HP1 γ and gives

access to activin responsive elements by *SMAD4-SMAD2/3* complexes [Massagué Joan and Xi Qiaoran, 2012]. In the pluripotent state BMP activated *SMAD1* contributes to the maintenance of the state but in absence of self-renewal signals, Nodal activated *SMAD* complexes induce differentiation genes. Furthermore, lineage identity factors target *SMAD* proteins to different subsets of genes which activates special differentiation programs. These mechanisms demonstrate the importance of *SMAD* mediated TGF- β signaling in ES cell pluripotency, differentiation and lineage regulation.

Epithelial-mesenchymal transition (EMT): EMT is the process in which epithelial cells lose their epithelial phenotypes such as cell polarity, cell-cell adhesion and gain a mesenchymal invasive phenotype. EMT underlies many developmental processes such as mesoderm formation, neural tube formation while pathological types of EMT cause fibrosis and cancer. EMT is fundamentally characterized by loss of cell-cell adhesion protein E-cadherin. A transcriptional network including *SNAIL1*, *SNAIL2*, *ZEB1*, *KLF4* *TCF3* and *TWIST* suppresses the E-cadherin promoter, *CDH1* [Thiery et al., 2009]. Together with WNT signaling, TGF- β -induced *SMADs* trigger EMT in epithelial cells through this mechanism. TGF- β induced *SMADs* activate *SNAIL1* and *TWIST1* suppressor expression while WNT effector *LEF1* associating with *SMAD1/4* similarly joins the suppression complex of *CDH1* promoter. BMP signaling, on the other hand, antagonizes EMT by activating pro-epithelial microRNAs *miR-200* and *miR-205* which favour mesenchymal-epithelial transition (MET) [Brabletz and Brabletz, 2010].

Tumor suppression and tumor progression: Context-dependent activity of TGF- β signaling plays a key role in tissue regeneration and homeostasis [Yang et al., 2010]. In principle, TGF- β as a strong tumor suppressor exerts its cytostatic effects in the context of normal epithelial cells. In case of a pre-malignant cell, TGF- β alters its suppressive effects even yielding apoptosis. However, it has been shown that TGF- β switches to a tumor enhancer in more advanced carcinoma cells. In some cancer cells, mutations inactivating *T β RII* and *SMAD4* functionality result in selection of malignant clones not responsive to suppressive effect of TGF- β . This insensitivity to TGF- β can create a TGF- β rich environment which can favor more tumor progression. On the other hand, some cancer cells still own TGF- β induced *SMAD4* activity but lack responses downstream of *SMAD* signaling. The genetic loss of tumor suppressive responsiveness result in further inductions of metastatic effects. This change of TGF- β functionality from tumor suppressor to tumor enhancer again points to the context-dependent nature of TGF- β signaling [Massagué, 2008].

TGF- β signaling and the fibrotic response: Fibrosis is the formation of excess fibrous connective tissue in an organ or tissue in a reparative or reactive process. Thus, the term is used to describe both the process of connective tissue deposition in healing and pathological state of excess deposition of fibrous tissue. After injury, the synthesis of the new connective tissue is crucial. During this process, fibroblasts proliferate and migrate into the wound and synthesize elevated levels of matrix proteins such as collagen and fibronectin [Leask and Abraham, 2004]. Specialized form of fibroblasts, known as myofibroblasts, express increased levels of α -smooth muscle actin (α -SMA) allowing to contract extracellular matrix which is

crucial for wound closure [Gabbiani, 2003].

Normal tissue repair is controlled by complex interactions of profibrotic and antifibrotic cytokines including TGF- β , connective tissue growth factor (*ctgf*), tumor necrosis factor- α (TNF- α), interferon- γ (IFN- γ). TGF- β , the central mediator of the fibrotic response, activates fibroblasts to synthesize and contract ECM. *Ctgf*, induced by TGF- β , acts as the downstream mediator of the effect of TGF- β on fibroblasts [Grotendorst, 1997, Leask et al., 2004]. TGF- β also induces the expression of an alternative splicing variant of the fibronectin, extra domain-A containing fibronectin which result in collagen type I and α -SMA expression [Serini et al., 1998]. Inversely, macrophages expresses the proinflammatory cytokine TNF- α which downregulates the expression of ECM genes during the wound healing process [Hill, 2000]. Similarly, T-cells release another proinflammatory cytokine, IFN- γ , suppressing collagen synthesis after injury [Ghosh et al., 2001].

After injury, quickly released TGF- β assists in the attraction of fibroblasts, neutrophils and macrophages which results in TGF- β release further [Kane et al., 1991]. Therefore, TGF- β and its receptor expressions are elevated during wound healing process in fibroblasts of human postburn scars, in keloids and keloid-derived fibroblasts [Chin et al., 2001, Schmid et al., 1998]. Increased levels of TGF- β appear in fibrotic lesions of scleroderma patients during formation of scar tissue but not in the established lesions [Querfeld et al., 1999, Abraham et al., 2000]. These observations suggest the function of TGF- β in the initiation of the fibrotic response, *in vivo*. Similarly, the treatment of fetal wounds with TGF- β induces scarring and wound closure. Conversely, anti-TGF- β antibody or antisense oligonucleotide treatments suppress scarring and matrix synthesis [Cordeiro et al., 2003]. Supporting that notion, TGF- β 1 deficient mice display impaired late-stage scarring due to reduced collagen deposition accompanied by a pronounced, generalized inflammatory response and tissue necrosis [Böttinger et al., 1997]. Taken together, these results suggests roles of TGF- β in fibrosis while challenges regarding the use of this cytokine as an antifibrotic target still remains. The utilization of TGF- β as the target may be restricted to only acute cases which would require the antifibrotic treatment for a limited period. However, the long period of application of this cytokine to treat chronic fibrotic disease still remains as a challenge. Therefore, thorough characterization of the responses downstream of TGF- β and its receptors is crucial to develop clinically proper antifibrotic therapies.

1.2.6 Connective tissue growth factor functions and mechanisms of regulation

In this work, we investigate TGF- β induced and *SMAD* mediated transcriptional regulation of the target gene, connective tissue growth factor (*ctgf*, also known as *CCN2*) in NIH/3T3 mouse fibroblast cells. This cell line on which first studies of TGF- β induced *ctgf* upregulation was performed can serve as an excellent model system [Grotendorst, 1997]. Upon treatment with TGF- β , these cells acquire a myofibroblastic phenotype and further contribute to increased matrix deposition. *Ctgf* is a 38-kDa cysteine-rich secreted matricellular peptide that belongs

to a group of proteins named CCN whose members share a similar structure [Moussad and Brigstock, 2000]. These proteins are secreted and sequestered in the extracellular matrix interacting with cell surface receptors such as integrins, growth factors, cytokines or extracellular matrix proteins [Abreu et al., 2002]. Therefore, *ctgf* can be considered as a modulator of activities for other growth factors such as TGF- β or VEGF. *Ctgf* activity heavily depends on the cellular context and its interactions with other molecules in the cellular environment [Samarakoon et al., 2010]. It plays a key role in a broad array of cellular processes such as adhesion, migration, extracellular matrix production as well as regulation of development, differentiation and tumorigenesis [Cicha and Goppelt-Struebe, 2009]. It is barely expressed in normal adult tissue while it is stimulated in fibrotic tissue, in wound healing, during development and in some types of cancer.

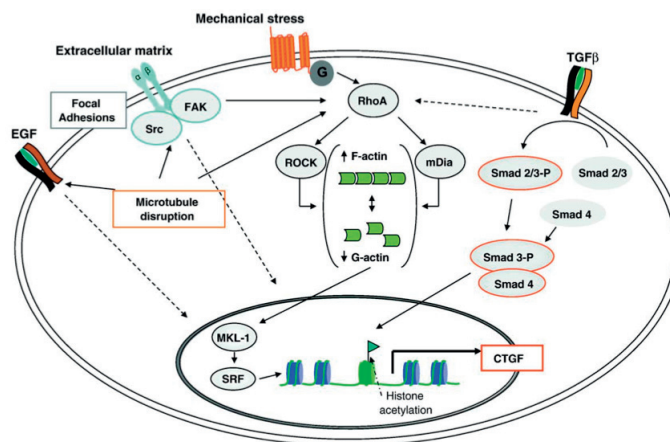


Figure 1.9 – Signaling pathways involved in the regulation of *ctgf* expression. Mechanical stress also regulates *ctgf* gene activity mediated by actin proteins modulating MKL-1, a required cofactor of SRF (serum response factor). Modified from [Samarakoon et al., 2010].

Ctgf is encoded by an immediate-early response gene, therefore it is expressed within minutes of stimulation and its upregulation does not depend on the newly synthesized protein [Hahn et al., 2000]. Treatment of cells with an inhibitor of protein synthesis (cycloheximide) results in a strong and long lasting upregulation of *ctgf* mRNA typical for immediate early response genes [Fowler et al., 2011]. Upregulation of *ctgf* mRNA upon treatment with protein synthesis inhibitor suggests the existence of short-lived proteins degrading *ctgf* mRNA to control *ctgf* expression levels.

Due to its multifunctional significance, *ctgf* expression is tightly regulated. *ctgf* mRNA synthesis occurs in basal state even in the absence of external stimuli when cells are transfected with *ctgf* promoter constructs [Cicha and Goppelt-Struebe, 2009]. In these conditions, promoter activity is observed for more than 24 hours even in the absence of external stimuli. Such basal regulation of *ctgf* involves soluble mediators together with mechanical forces (Figure 1.9). Its expression is predominantly regulated at the transcriptional level while few studies report the regulation of *ctgf* mRNA stability [Kondo et al., 2006]. Therefore, it provides

a suitable model system for the investigation of transcriptional response upon stimulus.

While *ctgf* was initially considered to be almost exclusively controlled by TGF- β , it has been shown that many other soluble mediators, including bioactive lipids, angiotensin II, VEGF, or lipoproteins as well as mechanical strain can upregulate *ctgf* expression [Cicha and Goppelt-Struebe, 2009]. It has become evident that mechanical signals cooperate with biochemical signals originating from growth factors and cytokines. Cells perceive their microenvironment through physical and mechanical cues, such as extracellular matrix (ECM) stiffness or confined adhesiveness. In response to mechanical strain, the cellular components of organ systems, particularly fibroblasts, endothelial, and smooth muscle cells upregulate the various growth factor synthesis including *ctgf* resulting in proliferative and fibrotic responses. *Ctgf* expression is upregulated in response to non-uniform shear stress while suppressed by laminar flow. In endothelial cells, *ctgf* upregulates in a RhoA-dependent manner in response to non-uniform shear stress [Cicha et al., 2009]. Induced *ctgf* gene expression depends on stress-associated restructuring within the cellular microfilament system. RhoA signaling and its downstream effectors ROCK and mDia result in elevated cellular F-actin structures and a corresponding decrease in monomeric G-actin (Figure 1.9). The ratio of G-actin and F-actin levels modulates the activity of certain serum response factor (SRF)-responsive genes including *ctgf* transcription [Vartiainen, 2008, Samarakoon et al., 2010]. SRF was shown to upregulate *ctgf* expression via its interaction with a response element about 4 kb upstream of *ctgf* promoter [Muehlich et al., 2007]. Similarly, the transcriptional regulators YAP/TAZ has been identified as downstream mediators of Wnt pathway mediating mechanical cues instructed by the cellular microenvironment to *ctgf* target gene induction [Dupont et al., 2011, Park et al., 2015, Tsutsumi et al., 2013].

In vivo studies indicated high correlation of the expression of *ctgf* transcripts with the expression or presence of TGF- β [Moussad and Brigstock, 2000]. *Ctgf* shows a much stronger induction upon TGF- β treatment compared to some other cytokines such as PDGF, FGF or EGF. Moreover, *ctgf* response upon TGF- β stimulation showed a prolonged profile even with a brief exposure to the ligand despite its relatively short-lived mRNA and protein, with mRNA half-life around 1-2h and protein being degraded within 3-4h. It has been shown that TGF- β induces *ctgf* expression directly by activating the transcription of the gene through its specific sequence element, TGF- β responsive enhancer (TGF- β RE), identified between nucleotides -244 and -166 upstream of the promoter [Grotendorst et al., 1996]. In adult normal fibroblasts, *ctgf* expresses only minimally in the absence of stimulation while in dermal fibroblasts obtained from lesional areas of scleroderma, *ctgf* is expressed constitutively. TGF- β RE control element plays a role in increased basal expression of *ctgf* in scleroderma fibroblasts [Leask et al., 2001, Holmes et al., 2001], mesangial cells [Chen et al., 2002].

TGF- β RE sequence differs from the TGF- β response elements described in other genes, including the SMAD recognition sequence. Therefore, induction of *ctgf* transcription by TGF- β occurs through SMAD mediation and action on the *ctgf* promoter via a functional SMAD binding site. This functional SMAD binding site, CAGACGAA located between nucleotides

–173 and –166 in the *ctgf* promoter, is necessary for *ctgf* induction in response to TGF- β in NIH 3T3 fibroblasts [Leask et al., 2001]. Mutations on this *SMAD* binding site led to complete removal of the *ctgf* response while increased *SMAD* levels enhanced *ctgf* expression significantly depending on *SMAD3* and *SMAD4* but not *SMAD2*. The *SMAD* element is not necessary for the basal high *ctgf* promoter activity seen both in dermal and scleroderma fibroblasts while *ctgf* upregulation depends on *SMAD3*-*SMAD4* complex binding to corresponding DNA sequence. Supporting this notion, transfection of the inhibitory *SMAD*, *SMAD7*, had little effect on basal expression while it significantly reduced *SMAD3*-*SMAD4* complex mediated *ctgf* promoter activity [Holmes et al., 2001].

The existing data suggest that *SMADs* regulate the TGF- β induced response of *ctgf*. However, these studies used static measurements which may not capture the dynamic nature of the pathway. Given the rapid expression of *ctgf* upon stimulation, it has remained a challenge to understand how it is transcriptionally regulated via *SMAD* signal transduction. Another important question that is yet to be fully resolved is the fate of target genes in response to the duration of signaling (transient versus sustained). Therefore, it is crucial to investigate how *SMAD* signaling modulates the target *ctgf* gene expression and transcriptional kinetics in single-cell dynamic measurements.

1.2.7 Single-cell dynamics of *SMAD*-mediated TGF- β response

Aforementioned quantitative analyses of the pathway indicated the average behavior of a cell population at certain time points which would not capture the true dynamics and the underlying heterogeneity of the response to stimuli. Yet, the long-term response to stimulus at the level of single cells needs further investigation. A major step in TGF- β signaling is the translocation of *SMAD* complexes from the cytoplasm to the nucleus. Single cell observations indicated that *SMAD2*-*SMAD4* complex formation and nuclear translocation of activated complexes display substantial cell-to-cell heterogeneity [Warmflash et al., 2012]. Heterogeneous signaling behavior upon stimulation was shown to be partially related to cell density and cell cycle stage [Zieba et al., 2012]. Moreover, some studies emphasized the importance of receptor dynamics determining the long-term signaling behavior of the TGF- β pathway [Vizán et al., 2013].

Single-cell studies have demonstrated that true signaling dynamics of the TGF- β pathway, while how cells interpret *SMAD* signaling and respond in single cells remains a challenging question mostly due to the lack of experimental systems allowing measurements of both *SMAD* dynamics and consequent transcriptional output in the same cells. To achieve that, some studies attempted to correlate signaling activity and protein expression in different pathways and organisms [Aymoz et al., 2016, Sung et al., 2014]. Similarly, several studies attempted to decode the contributions of *SMAD* dynamics to downstream response in TGF- β pathway [Sorre et al., 2014, Frick et al., 2017]. Sorre et al. demonstrated the influence of the speed of the ligand presentation TGF- β signaling outcomes such as *SMAD* translocation and

Chapter 1. Introduction

its target gene response using synthetic TGF- β reporters. However, the analysis in this study linking SMAD dynamics to its expression response lacks the single-cell resolution and uses a synthetic TGF- β targeted promoter constructs which might have not the same epigenetic controls as the TGF- β responsive elements in the endogenous gene. Similarly, Frick et al. studied SMAD-mediated target gene transcriptional activity and revealed that cells sense signaling state relative to background in a fold-change manner and encode in downstream responses. Yet, in this study, the target gene response analysis relied on a static measurement (single-molecule FISH) which may have not captured the full dynamics the response.

This dissertation study aims at understanding how transcription factor expression levels and its translocation dynamics quantitatively relate to target gene activation kinetics in response to stimulus. Therefore, using SMAD-mediated TGF- β signaling in fibroblast cells as the model system, we embarked on the task to decipher the quantitative link between SMAD2/4 dynamic profiles and its target *ctgf* gene response which would provide new insights into both TGF- β and transcription regulation fields.

2 Materials and Methods

2.1 Cell line generation and cell culture

2.1.1 Construction of lentiviral plasmid constructs

pLVTRE3G-NLuc-SMAD4 was constructed by inserting NLuc-SalI-SMAD4-STOP into pLVTRE3GMCS [Deluz et al., 2016] linearized with SalI using a two-fragment In-fusion (Clontech) reaction. pLVTRE3G-NLuc-SMAD2 was constructed by cloning into pLVTRE3G-NLuc-SMAD4 digested with SalI and NdeI (NEB) to remove SMAD4 (Figure 2.1 and 2.2). pLV-PGK-rtTA3G-IRESHygro was constructed as described previously [Mandic et al., 2017]. All constructs were verified by Sanger sequencing.

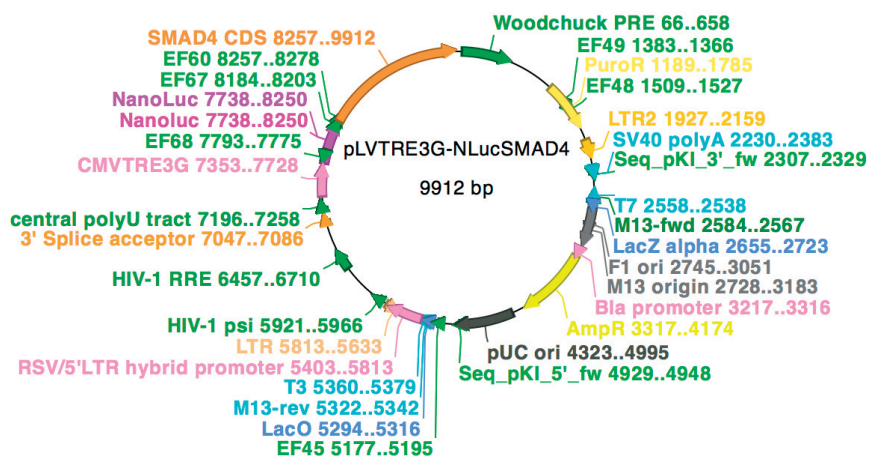


Figure 2.1 – Graphical representation of the pLVTRE3G-NLuc-SMAD4 vector.

concentrations spanning from 1 to 500 ng/ml. 1 day after seeding and doxycycline treatment, half of samples were added the NanoLuc substrate MT Cell Viability Assay Substrate (Promega, dilution 1:4000 in cell culture medium) for SMAD2/4 measurements while the other half were added 0.5 mM luciferin (NanoLightTechnology, 306A) for ctgf measurements. The luminometer system allowed us to supplement samples with TGF- β (5nM) for stimulation experiments. The recordings had a total duration up to 24 hours with time resolution of 6 minutes, and the collection of data was performed with the LumiCycle Data Collection software. Matlab was used for further analysis of obtained bioluminescence recordings.

2.2.2 Single-cell time lapse luminescence microscopy

One day before imaging, 2×10^5 cells, stably expressing short-lived luciferase and Nanoluc-SMAD2/4, were diluted 1:1 into non-luminescence NIH-3T3 cells (total 4×10^5 cells), seeded in a 2.3 cm glass bottom dish (FluoroDish, World Precision Instruments) and maintained in DMEM media supplemented with 10% Fetal Bovine Serum and 1% L-Glutamine – Penicillin – Streptomycin (PSG) antibiotics. Non-luminescent cells provided space for spatial separation of individual cells. High confluency was needed to keep fibroblast cells in G0 phase preventing cell divisions as much as possible. In this way, the effect of cell cycle on single cell responses, which can be an additional source of extrinsic noise, was minimized.

Luminescence images at the single-cell level were acquired using a Luminoview LV200 imaging system (Olympus) equipped with an electron multiplier charge-coupled device camera (ImagEM, Hamamatsu Photonics, EM-CCD C9100-13). Images were taken using a 60-fold oil-immersion magnification objective (OlympusUPlanSApo 60x, NA 1.35, oil immersion). The culture dish was maintained in a 37°C, 5% CO₂ humidified incubator and imaged up to during 24 hours. For dual-luminescence imaging of single cells, a 700nm LP filter (Chroma) for Fluc and 460/36nm band-pass filter (Chroma) for Nluc were used. Different exposure times spanning from 1 to 5 min for channels were used in preliminary experiments (data now shown) to determine the optimal exposure times for each channel ensuring high signal-to-noise ratio and avoiding saturation of signals in each frame. Considering also the importance of temporal resolution in the analysis of single cell traces, images throughout time-lapse experiments were acquired as following: a frame in the Fluc channel (exposure time: 2min), followed by a frame in the Nluc channel with a higher exposure time (3min), overall allow a temporal resolution of 5 min. Right before imaging, samples were added the NanoLuc substrate MT Cell Viability Assay Substrate (Promega, dilution 1:4000 in cell culture medium) for Nluc imaging and 0.5 mM luciferin (NanoLightTechnology, 306A) for Fluc imaging.

2.3 Extraction of single cell traces

In what follows, steps of analyzing image series obtained by luminescence microscopy are explained. Custom-written Matlab (MathWorks) scripts were used for preprocessing images, segmentation, tracking of individual traces, quality control of obtained traces and further

analysis. Single-cell data encompassed a range of conditions in which experimental parameters were varied including ligand concentration, SMAD abundance, ligand type and temporal profile reaching a large number of single cells for long-term luminescence imaging (more than 2000 cells and ~1 million time points, in total with replicates of each condition). The extraction of the single cell traces was based on the CAST (Cell Automated Segmentation and Tracking) software [Blanchoud et al., 2015], in which, we adapted the existing tool to our luminescence recordings. This required the optimization of analysis parameters for both segmentation and cell tracking since different imaging conditions and cellular contexts impact the image features. Different steps involved in the image analysis pipeline (e.g. segmentation, quantification and tracking) will be further explained (Section 2.3.1 to 2.3.4). CAST was developed as a set of custom MATLAB functions. Importantly, while its applicability on bioluminescence data is explained here, its modular implementation can be easily adapted to the analysis of a wide variety of time-lapse data [Blanchoud et al., 2015, Bieler et al., 2014]. CAST consists of steps that are configurable, editable and, once the corresponding parameters are manually tuned to the specificity of the analyzed recording, fully automatized. Together, these features provide a robust way to track the reporter signals in mammalian culture cells.

2.3.1 Pre-processing images

To ease the analysis of raw images obtained from luminescence imaging, the following preprocessing steps are applied. The background is obtained, for each image of the recording independently, by opening the raw image, smoothing it using a Gaussian kernel, and least-squares fitting of a 2D quadratic surface on the resulting intensities. Of note, the fitted background, that is then subtracted to the raw image, is centered (i.e. has a mean of zero) to remove trends from the image without altering the range of values. CAST also removes saturated pixels from cosmic rays (useful for bioluminescence when long exposure times are used) with a local histogram approach [Pyh, 2004], using robust statistics for the estimation of the location and scale (i.e. using the median instead of the mean and the *spell out*, MAD, instead of the standard deviation). CAST also remove defective pixels by detecting them as a signal varying more from the mean of the whole image than a defined threshold times the standard deviation. Finally, the entire stack of images can be normalized (i.e. scaled) to compensate for variation between recordings. The result of all these operations is then stored in a new image stack that is used for all the subsequent steps of the analysis.

2.3.2 Image segmentation

After preprocessing images, images are convolved with a family of cell-like filters (Figure 2.3). The resulting images are converted to binary masks using an adaptive threshold. The binary images are cleaned using morphological operations. Finally an automatic correction is done by splitting adjacent cells that have been recognized as one object (Figure 2.4).

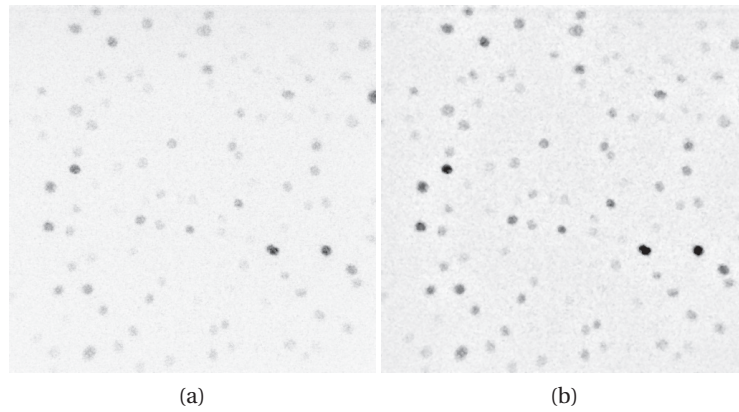


Figure 2.3 – Preprocessing images. *a) Raw image from Fluc channel representing nuclei to be preprocessed and segmented. Inverted gray scale images are used for better visualization. b) Preprocessed image ready for segmentation.*

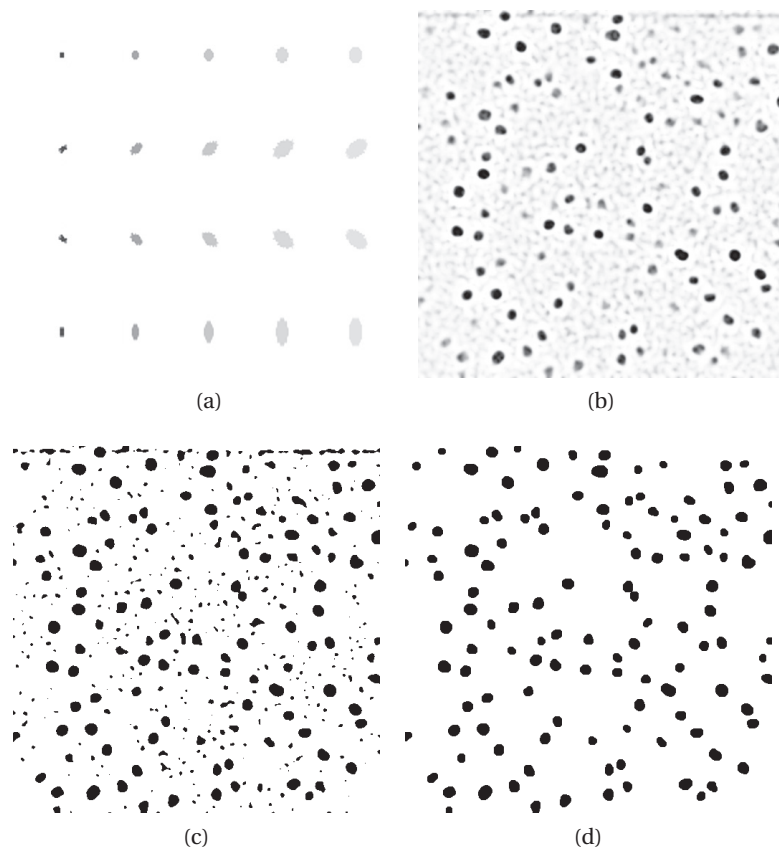


Figure 2.4 – Cell segmentation overview. *a) A collection of cell-like filters used before segmentation. b) Filtered image before segmentation. c) Filtered images are converted to binary masks using an adaptive threshold. d) After segmentation, binary images are cleaned using morphological operations.*

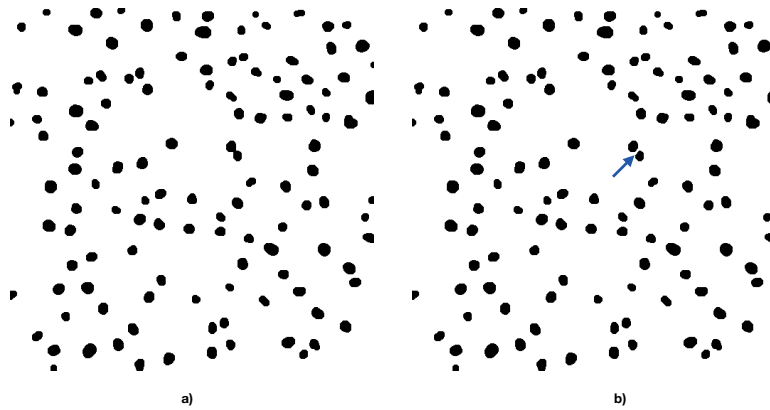


Figure 2.5 – Automated object splitting process detects and corrects merged cells. a) Segmented image before splitting process consisting of one merged pair is shown. b) After automated splitting process. Arrow shows the split cell pair.

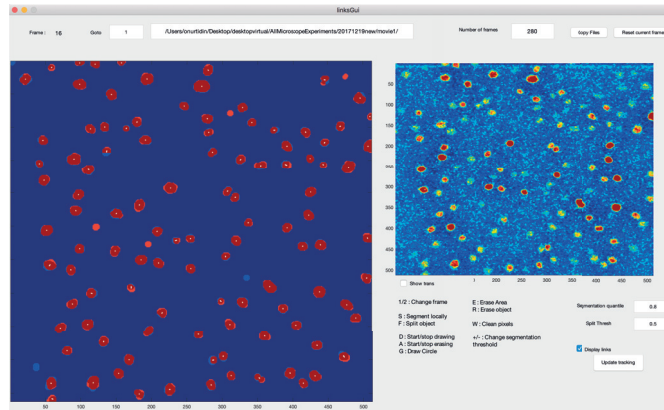


Figure 2.6 – Segmentation validation tool. The left panel displays both the current segmented frame and the previous one. Red areas correspond to parts of the image corresponding to the segmented objects in the two frames. Orange objects represent detection only in the current frame, and light blue ones display those only present in the previous frame. The white lines represent which objects are linked by local tracking between the two frames, thereby also showing the distance traveled between two frames. If needed, the user is able to execute a series of operations on the current segmented frames to correct automatic segmentation errors.

Robustness assessment of nuclear segmentation

The segmented images were manually corrected and validated using a custom Matlab user interface (Figure 2.6). When mistakes were detected, segmentation was optimized and repeated with altered parameters. The adaptive threshold was adjusted to encompass the whole nuclear signal. An example of a typical single cell segmentation is shown in Figure 2.7.

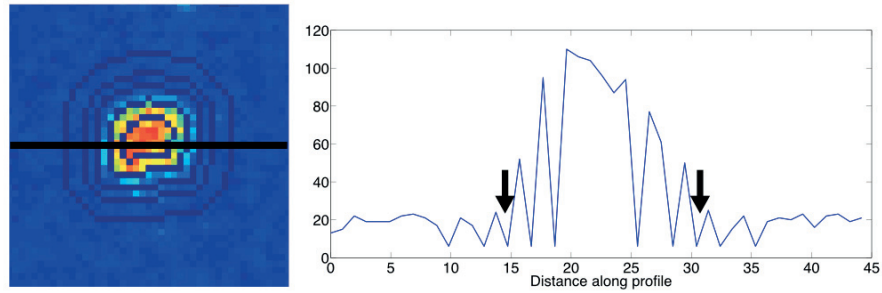


Figure 2.7 – Robustness assessment of nuclear segmentation. Signal intensity profile (right) along a path represented by black line (left) for a single cell is given. Circles represent different dilation and erosion of the segmented nuclei represented by the middle circular region. Adaptive thresholding was adjusted to include the whole nuclear signal measured from enclosed region (black arrows).

Robustness assessment of cytoplasmic segmentation

We segmented nuclear region as described previously and defined a cytoplasmic region using an annulus around the nucleus. Using a varying distance from nucleus to define a cytoplasmic region, nuclear to cytoplasmic signals for each case is quantified. This helped us to define the most robust distance for cytoplasmic annulus (Figure 2.8).

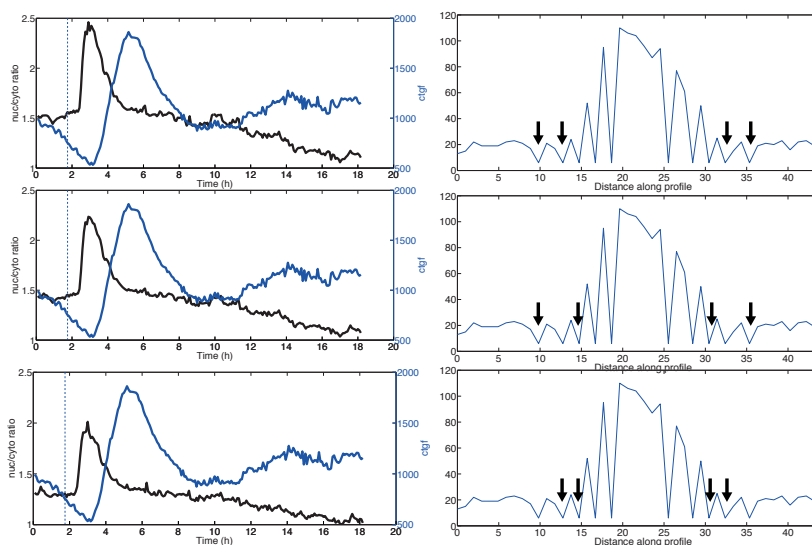


Figure 2.8 – Robustness assessment of cytoplasmic segmentation. For the cytoplasmic region quantification, a ring around nucleus was used. Three panels show the effect of different distance selection of ring borders from the nucleus on the quantified nuclear to cytoplasmic ratio (black plots). Rings that were not distant enough (1-5 pixels, lower and middle panel) from detected nuclei result in bias from the nuclear signal. Above a certain distance from the nucleus (around 8 pixels), nuclear-to-cytoplasmic ratio quantification stabilizes (upper panel). Arrows show the ring borders to obtain nuclear-to-cytoplasmic ratios on the left part. Nuclear segmentation is kept the same based on ctgf signal as explained in Figure 2.7.

2.3.3 Cell-tracking

Tracking of cells is performed using a custom, memory efficient, implementation of an algorithm that solves both the frame-to-frame linking of objects and the problem of assigning trajectories globally using the Hungarian algorithm [Jaqaman et al., 2008]. Our implementation uses sparse matrices for the previously proposed cost matrices to decrease the memory load of the algorithm, a major bottleneck in the standard implementation, hence permitting the efficient handling of a very large number of trajectories. Thus, CAST can efficiently handle the gap closing (i.e. linking similar objects that were not segmented in some frames), merging and splitting steps, and thus provides a powerful framework for accurately tracking cell lineages in mammalian culture cells. Importantly, the splitting step grants an automated approach for tracking both daughter cells of a division, thus providing a full lineage of the cells present in the recording. Finally, to remove spurious detection of cells, we implemented an additional optional step to filter out short trajectories before the gap closing, merging and splitting steps, thus preventing these glitches from being linked together into a spurious trajectory.

2.3.4 Assessment of the cross-talk between Fluc and Nluc channels

For dual-luminescence imaging of single cells, a 700nm LP filter (Chroma) for Fluc and 460/36nm band-pass filter (Chroma) for Nluc were used. These filters allowed to resolve the emission spectra of Nluc and Fluc (Figure 3.2). We further determined the crosstalk between Fluc and Nluc channels using Nluc-Fluc calibration cell line. No detectable bleed-through signal was observed in 460/36nm (Nluc) channel due to luminescence emitted by Luciferin. The luminescence measured by 700nm LP filter (Fluc) is shown as a function of luminescence measured by 460/36nm BP filter (Nluc) when the sample was treated with MT Cell Viability Assay Nluc Substrate alone (Figure 2.9). It is assumed that the intensity measured in 700nm LP channel was proportional to the intensity measured in the 460/36nm channel since it was due to leakage of luminescence from one filter to another. Taking this characteristic into account, the data were fitted linearly. Analyses of intensities from the two channels revealed 1.5% leakage (slope) of Nluc signal to Fluc channel. This fraction of Nluc signal was subtracted from Fluc signal measured in single cells, when analyzing dual-luminescence imaging frames.

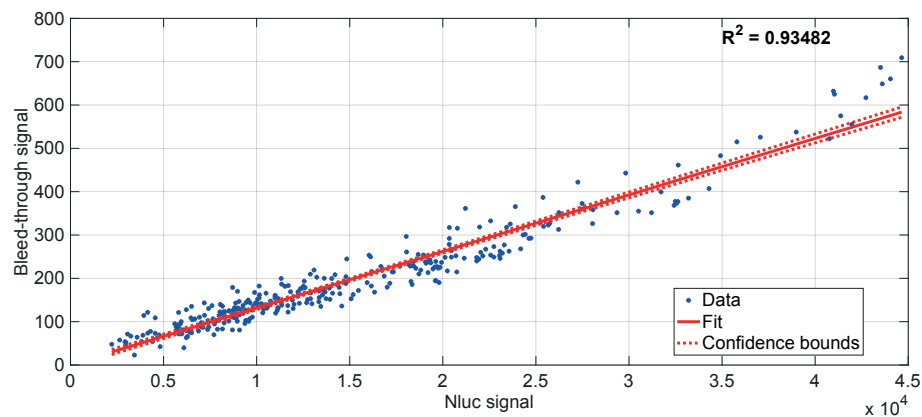


Figure 2.9 – Crosstalk analysis of dual-luminescence imaging system. Graph represents the luminescence measured by 700 nm LP filter (Fluc) in function of luminescence measured by 460/36 nm BP filter (Nluc) when the sample was treated with MT Cell Viability Assay substrate alone. The data represents nuclear signals measured from individual cells in both Nluc and Fluc channels. The intensity of both channels were analyzed after subtracting background values. (Linear coefficient: 0.015(%1.5), $R^2 = 0.93482$)

2.4 Clustering of single cell trajectories

We used *k-means clustering* approach to group cells with similar dynamic patterns. *k-means clustering* is an iterative, data-partitioning algorithm that assigns n observations to k clusters defined by centroids [Lloyd, 1982]. The number of clusters (k) is chosen before the algorithm is executed. Steps of *k-means clustering* algorithm with the modification of *k-means++ algorithm* for the initial seeding step is given below:

1. Randomly choose k cluster centers (*centroid*). One modification for this step is to use

k-means ++ algorithm for better cluster center initialization [Arthur and Vassilvitskii, 2007]

2. Repeat until cluster assignments do not change since doing so does not reduce the sum of distances, or the maximum number of iterations is reached.
 - For each point, calculate point-to-cluster centroid distance
 - Find the nearest centroid and assign each point to its closest centroid
 - Calculate the average of the newly assigned observations to obtain k new centroid locations

Step 2 in the algorithm requires a distance metric to check in every iteration with newly assigned centroid locations. We employed 1 minus correlation between samples as the dissimilarity metric instead of commonly used Euclidean distance for two reasons: 1) We mainly aimed to observe subgroups differing in long-term dynamics which came to our notice during qualitative assessments 2) Observed variability in the *ctgf* expression level may be misleading when using a clustering approach that relies on Euclidean distance and can ignore the dynamic features of the response.

2.5 Western Blot

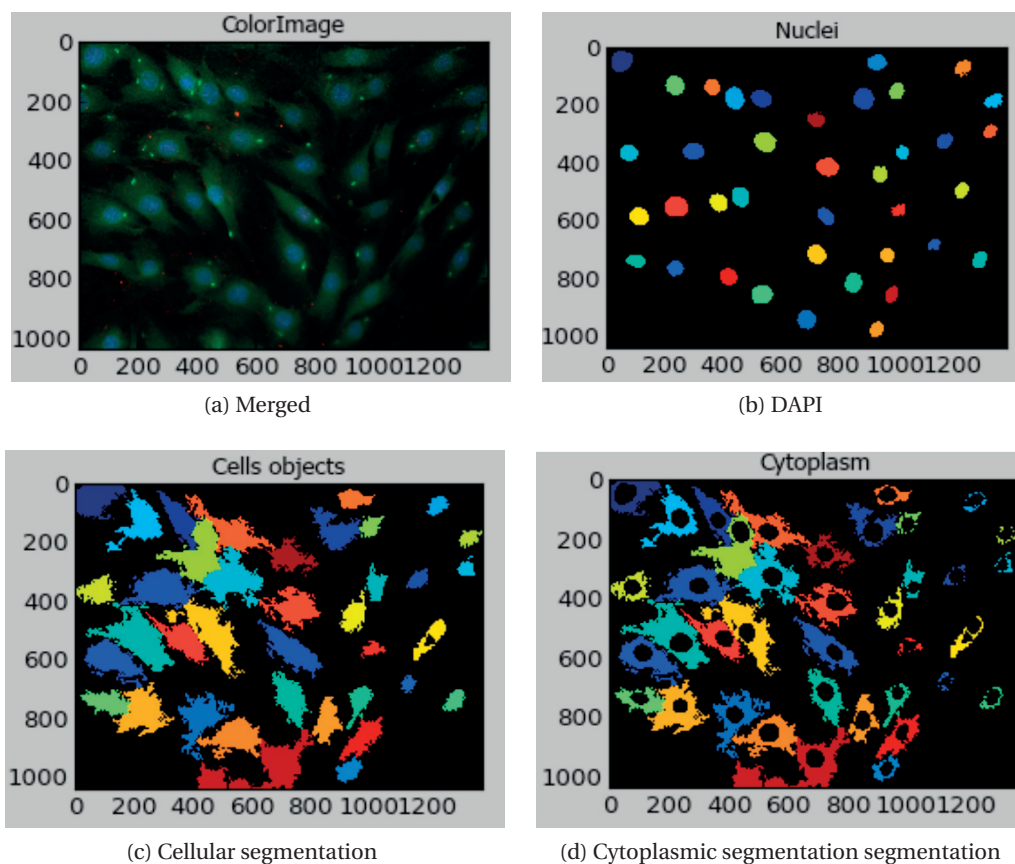
Cells were plated 1 day before experiments and grown until confluent. When seeding, cells were treated with a range doxycycline concentrations spanning from 1 to 20 ng/ml for doxycycline dose response investigation. For time-lapse TGF- β stimulation samples, cells were stimulated 1 day after seeding and collected with counting at indicated time points by trypsinization and centrifugation. Cells were then lysed in RIPA buffer (50 mM Tris pH 7.4, 1% NP-40, 0.5% NaDeoxycholate, 0.1% SDS, 150 mM NaCl, 2 mM EDTA), supplemented with 1mM PMSF (AppliChem A0999.0005) and Protease Inhibitors (Sigma P8340-5ML). Samples were left on ice for 30 minutes and then spun down at 14,000g for 15 minutes at 4°C. The protein concentration of the supernatant was determined by performing a Bicinchoninic acid assay (BCA, ThermoFisher 23235) and 15 ug of protein were mixed with Laemmli sample buffer (Invitrogen NP0007) and loaded on an SDS gel (BioRad 456-1094) for separation (SDS Running Buffer 25mM Tris, 190mM Glycine, 0.1% SDS). Proteins were subsequently transferred from the gel onto a nitrocellulose membrane using a dry transfer system (Merck IB21001, iBlot 2 Dry Blotting System). Antibodies against C-terminally phosphorylated SMAD4 (3108; Cell Signaling Technologies) (dilution 1:1,000), SMAD2/3 (610482; BD Transduction Labs) (dilution 1:1,000), and SMAD4 (B-8; Santa Cruz) (dilution 1:500) were used. The membrane was blocked with 5% bovine serum albumin or 5% milk (Roth T145.3) in TBS-T (for phosphorylated samples) or PBS-T followed by incubation with primary antibody overnight. The following primary antibodies with given dilutions were used; SMAD4 (ABE21 ; Merck, dilution 1:2000), pSMAD2 (D27F4; Cell Signaling Technologies, dilution 1:1000), SMAD2 (D43B4 ; Cell Signaling Technologies, dilution 1:1000). Membranes were subsequently washed shortly and incubated with

HRP-conjugated secondary antibodies in 5% milk in TBS-T (for phosphorylated samples) or PBS-T. The following secondary antibodies with given dilutions were used; antimouse-IgG-HRP (W402B; Promega, dilution 1:10.000), antirabbit-IgG-HRP (W401B; Promega, dilution 1:10.000). Before imaging, membranes were washed lastly with in TBS-T/ PBS-T. The protein bands were visualized using Clarity Western ECL Substrate (BioRad 170-5060). Images were captured using a Vilber-Fusion chemiluminescence system (Molecular Imaging Vilber Fusion FX7) and were analyzed using ImageJ.

2.6 Immunofluorescence

NIH-3T3 cells were fixed for 15 min with ice-cold 4% PFA (AppliChem A0877,0500) in PBS, permeabilized and blocked with chilled PBS-Triton (AppliChem A1388,0500) and 1% FBS for 30 - 60 min. Samples were incubated with the primary antibody in PBS and 1% FBS overnight at 4°C, washed twice in PBS, and incubated with the secondary antibody in PBS and 1% FBS for 45 - 60 min. Samples were then washed three times with 0.1% PBS-Tween (Fisher Scientific BP337- 500), incubated with 1 g/mL DAPI for 15 minutes, washed twice with 0.1% PBS-Tween and once with PBS.

To quantify signals in samples, a semi-automated image analysis pipeline built in Cell-Profiler software was used as depicted in Figure 2.10. Figure 2.10a shows DAPI and SMAD4 channels merged. DAPI staining was used to precisely locate nuclear regions (Figure 2.10c) and SMAD4 staining was used to define cellular borders and to determine cytoplasmic regions (Figure 2.10c and Figure 2.10d). Manual correction was performed for erroneous detections. Based on defined nuclear and cytoplasmic regions, signal intensities from control and temporal samples were obtained. Background subtraction was performed using control samples.



(a) Merged
(b) DAPI
(c) Cellular segmentation
(d) Cytoplasmic segmentation

Figure 2.10 – Quantification of immunofluorescence samples with the automated CellProfiler pipeline. (a) Merged image of DAPI and SMAD4. (b) Nuclear segmentation based on DAPI signal. Nuclei touching image boundaries were not segmented. (c) Cell contour segmentation based on SMAD4 signal (d) Defined cytoplasmic regions. Images are taken from Cell Profiler software [Carpenter et al., 2006].

3 Results

3.1 Simultaneous monitoring of transcription factor signaling and target gene activation in single living cells

3.1.1 Generation of stable cell lines for dual-luminescence imaging

Suter et al. previously generated a lentiviral gene trap cell line with an insertion of a short-lived firefly luciferase into the 5th exon in one allele of the connective tissue growth factor (*ctgf*) gene. This cell line allowed to monitor the transcriptional profile of the endogenous *ctgf* gene both in steady state [Suter et al., 2011] and upon stimulation [Molina et al., 2013].

The gene trap construct elements (Figure 3.1) were assembled from 5' to 3' [Friedrich and Soriano, 1991]. The EMCV internal ribosomal



Figure 3.1 – Schematic of the lentiviral gene-trap vector to generate the gene-trap cell line

entry site (IRES) initiates independent translation of the protein [Jang and Wimmer, 1990]. F2A (Foot-and-mouth disease virus2A) mediates the cleavage into separate Bsd (blasticidin) and NLSluc proteins [Ryan and Drew, 1994]. Nuclear localization signal (NLS) increases the local luciferase (luc) concentration for better luminescence signal. ARE (AU-rich element) controls mRNA stability [Winzen et al., 1999] while a destabilized version of firefly luciferase (Promega Corp.) was used for protein instability. This vector was used to generate the gene trap cell line after Bsd selection.

Nanoluciferase is a small luciferase subunit (19kDa) from the deep sea shrimp *Oplophorus gracilirostris*. Optimization of this protein structure together with development of a novel substrate (furimazine) resulted in a novel bioluminescent system (Promega). It has been shown that the Nluc system provides several advantages over other established systems such as smaller size, enhanced stability and around 100-fold increase in luminescence and lower background (Figure 3.2). Its smaller size results in minimal disturbance and bias in its experimental hosts. Moreover, the spectral profile of Nluc displays an emission maximum of 460 nm significantly distinct from firefly luciferase emission (Figure 3.2b). Thus, with its

Chapter 3. Results

superior features and well separated spectra, Nluc suits for applications involving other longer wavelength reporters (e.g., firefly luciferase in our case).

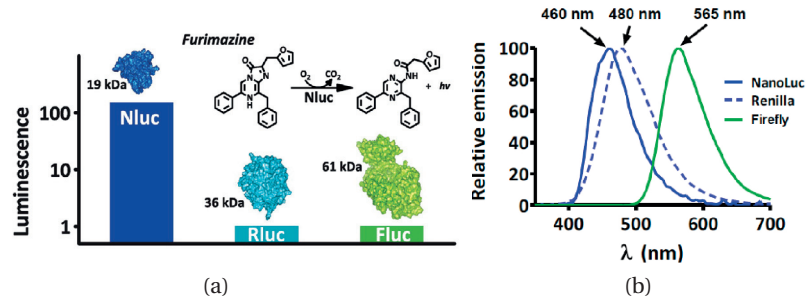


Figure 3.2 – Comparison of luminescence systems. a) Luminescence systems performance comparison based on molecular size and luminescence efficiency. b) Emission spectrum for different bioluminescence systems with Firefly, Renilla and Nanoluc luciferases. The distinct emission spectra of Nanoluc and Firefly luciferase allows simultaneous monitoring of two channels. Adapted from [Hall et al., 2012].

We established two Tet-On inducible stable cell lines each expressing a fusion protein of a Nluc reporter to either *SMAD4* or *SMAD2*, the two main transcription factors downstream of TGF- β signaling pathway. These cell lines that also contain a short-lived firefly luciferase reporter for the expression of the target endogenous *ctgf* gene allowed us to perform dual-luminescence imaging (Section 2.2.2, Figure 3.3, Figure 3.4).

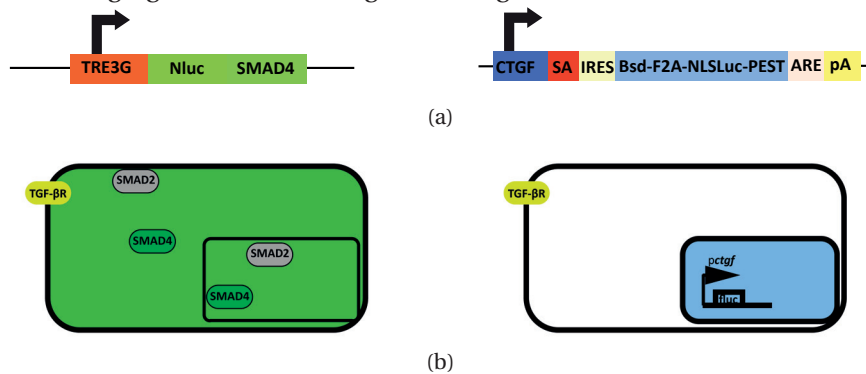


Figure 3.3 – Schematic illustration of the dual-luminescent reporter system. a) Nanoluciferase was fused to *SMAD4* and expressed under the control of doxycycline inducible promoter TRE3G (left). Schematic representation of the luciferase (*Luc*) fusion constructs used in the experiments (right, [Suter et al., 2011]). Constructs used in the experiments Bsd: Blastocidin-deaminase; F2A: foot-and-mouth virus co-translationally cleaved peptide; NLS: Nuclear localization signal; PEST: destabilizing sequence. Using this gene trap lentivector, short-lived NLS-firefly luciferase was integrated into endogenous locus and expressed under endogenous promoter. Arrows represent promoters. b) In unstimulated condition, *SMAD4* is localized both in the nucleus and cytoplasm and *ctgf* is expressed at a basal level localizing into nucleus due to fused nuclear localization signal.

3.1. Simultaneous monitoring of transcription factor signaling and target gene activation in single living cells

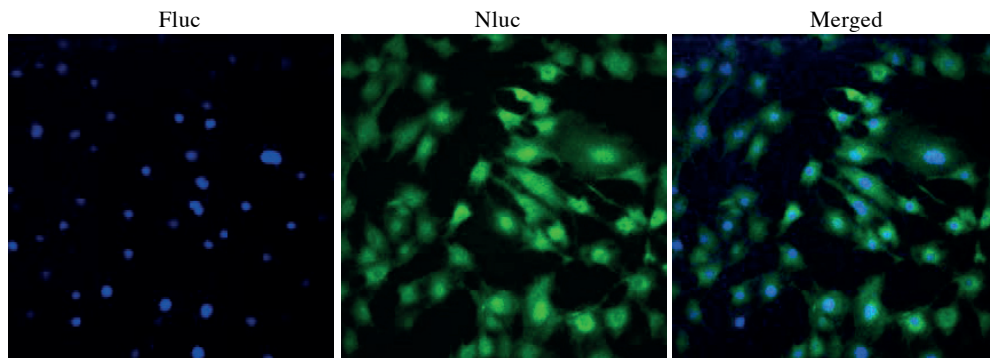


Figure 3.4 – Dual-luminescence detection of SMAD4 and *ctgf* in NIH 3T3 cells. Image shows one frame obtained from luminescence movies: Fluc (left, blue) channel, Nluc channel (green, middle), merged image (right). Fluc signal is localized to nucleus via an NLS signal. Nluc-SMAD4 signal is distributed inside the whole cell in the basal state. Nuclear Fluc signal is used to segment nuclei of cells in luminescence imaging.

Nluc-SMAD2/4 fusion reporter proteins were designed to be expressed with an inducible system. For this purpose, cells were transduced with two lentiviral vectors. The first one expressed the reverse tetracycline transactivator rtTA3G while the second one expressed reporter fusion proteins under TRE3G promoter (Clontech). This promoter consists of repeats of tet operator sequence named as tetracycline response element (TRE). In absence of doxycycline, a tetracycline analog, rtTA3G protein is not able to bind and activate TRE3G promoter. Moreover, TRE3G promoter does not have binding sites for endogenous mammalian transcription factors. Therefore, the inducible promoter TRE3G provides for very low basal expression. When cells are cultured in the presence of doxycycline, the transactivator protein undergoes a conformational change that allows it to bind to tet operator (tetO) sequences located within TRE3G promoter and activate expression (Figure 3.5).

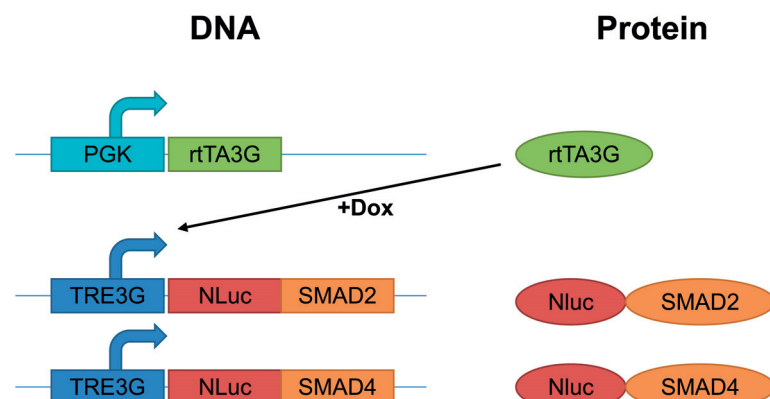


Figure 3.5 – Schematic description of Tet-On 3G inducible system. The Tet-On 3G system allows inducible gene expression only in the presence of doxycycline. The reverse tetracycline transactivator rtTA3G is constitutively expressed under PGK promoter. In the presence of doxycycline, rtTA3G protein binds and activates TRE3G promoter expressing Nluc-SMAD4/2 fusion proteins.

This Tet-on inducible expression system allows precisely regulated control of transgene expression which is reversible and reproducible [Gossen and Bujard, 1992]. It provides very low basal expression and high maximal expression after induction. Such tunable expression

Chapter 3. Results

profile offered the possibility to tune expression levels *SMAD* reporter in experiments which will be covered in the next sections. Therefore, it was crucial to characterize expression levels of the reporter before performing single-cell experiments.

3.1.2 Validation and characterization of dual-luminescence reporters

In order to avoid high overexpression of the reporter and to mimic the endogenous *SMAD* levels and localization, both parental cell line and *Nluc-SMAD* reporter clones were characterized regarding *SMAD* localization and expression levels.

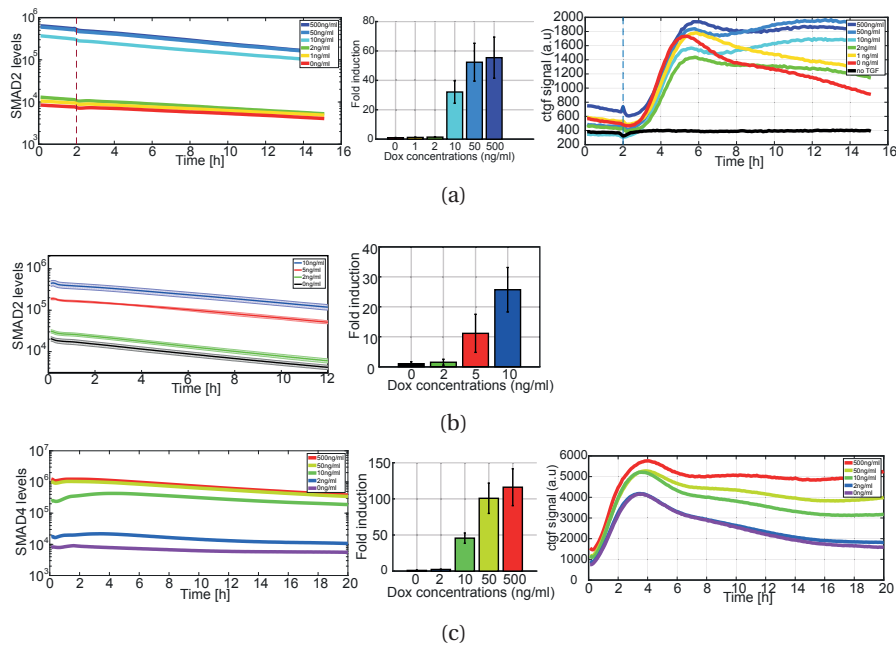


Figure 3.6 – Doxycycline induced *SMAD* reporter levels and its effect on *ctgf* response upon *TGF- β* stimulation in population-level measurements. In lumicycle measurements, Firefly and Nanoluciferase expression levels were measured in parallel in response to a range of dox concentrations. a) Samples cultured with indicated doxycycline concentrations are stimulated with *TGF- β* (5 nM) as indicated by dashed black line. *SMAD2* expression levels varied gradually depending on doxycycline concentration. *TGF- β* stimulation didn't have any impact on *SMAD2* levels (left). *SMAD2* levels dose-response was calculated averaging signals on the left panel in triplicates (middle). Upon *TGF- β* stimulation, the mean *ctgf* response of triplicates corresponding to *SMAD* levels in the left and middle panel. b) Samples cultured with lower doxycycline concentrations to characterize the expression induction from 2ng/ml to 10ng/ml (left and middle panel). c) The same experiment and quantification using *Nluc-SMAD4* cell line (*TGF- β* stimulation at $t=0$). Samples are color coded as shown in legends.

The doxycycline-dose responses of both *Nluc-SMAD4* and *Nluc-SMAD2* reporter levels were characterized using Lumicycle measurements. When seeding, cells were treated with a range doxycycline concentrations spanning from 1 to 500 ng/ml. Expression levels of both *Nluc-SMAD4* and *Nluc-SMAD2* reporters, in comparison to uninduced leakage expression, were measured. Samples were also stimulated with *TGF- β* to check how samples with different *SMAD* reporter levels respond. Regarding *SMAD* levels, *Nluc-SMAD2* cell line showed up to

3.1. Simultaneous monitoring of transcription factor signaling and target gene activation in single living cells

50-fold induction (Figure 3.6a and Figure 3.6b) while Nluc-SMAD4 covered more than 2 orders of magnitude fold-induction depending on doxycycline concentration (Figure 3.6c). Both reporters have a similar dose response profile that has a profound increase from 2 ng/ml to 10 ng/ml of doxycycline concentration (Figure 3.6a and Figure 3.6c, left and middle panel). Therefore, we also repeated the same experiments covering lower dose range of doxycycline (0-2-5-10 ng/ml). This set of samples revealed the gradual transition in expression levels (Figure 3.6b, left and middle panel). Gradually higher SMAD levels resulted in slight differences in ctgf response profiles upon TGF- β stimulation. In SMAD2 samples, there was no clear enhancement in ctgf responses with the increased SMAD2 levels (Figure 3.6a, right panel). Yet, in case of SMAD4, the increasing levels of SMAD4 resulted in slightly higher initial response levels and more sustained long term ctgf signals. In uninduced samples and the sample induced with low dose of doxycycline (2 ng/ml), the response was transient. Starting from certain level of doxycycline concentration (10 ng/ml), the response gradually became less transient (Figure 3.6c, right panel).

The dose response of doxycycline was previously characterized; however, these results only showed the comparative fold-induction of transgene expression relative to its leakage expression. Therefore, it was essential to compare ectopic expression of SMAD reporters with endogenous levels. Characterization of endogenous SMAD2/4 and Nluc-SMAD2/4 levels by Western blot showed that for low doses of doxycycline (2 ng/ml), Nluc-SMAD reporter levels are either on the order of endogenous SMAD levels (in case of SMAD4) or much less than the endogenous levels (in case of SMAD2). Starting from doxycycline concentration of 10ng/ml, the exogenous levels reach and exceed endogenous levels with the increasing doxycycline levels (20 ng/ml) (Figure 3.7). This observation agrees with the previous dose-response profile of dox induction (Figure 3.6).

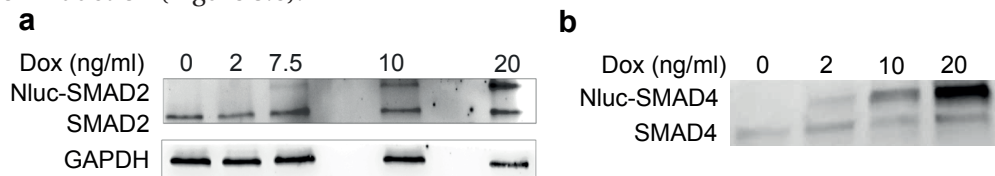


Figure 3.7 – Western blot analysis of endogenous and Nluc-tagged SMAD2/4 in the stable clonal reporter cell lines. Samples were treated with doxycycline using indicated concentrations. a) Nluc-SMAD2 cell line. GAPDH is shown as a loading control. b) The same experiment is repeated for Nluc-SMAD4 cell line using indicated doxycycline concentrations.

Our reporter system allowed us to operate in different regimes of SMAD levels. Previous studies used different expression levels of exogenously tagged SMADs in comparison to endogenous levels ranging from subendogenous levels ($\times 0.5$, [Strasen et al., 2018]) to over expressed regime (>2 , [Warmflash et al., 2012]). We started our single-cell microscope experiments with low doxycycline concentration (2ng/ml) to minimize perturbations of the signalling network with ectopic expression. Such expression level provided the minimal signal that can be detectable in the luminescence microscope while preserving the system as close as possible to its endogenous state. However, the tunability of doxycycline-inducible promoter was exploited to adjust SMAD expression levels in further experiments in the study. This allowed us to investigate the effect of SMAD abundance on the target gene response upon

stimulation. Taken together, these results validate that our reporter system covered a range expression levels and can be utilized for real-time stimulation experiments.

3.1.3 Single-cell imaging measurements of SMAD2/SMAD4 translocation and *ctgf* expression response in TGF- β stimulation experiments

Aforementioned population averaged measurements initially gave some basic understanding of the newly designed SMAD reporters. Yet, they did not reveal the detailed cellular dynamics and variability. Only dynamic single-cell measurements would allow us to dissect those aspects of the response to TGF- β stimulation. However, before performing TGF- β stimulation experiments, the behaviour of the reporter system regarding SMAD localization and the target gene expression in the absence of any stimuli was checked (Figure 3.8a). In the basal state, a roughly homogeneous localization of both SMAD2 and SMAD4 between nucleus and cytoplasm, which was quantified as the ratio of nuclear to cytoplasmic signal, was observed. This localization is due to continuous shuttling between the nucleus and the cytoplasm and dictated by the relative strengths of the nuclear import and export signals [Pierreux et al., 2000]. On the other hand, *ctgf* expression levels, revealed a slightly decaying profile for the initial 1-2 hours followed by a non-changing profile on average, while exhibiting random bursts of activity in some single cells agreeing with its previously observed steady-state characteristics [Suter et al., 2011]. The initial decay of *ctgf* expression can be attributed to adaptation to mechanical stress exerted during initial periods of the experiment since physical and mechanical cues were shown to induce *ctgf* expression together with other soluble signals [Chaqour and Goppelt-Strube, 2006, Dupont et al., 2011]. After such adaptation, both *ctgf* expression and SMAD localization baseline remained constant for hours. Therefore, stimulation experiments were performed 1-2 hours after imaging started when this mechanotransduction effect diminished, providing a flat baseline.

Single cells exhibited substantial heterogeneity even in basal *ctgf* expression levels which may be attributed to both the stochastic nature of its transcription [Suter et al., 2011] and involvement of *ctgf* in different cellular processes [Samarakoon et al., 2010]. Therefore, it was checked if initial SMAD4 expression levels have any influence on *ctgf* expression level in the absence of stimuli. Therefore, we designed an experiment in which doxycycline induction of SMAD4 levels was utilized (Figure 3.8b). The experiment was performed with low initial ectopic expression levels of SMAD4, thereby only expressing endogenous SMAD4, using no doxycycline as characterized before (Figure 3.7). Right before imaging, addition of 20ng/ml doxycycline resulted in a significant amount of increase in the SMAD4 reporter expression, more than one order of magnitude, in SMAD4 levels along 5 hours (Figure 3.8b, upper and lower panels). This marked increase, which would also possibly level up the total SMAD4 abundance in cells, did not lead to any response or expression level increase in *ctgf* expression (Figure 3.8b, middle and lower panels). Similar results were also obtained with Nluc-SMAD2 cell line. This observation suggests that SMAD2/4 abundance alone does not induce any response in the target gene in the absence of stimulation.

3.1. Simultaneous monitoring of transcription factor signaling and target gene activation in single living cells

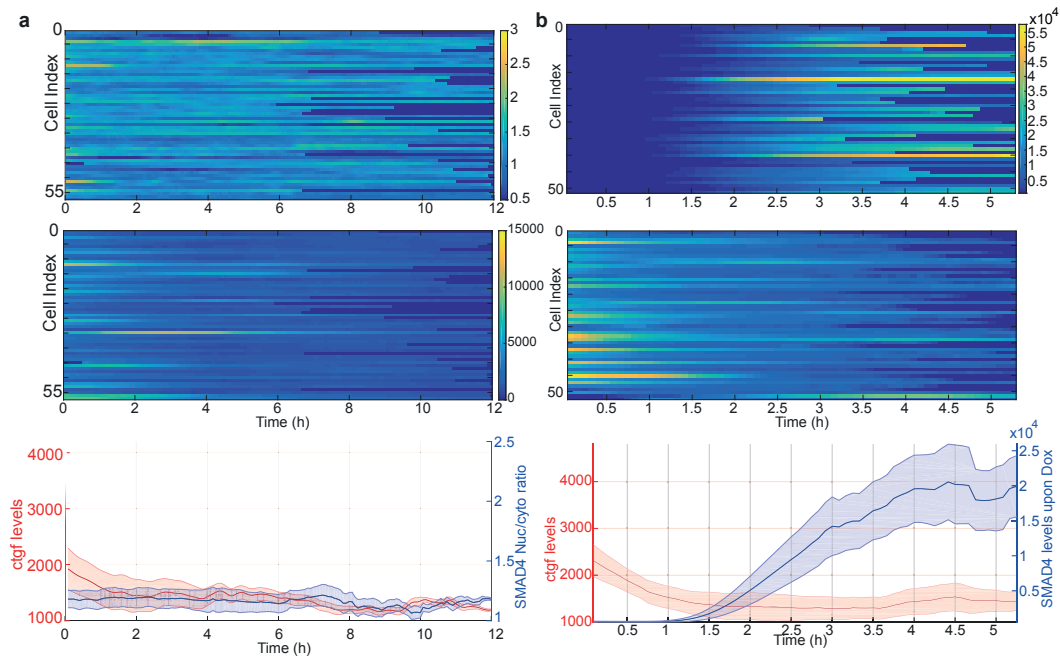


Figure 3.8 – Dual-luminescence reporter system control experiments. *a)* Unstimulated cells reveal flat baseline of SMAD localization and *ctgf* expression. Heat maps of SMAD4 (upper) and *ctgf* expression (middle) levels in unstimulated individual cells over time were shown. Population average shows baseline of SMAD4 nuc/cyto ratio localization of SMAD4 (blue) and *ctgf* expression levels (red, bottom $N=58$). *b)* SMAD4 levels alone do not induce any expression response in the absence of stimulation. Heat map of SMAD4 (upper) and *ctgf* expression (middle) levels in individual cells over time. Cells were treated with 20 ng/ml doxycycline at the time point indicated by dashed black line. They were imaged up to 6 hours but not longer due to high expression levels of SMAD4 reached. Population average shows the significant induction of SMAD4 levels (blue) and steady profile of *ctgf* expression levels (red, bottom $N=52$).

After validation of the dual reporter system and characterizing its behaviour in case of no stimulation, single-cell time-lapse bioluminescence experiments monitoring SMAD translocations and the target gene expression in response to stimulation were performed (Figure 3.9). To do that, cells were imaged with luminescence microscopy up to 24h, upon a high dose (5 nM) of TGF- β stimulation (Figure 3.9). To verify the activation of SMAD signal transduction pathway and to relate time-courses of reporter cell lines to endogenous SMAD dynamics, a combination of western blot analysis and immunofluorescence experiments was performed (Figure 3.11).

Luminescence microscopy movies were analyzed using a custom-made automated image analysis pipeline in Matlab (Section 2.3). With this tool, all individual cell nuclei are segmented and tracked throughout the entire experiment using nuclear Fluc signal. Using segmented nuclei regions, *ctgf* expression signal, localized to nucleus via nuclear localization signal, is calculated. For SMAD translocation quantification, nuclear and cytoplasmic SMAD signal is measured separately and nuclear/cytoplasmic ratio is calculated. When quantified in single cells, levels of nuclear and cytoplasmic SMAD signal varied. Upon stimulation, cells with higher basal levels of nuclear SMAD4 responded more strongly, reaching a higher final level while cells with lower basal levels of nuclear SMAD4 responded less in absolute levels.

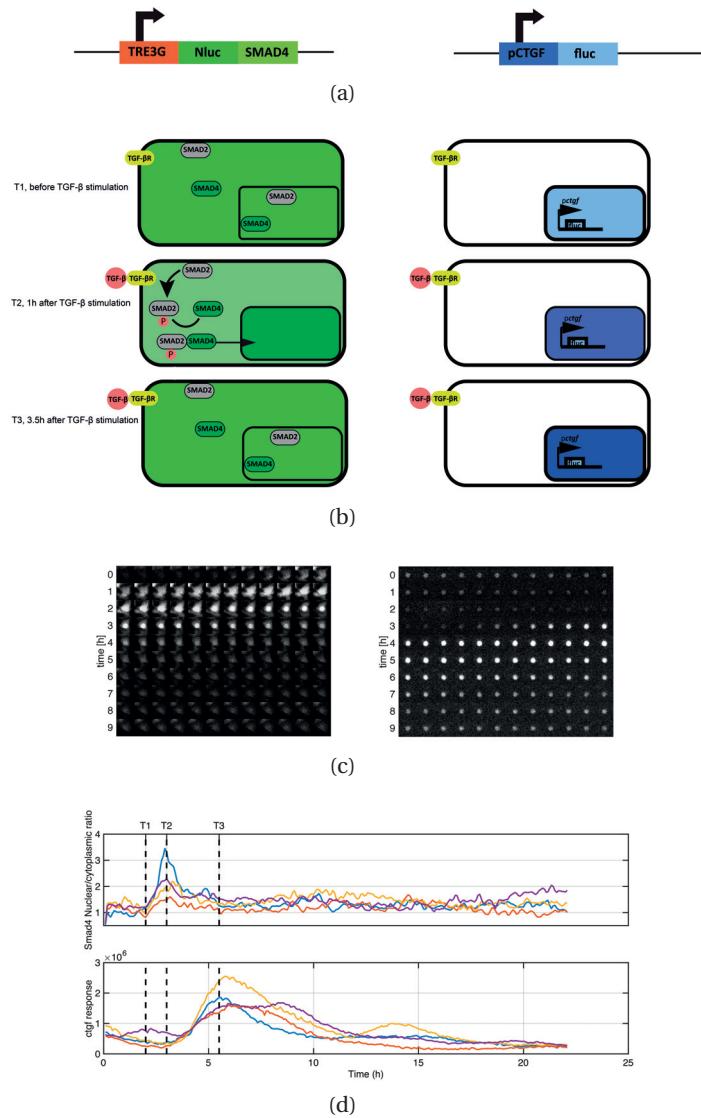


Figure 3.9 – Single cell quantification of dual-luminescence imaging movies in TGF- β stimulation experiments. *a*) Nanoluciferase is fused to SMAD4 and expressed under the control of doxycycline inducible promoter TRE3G (left). Using a gene trap lentivector, short-lived NLS-luciferase is integrated into endogenous locus and expressed under endogenous promoter (right). Arrows represent promoters. *b*) Before TGF- β stimulation, SMAD4 is localized both in the nucleus and cytoplasm (T1) and *ctgf* is expressed at a basal level. TGF- β stimulation results in a transient nuclear enrichment of SMAD4 signal while nucleary localized *ctgf* expression signal increases. T2 and T3 represent time points 1h and 3.5h after TGF- β stimulation at which nuclear SMAD localization and *ctgf* response reach their maxima respectively. *c*) Time series images of a tracked single cell on both Nluc (left) and Fluc (right) channel in the luminescence microscopy. Time goes from top left to bottom right, each of the 10 rows represent 1 hour. Images were taken every 5 minutes. *d*) Quantification of SMAD4 translocation (upper panel) and *ctgf* expression (lower panel) dynamic profiles in four individual cells stimulated with TGF- β (5 nM) at time T1.

Therefore, nuclear/cytoplasm SMAD ratio, rather than the absolute signal, is a reasonable choice against variable levels of transgene expression (Figure 3.9c and Figure 3.9d). Especially

3.1. Simultaneous monitoring of transcription factor signaling and target gene activation in single living cells

when the reporter is expressed at subendogenous levels (2ng/ml dox, Figure 3.7), the ratio quantification can be a good proxy for the assessment of SMAD translocation levels upon stimulation. Indeed, cells displayed a more precise response to ligand stimulation in the ratio quantification (Figure 3.10). Therefore, nuclear enrichment relative to cytoplasmic levels, either SMAD nuclear/cytoplasmic ratio or nuclear minus cytoplasmic levels, are used further in this study to assess the translocation response to stimulation.

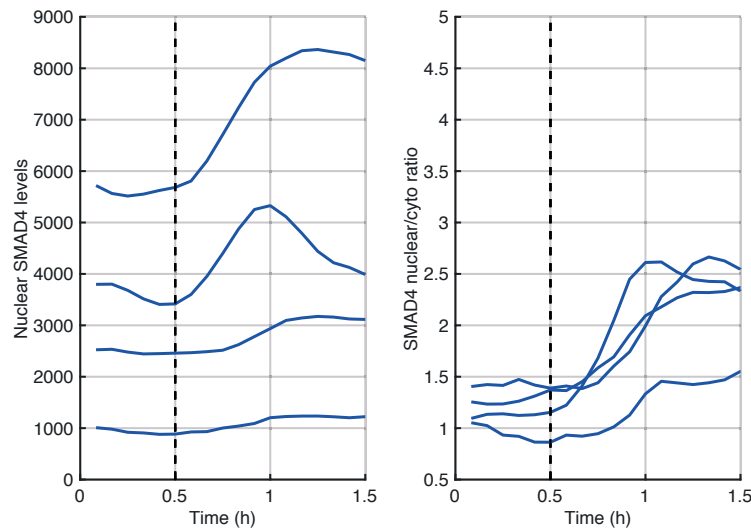


Figure 3.10 – Nuclear/cytoplasmic SMAD ratio is a more precise response to ligand stimulation. The dashed line indicates when TGF- β was added. Each line is a trace from a single cell, plotted as the absolute nuclear level (left) or relative to its cytoplasmic level (nuclear/cytoplasmic ratio, right).

Immunofluorescent staining for endogenous SMAD4 in parental cell line displayed strong overlap of SMAD4 localization and translocation profile upon stimulation as the reporter (Figure 3.11a and b). Slight differences between endogenous profile obtained from immunofluorescence quantification and SMAD reporter profiles in single-cell luminescence imaging may source from the distinct nature of experiments and different image analysis pipelines used for quantification (Section 2.6 and Section 2.3). Similarly, Western blot analysis comparing pSMAD2 levels as a function of time showed the transient induction profile agreeing with SMAD4 translocation profile measured in reporter cell lines (Figure 3.11c and d).

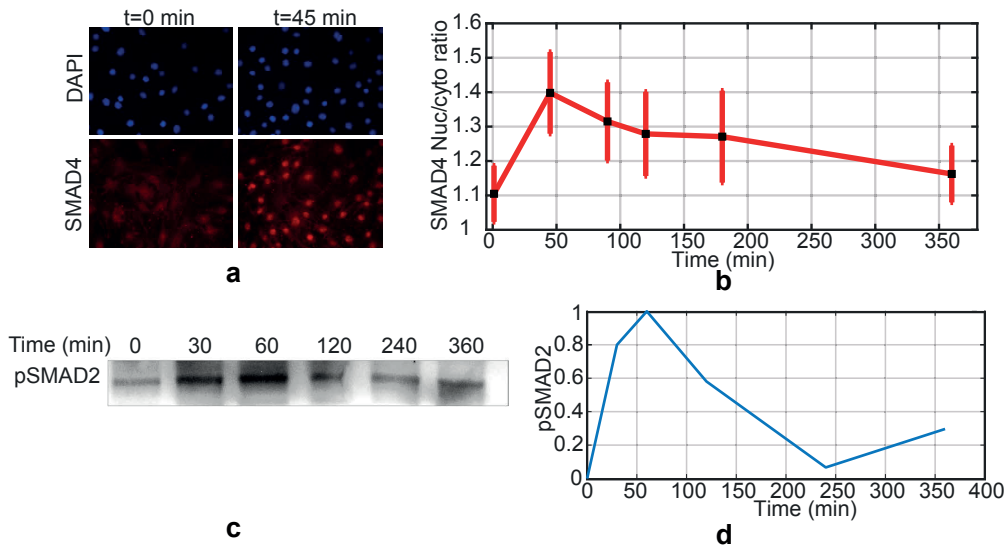


Figure 3.11 – Validation of the reporter system using immunofluorescence and western blot. a) Immunofluorescent staining for SMAD4 at the indicated times after continuous application of 5nM TGF- β b) showing transient enrichment of endogenous SMAD4 in the nucleus characterized by SMAD4 nuclear/cytoplasmic ratio at different time points (typically ~100 cells for each time point). c) Western blots of total cell lysates from cells exposed to 5nM TGF- β for the indicated times show transiently elevated pSMAD2 levels (exogenous pSMAD2 is not detected due to low doxycycline induction (2ng/ml)). Signals were normalized to GAPDH staining in each well. The sample at t=120 min had a low loading problem which was still corrected with GAPDH normalization. d) GAPDH normalized signals were further normalized with min-max scaling assuming that sample was at the low basal limit before TGF- β stimulation (performed in monoplicate).

It is also verified that the localization of SMADs and ctgf response levels were both dependent on the stimulation. In contrast to unstimulated cells or cells treated with inhibitor SB-431542 (Figure 3.12 and Figure A.1), we observed robust nuclear shuttling of SMAD4 and SMAD2 and the subsequent transcriptional response of ctgf upon TGF- β stimulation. Therefore, dual-color luminescence imaging allows simultaneous recording of transcription factor shuttling and transcriptional responses at high temporal resolution.

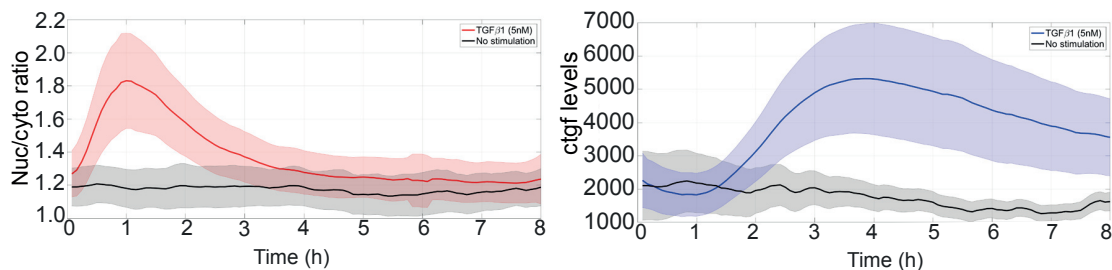


Figure 3.12 – Dual-luminescence system performance in single-cell TGF- β stimulation experiments. After quantification of the time-lapse movies, a) the ratio of nuclear-to-cytoplasmic SMAD4 ratio (red) and b) ctgf expression (blue) are plotted as a function of time. The response of cells with the functional reporter and stimulated with 5 nM TGF- β ($N = 301$) was compared to unstimulated cells ($N = 58$, gray). The solid lines indicate the population average of single cells and the shaded area the standard deviation. Data is shown up to 8 hours to emphasize activation both in SMAD4 and ctgf.

Similarly, signal-to-noise ratios for newly established SMAD translocation reporters

3.2. Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

were calculated based on single cell SMAD translocation traces (Figure 3.13). High SNR even with modest levels of translocation indicates that reporters are suitable to measure SMAD translocation in response to stimuli. Taken together, these results suggest that we have established a functional dual-luminescent reporter system to measure SMAD mediated TGF- β pathway at two different nodes.

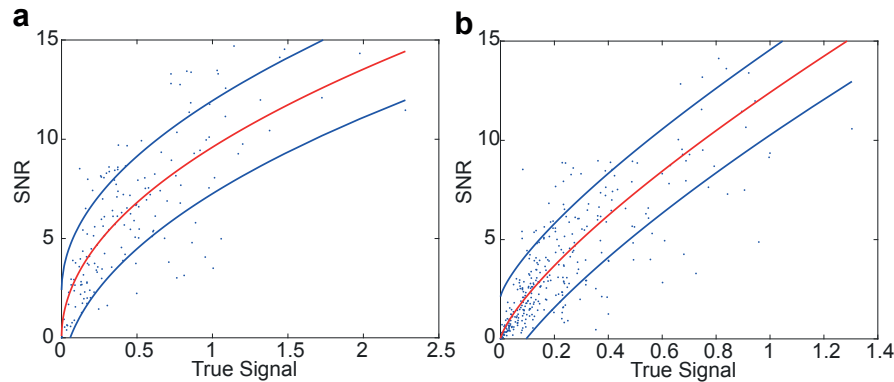


Figure 3.13 – Signal-to-noise ratio assessment of SMAD translocation reporter measurements for a) SMAD4 and b) SMAD2. Each data point represents a single cell. The signal on x-axis is the amplitude of the translocation peak compared to basal SMAD nuclear/cytoplasmic ratio. The SNR was calculated as the maximum ratio change, normalized to the baseline ratio, divided by the standard deviation of baseline signal levels (5 frames) before TGF- β stimulation. The data were fitted by using the relationship $SNR = Signal^{1/2}$. Blue lines indicate the 95% confidence intervals.

3.2 Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

Single cells in previous stimulation experiments responded to the stimulation with a profound cellular variability (Figure 3.9, Figure 3.11, Figure 3.12). Both SMAD translocation and *ctgf* response varied from cell to cell temporally and in amplitude. To assess such heterogeneity thoroughly, TGF- β stimulation experiments were repeated several times to reach enough statistical power, typically hundreds of cells for each condition. Next, the variability in the response and the contribution of SMAD signaling on *ctgf* response heterogeneity were investigated.

Using the described single-cell quantification methodology, we extracted single-cell information from repeated experiments and combined hundreds of cells (Figure 3.14, typically ~30 cells from each imaging). For each single cell (~300 cells), SMAD4 translocation and *ctgf* response were quantified and displayed in two separate heat maps where each row corresponded to the same individual cell (Figure 3.14a). Single cell traces were min-max scaled, meaning that the data was scaled to a fixed range, 0 to 1, to indicate different levels of activation in the same scheme. Yet, absolute quantification of SMAD translocation and *ctgf* expression is utilized for further analysis (Figure 3.14b).

To investigate also temporal features of the response, certain time points on single cell traces such as SMAD translocation peak, the target gene response initiation and the target gene

Chapter 3. Results

response peak were manually annotated. Temporal distributions of the single cell traces were checked to assess the variability of each feature. The quantification of nuclear to cytoplasmic ratio of the SMAD4 translocation reporter signal revealed an immediate translocation response into nucleus which directly starts at 5 min and reaches a peak about 1h after TGF- β stimulus (Figure 3.14a, upper panel and Figure 3.14c, upper panel, green). This suggests the immediate phosphorylation of SMAD2/3 (Figure 3.11), their interaction with SMAD4 and formation of R-SMAD-SMAD4 complexes which leads to its translocation from the cytoplasm to the nucleus.

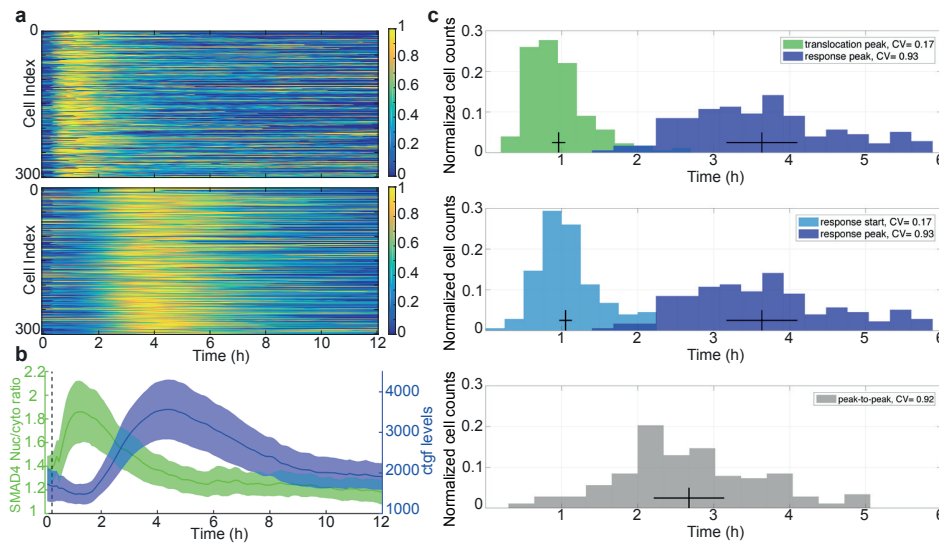


Figure 3.14 – Single cell temporal variability of SMAD4 translocation and ctgf expression dynamics. *a*) Heat map of single cell SMAD4 translocation (upper panel) and the consequent ctgf expression (lower panel) upon TGF- β stimulation (5nM). Results are a combination of several stimulation experiments. Min-max scaling, ranging from 0 to 1, was used to represent each single cell data. *b*) Population average of the data shown in heat maps with standard deviation (SMAD4 nuclear to cytoplasmic ratio (blue) and ctgf signal (green)). *c*) Nuclear localization of SMAD4 reaches a peak with a mean temporal delay of ~60 min after TGF- β stimulation (green). The target gene expression starts to increase with a temporal delay of ~60 min after TGF- β stimulation (blue). The target gene expression reaches a peak with a temporal delay of ~210 min after TGF- β stimulation (purple). The distribution of the time delay between SMAD4 nuclear localization peak and ctgf expression peak (gray, mean ~150min). Legends indicate the coefficient of variation (horizontal cross line) while the vertical cross line displays the mean.

Concurrently, the immediate early gene, ctgf, initiates a response which reaches a maximum expression only 3.5 h from stimulation after SMAD4 localization has already returned to almost its uninduced profile (Figure 3.14a, lower panel and Figure 3.14c, purple). Individual cells displayed relatively little variability in their timing of SMAD4 translocation (Figure 3.14a) but strong variability in its amplitude (Figure 3.15a-c), suggesting variable transduction efficiency of TGF- β signaling to SMAD4 shuttling. Both timing and amplitude of the ctgf transcriptional response displayed a broad distribution, reflecting large cell-to-cell variability in upregulation of the ctgf gene (Figure 3.14c and Figure 3.15b-d).

3.2. Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

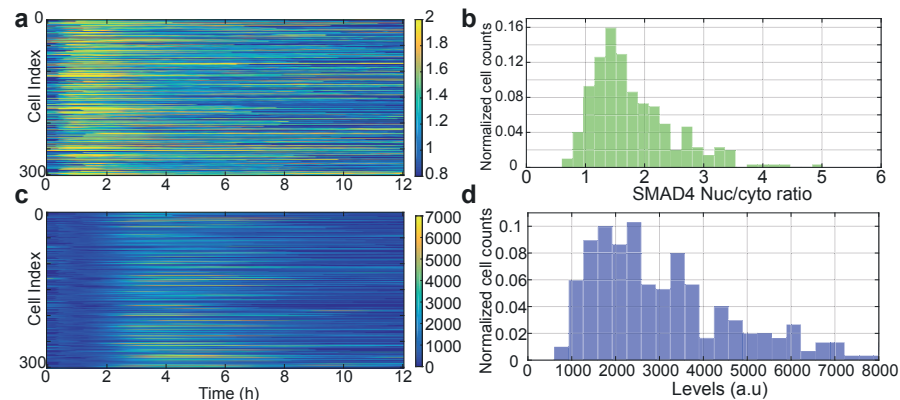


Figure 3.15 – Single cell variability of SMAD4 translocation and *ctgf* expression levels. Heat map of single cell a) SMAD4 translocation and b) the consequent *ctgf* expression upon TGF- β stimulation (5 nM). Data is represented as the absolute single cell quantification. c) Single-cell distribution of SMAD4 nuclear to cytoplasmic ratio calculated at translocation peak. d) Single-cell distribution of maximal *ctgf* responses calculated by subtracting basal nuclear *ctgf* levels from its peak value in response to TGF- β .

SMAD4 nuclear localization peak temporally varied from 0.5h to 1.5h in individual cells while the gene expression response covered a wider temporal range from 2h to 6h. This variability is also reflected in the peak-to-peak temporal distribution (Figure 3.14c). SMAD translocation process, directly downstream of TGF- β stimulation, encompasses the activation of the pathway on the membrane, SMAD-phosphorylation, heteromerization with co-SMADs and the import of the complex into the nucleus. On the other hand, the gene expression output may have many other sources of variability such as the transcription itself, the export of mRNA, its translation and the feedback mechanisms shaping the response further.

SMAD4 undergoes continuous nucleocytoplasmic shuttling prior to TGF- β signaling which determines its basal localization [Pierreux et al., 2000]. SMAD4 accumulation in the nucleus peaking at 1h is due to formation of complexes with the R-SMADs, SMAD2 and SMAD3. This increase in nuclear localization is attributed to the increased interaction with DNA and reduced export levels. R-SMAD dephosphorylation results in dissociation of SMAD4 from complexes and export from the nucleus to the cytoplasm. Regarding localization quantification, SMAD4 localization before induction seemed to be homogeneous all around the cell which may slightly deviate from the previous reports [Warmflash et al., 2012]. Such deviation could be explained with differences such as expression levels of SMAD4 reporter, cell types used, single-cell quantification strategy and the type of microscopy used in different studies. In addition, single cells displayed intercellular variability in SMAD localization prior to induction. Such variability could arise from varying basal activity of the pathway or from other extrinsic sources of noise such as cell cycle, cellular microenvironment leading to different SMAD4 localization in individual cells [Strasen et al., 2018].

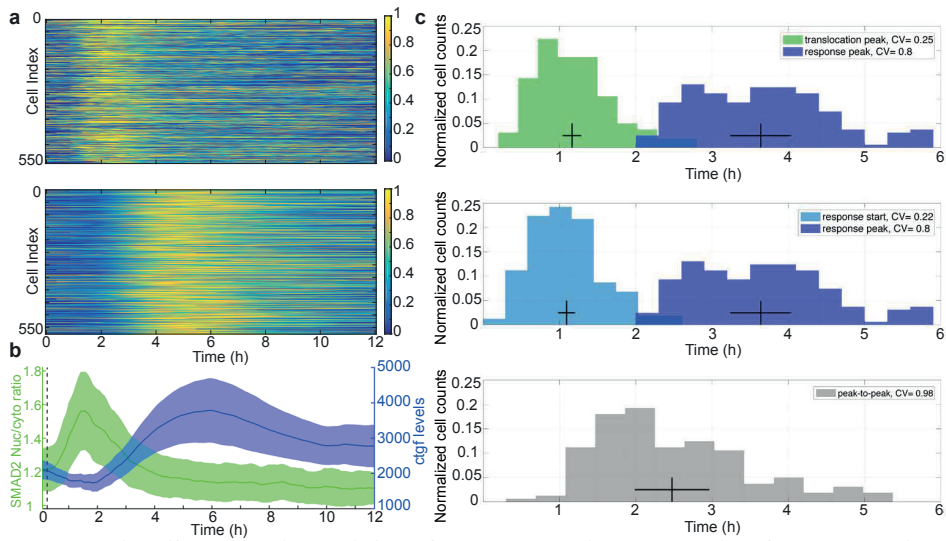
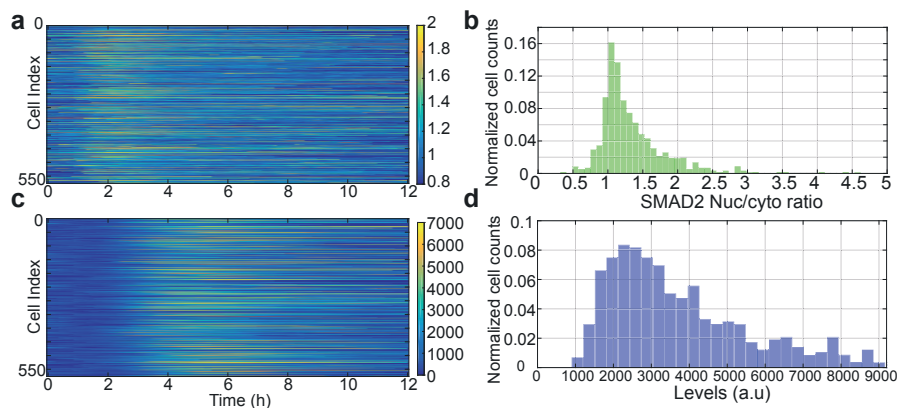


Figure 3.16 – Single cell temporal variability of SMAD2 translocation and ctgf expression dynamics. a) Heatmap of single cell SMAD2 translocation and the consequent ctgf expression upon TGF- β stimulation (5 nM). Min-max scaling was used to represent each single cell data. b) Population average of the data that was shown in heat maps with standard deviation (SMAD2 nuclear to cytoplasmic ratio (green) and ctgf signal (blue)) c) Nuclear localization of SMAD2 reaches a peak with a temporal delay of ~70 min upon TGF- β stimulation (green). The target gene expression starts to increase with a temporal delay of ~70 min after TGF- β stimulation (blue). The target gene expression reaches the peak with a temporal delay of ~210 min after TGF- β stimulation (purple). The distribution of the time delay between SMAD2 nuclear localization peak and ctgf expression peak (gray, mean ~150min). Legends indicate coefficient of variation (horizontal cross line) while vertical cross line displays the mean.

In parallel, we also conducted the same experiments to assess the partner signal transducer (SMAD2), behaviour using Nluc-SMAD2 translocation reporter cell line (Figure 3.16 and Figure A.2). Since the activated SMAD4-SMAD2 complexes translocate together into nucleus, similar translocation dynamics and gene expression response profiles were obtained as SMAD4 clone. SMAD2 nuclear localization reached the peak around 70 min which coincided with the initiation of the expression response. The ctgf response peaked at 3.5-4 hours after stimulation.



3.2. Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

Figure 3.17 – Single cell variability of SMAD2 translocation and the *ctgf* expression levels. Heat map of single cell a) SMAD2 translocation and b) the consequent *ctgf* expression upon TGF- β stimulation (5nM). Data represents the absolute single cell quantification. c) Single-cell distribution of SMAD2 nuclear to cytoplasmic ratio calculated at translocation peak. d) Single-cell distribution of maximal *ctgf* responses calculated by subtracting basal nuclear *ctgf* levels from its peak value in response to TGF- β .

Average SMAD2 and SMAD4 translocation dynamics upon stimulation agreed with the biochemical measurements in previous studies [Vizán et al., 2013] while we observed profound heterogeneity of SMAD translocation in single-cells. However, the gene expression response showed a higher cell-to-cell variability compared to SMAD translocation throughout the experiment (Figure 3.14 to Figure 3.17). SMAD localization has relatively low variability in the basal state while it reaches higher values after TGF- β stimulation due to differential responses in individual cells. On the other hand, *ctgf* expression levels show substantial heterogeneity even in basal state which can be due to both its high responsivity to cellular microenvironment (Samarakoon et al. [2010], Section 1.2.6) and its bursty nature even in steady-state [Suter et al., 2011]. Its expression can be stimulated by shear stress, serum shock, or TGF- β in cultured fibroblasts [Samarakoon et al., 2010, Molina et al., 2013]. Therefore, conditions such as serum concentration, cellular density contribute to its variability. However, TGF- β stimulation causes a synchronized response of individual cells which reflects into lowered, yet still higher than SMAD, cellular variability (Figure 3.14 and 3.16).

3.2.1 *ctgf* dynamics decompose into two distinct classes: transient and sustained

Single cells displayed substantial amount of cell-to-cell variability in SMAD translocation and *ctgf* expression response both in dynamics and amplitude. Inspection of individual cells revealed a more detailed profile of *ctgf* responses. The majority of individual *ctgf* responses were nicely defined transients, reaching a peak in 3-4 hours upon stimulation, termed as ‘initial response’, and returning to basal levels in around 8 hours (Figure 3.18, blue traces) while some portion of the whole population differed from such profile in their responses. These individuals responded in a sustained manner rather than transient as defined earlier, characterized by a relatively weaker initial response and a longer-lasting activity. In this subpopulation, cells typically also displayed a distinct second wave of response, termed as ‘final response’ for the rest of the study, before returning to their basal values (Figure 3.18, red traces) while this final response was temporally less synchronous among individuals. Such sustained or oscillatory profiles, previously reported to be possible in TGF- β induced SMAD signaling [Strasen et al., 2018], have not been clearly investigated in relation to its consequent target gene responses so far. Therefore, further analyses were crucial to clearly define individuals belonging to these two distinct sub-populations.

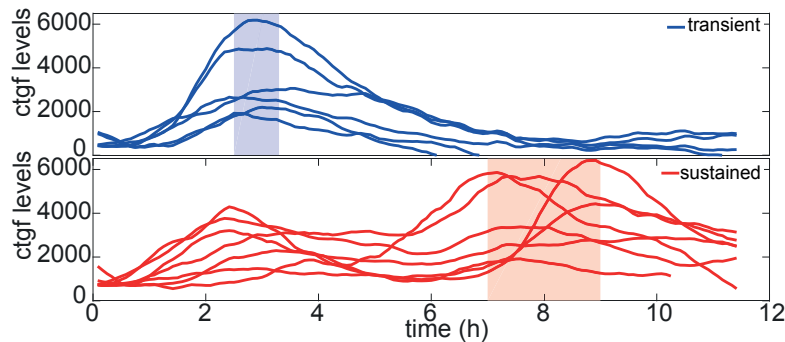


Figure 3.18 – Inspection of individual *ctgf* responses revealed two distinct profiles. Sample single-cell *ctgf* trajectories from TGF- β (5 nM, at $t=0$) stimulation experiments are shown ($N=6$ -blue traces, $N=7$ -red traces). Some cells responded in a transient manner (upper panel, blue single cell traces) with only ‘initial response’ (blue shaded region) while the others responded in a sustained manner (lower panel, red single cell traces), typically and additionally exhibiting a ‘final response’ (red shaded region).

In order to dissect these qualitatively different *ctgf* responses, individual *ctgf* traces were categorized into two classes using k-means clustering. Due to pronounced cell-to-cell variability in absolute *ctgf* levels and the weakness of Euclidean metric to resolve dynamic behaviour of traces, we did not use Euclidean distance as the similarity/dissimilarity measure. Instead, we use correlation as the k-means distance metric to determine subgroups with similar dynamic patterns. Moreover, previous qualitatively observed distinct classes mainly differentiate in their dynamics rather than amplitudes. We, therefore, decided to employ two clusters to represent best the dynamical heterogeneity observed. Constraining the number of clusters to two would make our results remain biologically significant since some previous studies also claimed two main response groups in TGF- β induced SMAD signaling, transient and sustained [Zi et al., 2011, Strasen et al., 2018]. Moreover, categorizing single cells into bigger numbers of classes would reduce the statistical power by further lowering number of cells in each cluster.

Indeed, we identified two cellular subpopulations of *ctgf* traces that show qualitatively distinct behaviour as observed manually earlier (Figure 3.19). Figure 3.19a displays the mean *ctgf* response of cells in the identified categories (solid lines) and the variability in that category (error bars). The first group exhibited a transient behaviour returning back to its basal values in around 8 hours (Figure 3.19a, blue) while the second group showed sustained and even oscillatory response (Figure 3.19a, red). The sustained population represents only a small portion of all individuals (13%-sustained vs 87%-transient).

As stated earlier, we made a heuristic choice of two clusters when applying *k-means* algorithm as the initial choice since our initial observations of *ctgf* responses in different samples gave us the notion of two classes although we do not exclude the possibility of having more number of clusters. Yet, our clustering approach, even with two classes, provided promising results when the performance of the clustering is evaluated using silhouette score. Silhouette plots provided a graphical representation of how well *ctgf* expression response

3.2. Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

of each cell belonging to a cluster corresponded to other cell traces in the same cluster rather than the other cluster [Rousseeuw, 1987]. The silhouette score ranges from -1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. The silhouette scores are calculated using 1-correlation as the distance metric. Results showed a very high percentage of cells with positive silhouette scores suggesting the nice separation of clusters (Figure 3.19b).

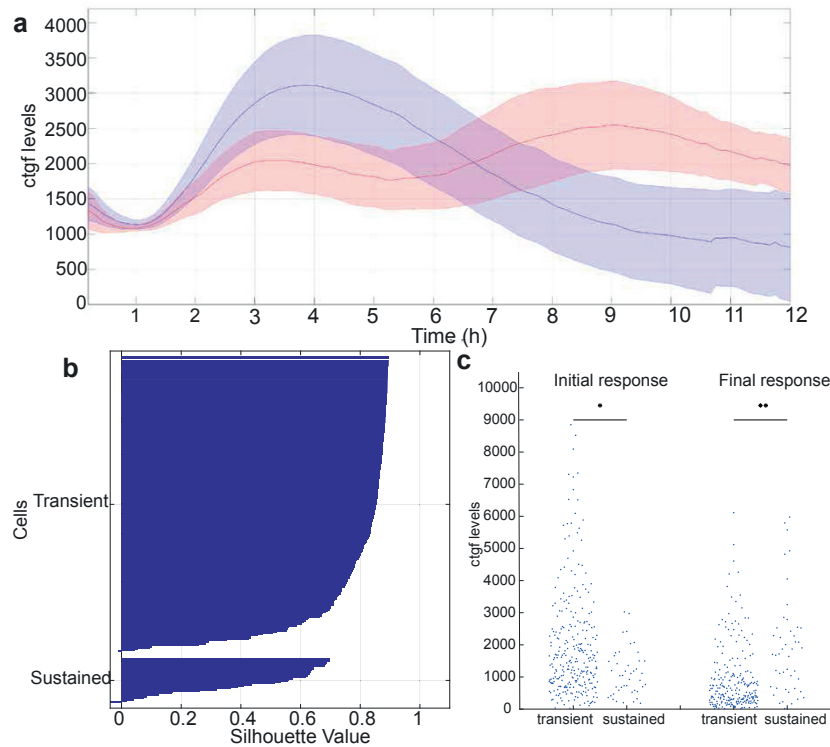


Figure 3.19 – Ctgf response dynamics decompose into two distinct response classes. a) Single-cell traces from TGF- β (5 nM) stimulation experiments (Nluc-SMAD4 cell line, as in Figure 3.14) were categorized into two classes using k-means clustering (N=301 cells). Time courses of the identified response classes showed qualitatively distinct behavior. The solid lines indicate the mean of the cell population and the shaded area the standard deviation of the mean. b) Silhouette plot of cells sorted according to ctgf expression dynamics. Plots represent how similar each cell to its own cluster (in our case, transient and sustained) compared to other clusters. Positive silhouette scores indicate that ctgf responses are more similar to their own group while negative scores suggest that the corresponding trace is closer to the other group. c) Heterogeneity in initial and final ctgf response levels, quantified at the peak point, for transient and sustained classes. ($p < 0.05$; ** $p < 0.0005$, t-test).*

While predominantly distinct in their dynamics, identified subpopulations also differed in their absolute levels of responses. Cells belonging to the sustained cluster displayed slightly, but still significant, lower levels of the initial response (Figure 3.19c). This observation might already suggest an existing differential cellular state among individuals that already reflects into the initial response level. As expected, the same sustained subpopulation showed higher expression levels compared to the transient category, at longer time scales (in between 8-12 hours). Taken together, both the response levels and dynamics differed in two identified

clusters indicated by absolute and normalized (min-max scaled) signals (Figure 3.20).

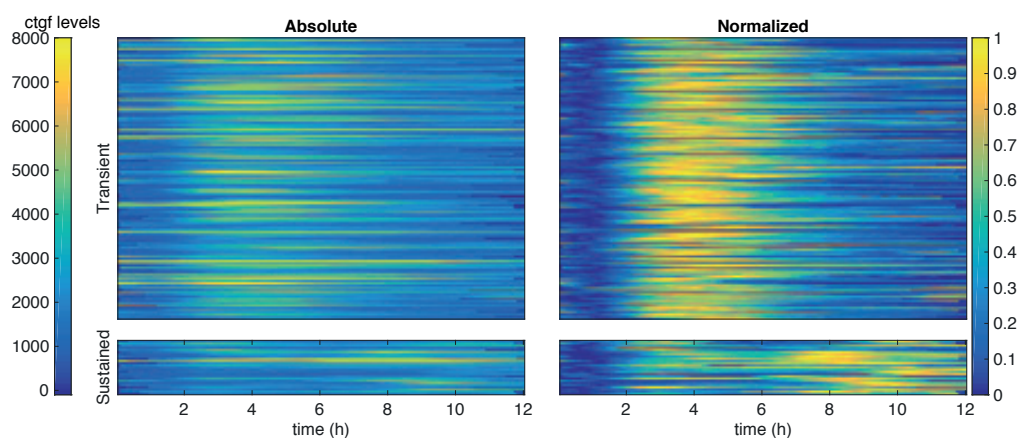


Figure 3.20 – Decomposition of cell-specific *ctgf* responses into transient and sustained classes. Heatmaps for *ctgf* traces belonging to transient (upper panels) and sustained (lower panels) are shown, quantified as both absolute (left) and normalized (right) levels (min-max scaled, 0 to 1). TGF- β (5 nM) is applied at $t=0$ and *ctgf* responses in *Nluc-SMAD4* cell line is quantified (Dox concentration, 2ng/ml).

The same clustering approach is also applied in *Nluc-SMAD2* cell line, on the single-cell data of TGF- β stimulation experiments (Figure 3.16) which indicated similar results (Figure A.3 and Figure A.4, TGF- β (5 nM), doxycycline concentration (2ng/ml)). Both the mean profile of identified subpopulations and the percentage of sustained category did not change much which suggests that identified distinct behaviours are independent from the reporter cell line type.

3.2.2 Long-term dynamics of *ctgf* response is influenced by SMAD4, but not SMAD2, levels

After decomposing responses into two distinct classes, we aimed to determine if SMAD signaling has any influence on the distribution of traces into those classes. Thus, we hypothesized that SMAD signaling components, SMAD4/2 in our case, may have certain variability among individuals such that these differences can manifest themselves in distinct *ctgf* responses. To check such relation, single-cell SMAD signaling traces obtained from dual-luminescence imaging measurements were utilized. Interestingly, both SMAD4 and SMAD2 translocation dynamics did not display a significant difference among transient and sustained *ctgf* response classes (Figure 3.21, left). In addition, even in case of a sustained *ctgf* response, cells belonging to this class showed a transient SMAD signaling upon stimulation. However, SMAD4, but not SMAD2, levels differed significantly between two classes which may suggest that high SMAD4 levels may contribute to sustainability of the response in certain proportion of cells (Figure 3.21, right).

3.2. Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

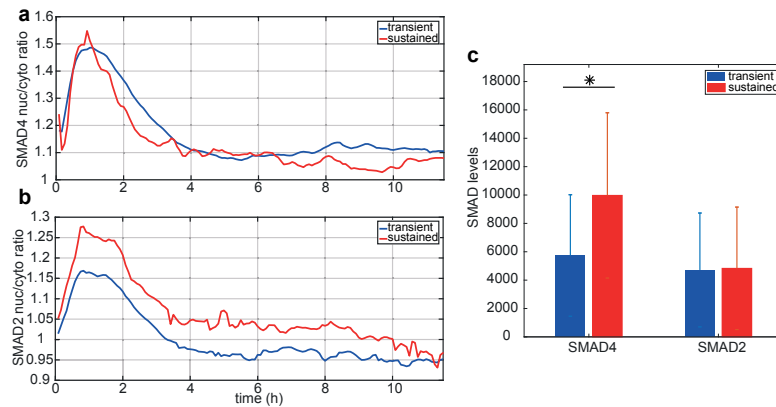


Figure 3.21 – Not SMAD shuttling dynamics but SMAD4 levels differ between transient and sustained *ctgf* response to TGF- β stimulation. Average SMAD nuclear to cytoplasmic ratio trajectories belonging to two classes identified with clustering (blue-transient, red-sustained), in *Nluc-SMAD4* (a) and *Nluc-SMAD2* (b) cell lines. c) Comparison of SMAD levels in the transient and the sustained classes. SMAD4 levels are significantly higher in the sustained class while SMAD2 levels do not differ. SMAD levels are quantified as the mean signal during 12 h after TGF- β (5 nM) stimulation ($t=0$). (* $p<0.005$; t -test)

To further validate the identified link between SMAD4 levels and the target gene responses to TGF- β stimulation, we designed experiments to vary SMAD levels in reporter cell lines by exploiting the doxycycline-inducible promoter. We repeated the previously described TGF- β stimulation experiments with the same high dose of TGF- β stimulation (5 nM) while using a range of doxycycline concentrations (0-200 ng/ml) (Figure 3.22). This allowed us to vary SMAD reporter levels from subendogenous levels to over expressed regime as characterized before (Figure 3.6 and Figure 3.7).

Population averages obtained from single-cells, treated with increasing doses of doxycycline, revealed a pattern in long-term *ctgf* response dynamics. With the increasing SMAD4 levels, population averages showed more sustained and even oscillatory *ctgf* responses upon stimulation (Figure A.7). This was better visualized when *ctgf* responses were normalized using min-max scaling. Starting from 20 ng/ml doxycycline treatment, population averages for *ctgf* response becomes less transient and more sustained. Higher-doses of doxycycline treatment (50-200 ng/ml), hence higher SMAD4 levels, resulted in even oscillatory responses (Figure 3.22a). On the other hand, SMAD2 samples did not exhibit such behaviour since all samples, irrespective of their induced SMAD2 levels, responded in a transient manner (Figure 3.22b). These observations agree with our previous Lumicycle measurements revealing also differential effect of SMAD2 and SMAD4 levels on the target gene response dynamics upon

TGF- β stimulation (Figure 3.6).

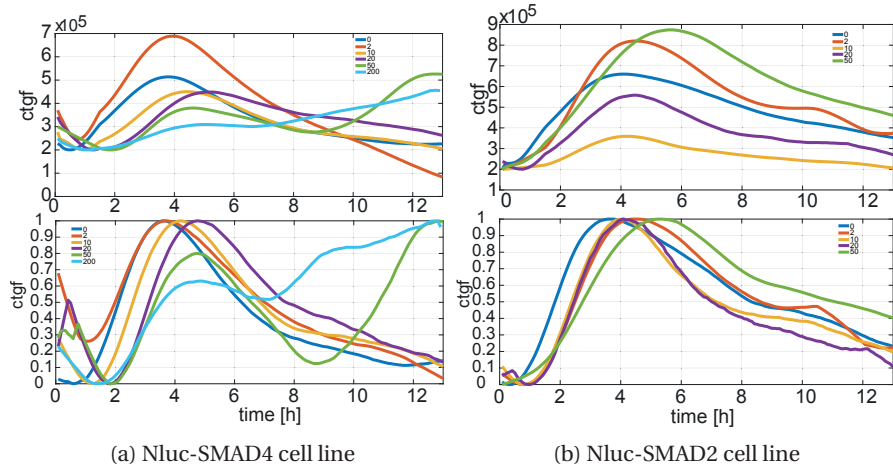


Figure 3.22 – Long-term dynamics of ctgf response is influenced by SMAD4, but not SMAD2, levels. Population time courses for ctgf responses upon TGF- β stimulation in samples treated with varying doses of doxycycline (0-200 ng/ml). TGF- β stimulation is at $t=0$. a) Nluc-SMAD4 cell line showed distinct long-term ctgf dynamics depending on SMAD4 levels (upper panel). Traces were normalized using min-max scaling (lower panel). b) In Nluc-SMAD2 cell line, no difference was found between samples regarding long-term dynamics (absolute traces-upper panel, normalized traces-lower panel).

To dissect further the identified link occurring in SMAD4, but not in SMAD2 samples, single-cell time series data obtained from different doxycycline dose treatment experiments were pooled for k-means clustering approach described earlier. This allowed us to analyze higher number of cells ($N=1576$) and gave a complete picture of response profiles, ranging from strongly transient ones to sustained and even oscillatory profiles. Similar to previous observation, the clustering revealed two distinct classes mainly differing in their long-term dynamics (Figure 3.23). Due to inclusion of many replicates of TGF- β stimulation experiments in different doxycycline conditions led to a wide range of ctgf response levels in absolute values (Figure 3.23, left). Yet, normalized profiles better revealed the underlying heterogeneity of dynamics detected (Figure 3.23, right). Of note, data represented in heatmaps display where the most striking difference lies in between two classes, within 12 hours after TGF- β stimulation. Yet, in later time points, the sustained class individuals also returned to their basal levels with varying dynamics (Figure A.6). Still, the sustained profile, which can also be considered as two waves of activation, differs from the typically expected transient response, thereby needs further investigation.

3.2. Cellular heterogeneity in transcription factor translocation and the target gene response dynamics

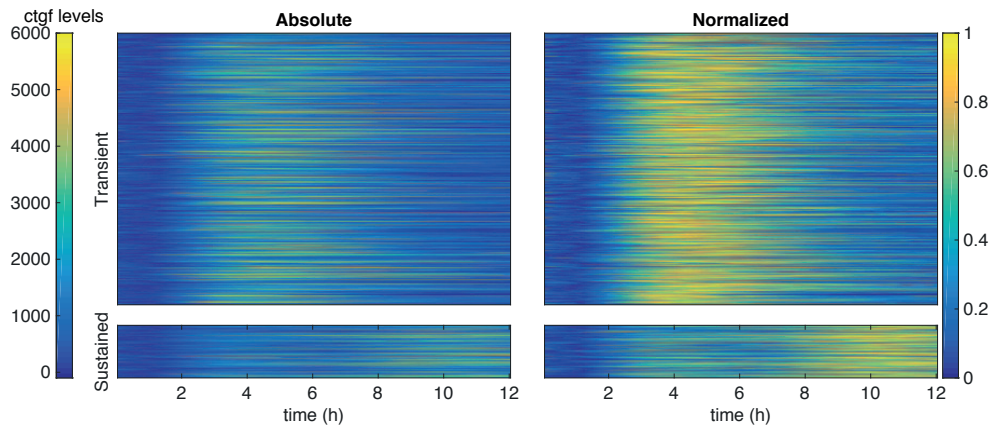


Figure 3.23 – Decomposition of cell-specific ctgf responses into transient and sustained classes. Cells treated with varying doxycycline concentrations as indicated in Figure 3.22 were all included in the analysis ($N_{total} = 1576$, $N_{transient} = 1213$ (77%), $N_{sustained} = 363$ (23%)). Heat maps for ctgf traces belonging to transient (upper panel) and sustained (lower panel) are shown using absolute (left) and normalized (min-max scaling, 0 to 1, right) levels of ctgf responses.

Neither population level measurements nor average signals of individual cells decipher the relation between SMAD levels and the response dynamics. Thus, this link is checked in single-cell traces by determining the occurrence, fraction of cells, of sustained dynamic profiles based on doxycycline induction levels. The clustering approach provided a clear separation of two classes (Figure 3.24a). In addition, distribution of each classes into varying doxycycline dose samples gave some clue about contribution of SMAD levels on such categories of target gene responses. Increased doses of doxycycline, consequently higher SMAD4 levels, altered the portion of sustained individuals in the population, from 10% to 50% in the most extreme case. The increase in sustained occurrence overlapped with the profile of relative SMAD reporter levels compared to endogenous counterparts. Starting from doxycycline concentration of 20 ng/ml at which SMAD4 reporter levels reached and even exceeded endogenous levels as measured earlier, the occurrence of the sustained portion is altered reaching its highest value (50%) at doxycycline concentration of 200 ng/ml (Figure 3.24b, right and Figure A.8). Yet, SMAD2 levels did not display any regular trend as sustained proportion varied between 10% and 23% in different samples (Figure 3.24b, left).

Chapter 3. Results

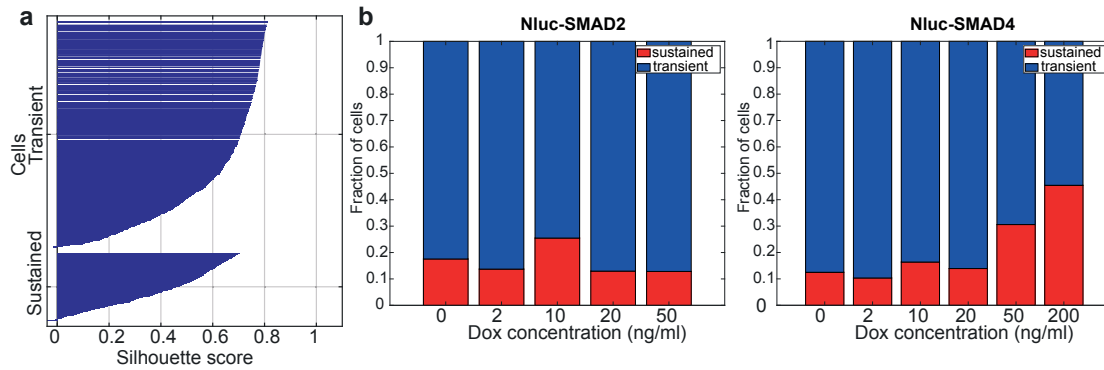


Figure 3.24 – Distribution of *ctgf* response clusters depending on SMAD levels. a) Silhouette plot of cells sorted according to *ctgf* expression dynamics. b) Bar plots indicate the fraction of each cluster (yellow-transient, blue-sustained) in the corresponding doxycycline dose inducing either SMAD2 (left) or SMAD4 (right) levels.

Of note, not all experiments in doxycycline dose analysis have the second channel, Nluc. Only experiments with doxycycline concentrations of 2 ng/ml and 10 ng/ml which allowed us to directly check if, in single cells, SMAD levels have any significant difference between two identified response clusters. Indeed, SMAD4 levels, in both conditions, displayed higher values in the sustained class compared to its transient counterparts (Figure 3.24).

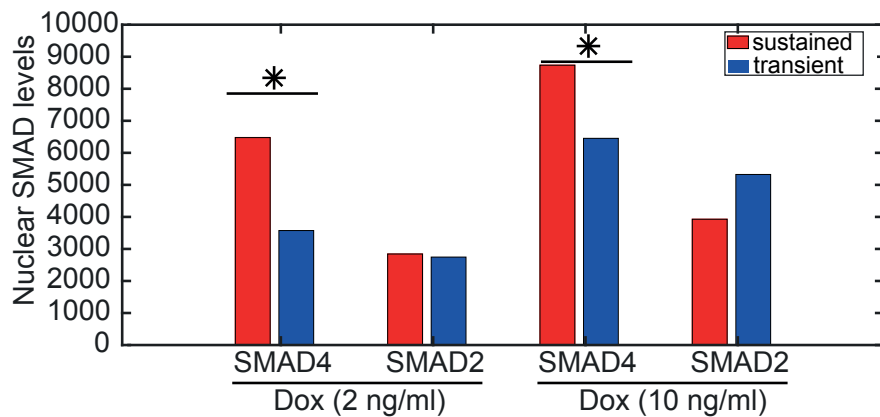


Figure 3.25 – Comparing SMAD levels in the transient and the sustained classes in samples treated with 2 ng/ml and 10 ng/ml doxycycline. SMAD4 levels are significantly higher in the sustained class while SMAD2 levels do not differ significantly. SMAD levels are quantified as the average signal during 12 h after TGF- β (5 nM) stimulation ($t=0$). (* $p<0.05$; t -test)

The experimental approach and analysis of single-cell data in previous sections allowed us to analyze the *ctgf* response in relation to SMAD profile in hundreds of genetically identical reporter cells. Heterogeneous *ctgf* responses were deciphered by the decomposition of traces into two clusters, transient and sustained. Modulating SMAD levels, the influence of SMAD levels on the distribution of cells into the defined classes was determined. It was observed that altered SMAD4 levels led to a shift in response toward sustained profile while SMAD2 levels did not have any impact on the *ctgf* dynamics. To verify if these observations were not sourced

3.3. TGF- β dose modulation on both SMAD signaling and *ctgf* response

by any perturbation due to doxycycline treatment itself [Moullan et al., 2015] and to eliminate any alternative interpretation of the results, a non-inducible doxycycline control cell line was produced. The same type TGF- β stimulation experiments, using this non-functional cell line, did not reveal any differential effect of doxycycline treatment on the *ctgf* response. Both the sample induced with high dose of doxycycline (200 ng/ml) and non-induced sample exhibited similar *ctgf* response amplitude and dynamics in response to TGF- β stimulation (Figure A.5). This non-inducible doxycycline control cell line expressing only rtTA3G without any target gene also did not exhibit altered *ctgf* response, confirming the absence of the perturbation due to reporter system on *ctgf* dynamics.

These results, in overall, agree with our previous observations at population level showing differential effect of SMAD2 and SMAD4 on the long-term *ctgf* dynamics (Figure 3.6). Furthermore, single-cell assessment of both SMAD levels and *ctgf* responses revealed the extent at which SMAD levels influence responses, evaluated by counting the occurrence of different behaviours, which would not be possible with bulk measurements.

3.3 TGF- β dose modulation on both SMAD signaling and *ctgf* response

TGF- β plays a key role in a broad array of cellular activities in a dose-dependent manner. Mammalian cells interpret the concentration of TGF- β in the surrounding environment and translate this information into downstream signals that direct their proliferation, survival or death in different cellular contexts. While some early studies assume constant TGF- β concentration throughout experiments, it has been shown that TGF- β is internalized and degraded by cells. This ligand depletion emerges as one of the main determinants of the downstream signalling [Vizán et al., 2013]. Moreover, units of ligand molecules per cell rather than just the ligand concentration was shown to be a better representation of TGF- β ([Zi et al., 2011]). Therefore, it was crucial to investigate how cells respond to different ligand doses and how changes in ligand exposure influence SMAD signalling and consequent gene expression dynamics.

In previous experiments, a high-dose of TGF- β (5 nM) was used which led to response of all cells both on SMAD and *ctgf* level to different extents. Therefore, when designing dose-response experiments, this dose (5 nM), also known to be high in comparison to existing literature, was considered as the upper limit which was also further validated. Before single-cell level investigation, the influence of TGF- β dose on *ctgf* response profile in population level lumicycle experiments was studied. An experiment to investigate both the dose effect of TGF- β on the response and its putative depletion dynamics in long-term experiments was designed. To do that, samples were stimulated cells on culture dishes with three different doses of TGF- β covering two orders of magnitude of TGF- β dose (5-50-500 pM) which were again stimulated after 24 hour with the highest dose used (500 pM) (Figure 3.26).

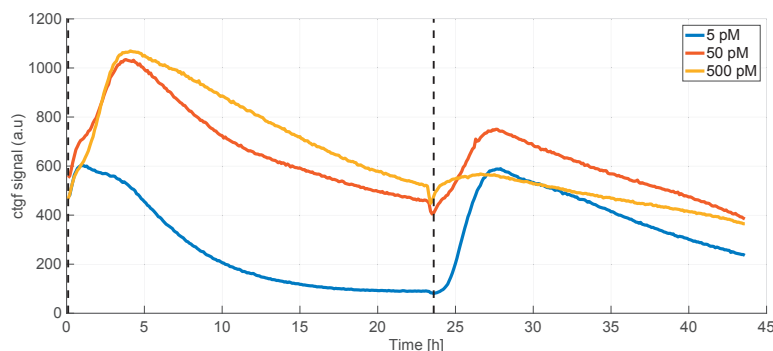


Figure 3.26 – Determination of the concentration dynamic range for TGF- β in population-level measurements. Samples were recorded for up to 2 days in lumicycle. Traces correspond to the average over 4 replicates. Stimulations are represented by dashed black lines (the first at $t=0$, the second at $t=24h$). The first stimulation doses varied, as indicated in legends, while the second stimulation was 500 pM.

The first responses scaled well with the TGF- β dose; low dose (5 pM) led to a shorter and weaker response while higher doses (50-500 pM) showed much higher and longer responses. High dose samples showed very similar response profiles, slightly differing in their return profile to basal levels, suggesting that these doses of stimulation already represent the saturating stimulation levels. This further validates our choice of 5 nM as the saturating high dose for single cell experiments. The second stimulation of the same dose (500 pM) also revealed other characteristics related to depletion of the ligand in medium such that samples with initial lower dose stimulations showed higher responses in the second stimulation. Yet, the highest dose sample (500 pM) in the first stimulation showed a very weak response to the second stimulation (Figure 3.26). First, it may point to still remaining amounts of TGF- β in the medium which had not been internalized or degraded. Second, cells that experienced high doses initially may have lost their responsiveness due to the internalization of ligand-receptor complexes which still keeps them impotent to further stimulations. Yet, initially low-dose applied samples responded to the second stimulus stronger inversely correlated with the dose of the first stimulation. This suggests that these samples recovered their potency to the ligand quicker due to the lower dose of the first stimulation which likely led to easier receptor turnover. In addition, this experiment roughly revealed the range of stimulation doses which we further could use for single-cell experiments. Similar results are obtained when double stimulation is applied on samples in the luminescence microscope (Figure 3.27).

3.3. TGF- β dose modulation on both SMAD signaling and *ctgf* response

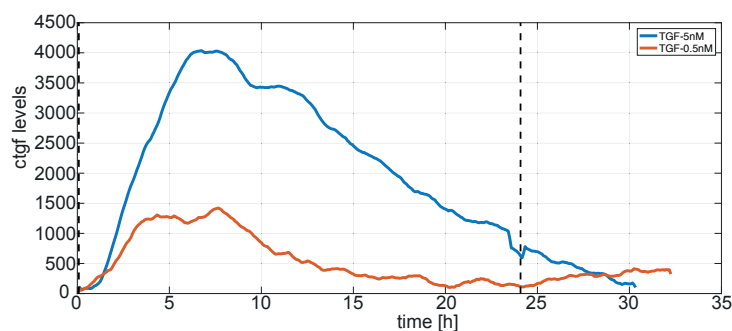


Figure 3.27 – Consecutive stimulation with high doses of TGF- β revealed refractoriness in the response. Samples were treated with high dose of TGF- β (5 nM) twice in 24 hours. Both samples showed refractoriness while the sample initially treated with lower dose (0.5 nM) exhibited a small response to the second stimulation. Stimulations are represented by black dashed lines.

These results overall revealed stimulation dose levels that substantially change both signaling and the target gene transcriptional response levels. Some studies emphasized that TGF- β molecules per cell is the main proxy for the TGF- β response [Feng et al., 2016]. Here, TGF- β molar concentration is used as the dose parameter since we used the same number of cells with the same amount of medium volume for each dose condition. When converted to this unit, our doses cover a three orders of magnitude range varying from 1.2×10^4 to 1.2×10^7 TGF- β molecules per cell. This range provided enough dynamic range covering doses from non-responding regime to the response saturation [Zi et al., 2011].

Next, different doses, previously tested on lumicycle measurements, of TGF- β (0.005-0.05-0.5-5 nM) were applied to samples to check their gene expression response together with signaling (SMAD2/4) profile in single-cells. These experiments revealed dose-dependent (analog) profile of the response characterized by gradually altered translocation signal and the target gene activity (Figure 3.28, Figure 3.29, Figure 3.30 and Figure 3.31). Expression responses exhibited slightly different profiles compared to lumicycle measurements such that 0.005 nM (5 pM) sample on average did not show any detected response both on SMAD and *ctgf* level which can be attributed to heterogeneity of single cells. Comparison of the data between untreated samples (Figure 3.8a) and those treated with the lowest concentration of TGF- β (5 pM) confirms that the concentration range used here was sufficient to capture the minimal responses from low doses of TGF- β . At this low stimulation level, cells either showed transient very weak response returning to basal levels quickly or no response both on SMAD and *ctgf* levels. At intermediately high stimulation levels (50 pM), average of single cells showed transient responses on signaling and *ctgf* response level. This intermediate level of stimulation represents an interesting non-saturated regime that cells respond in a dynamic range that can reveal information about SMAD mediated *ctgf* response levels. Starting from the ligand concentration of 500 pM, cells reached their maxima for signaling and target gene transcriptional response to TGF- β stimulation.

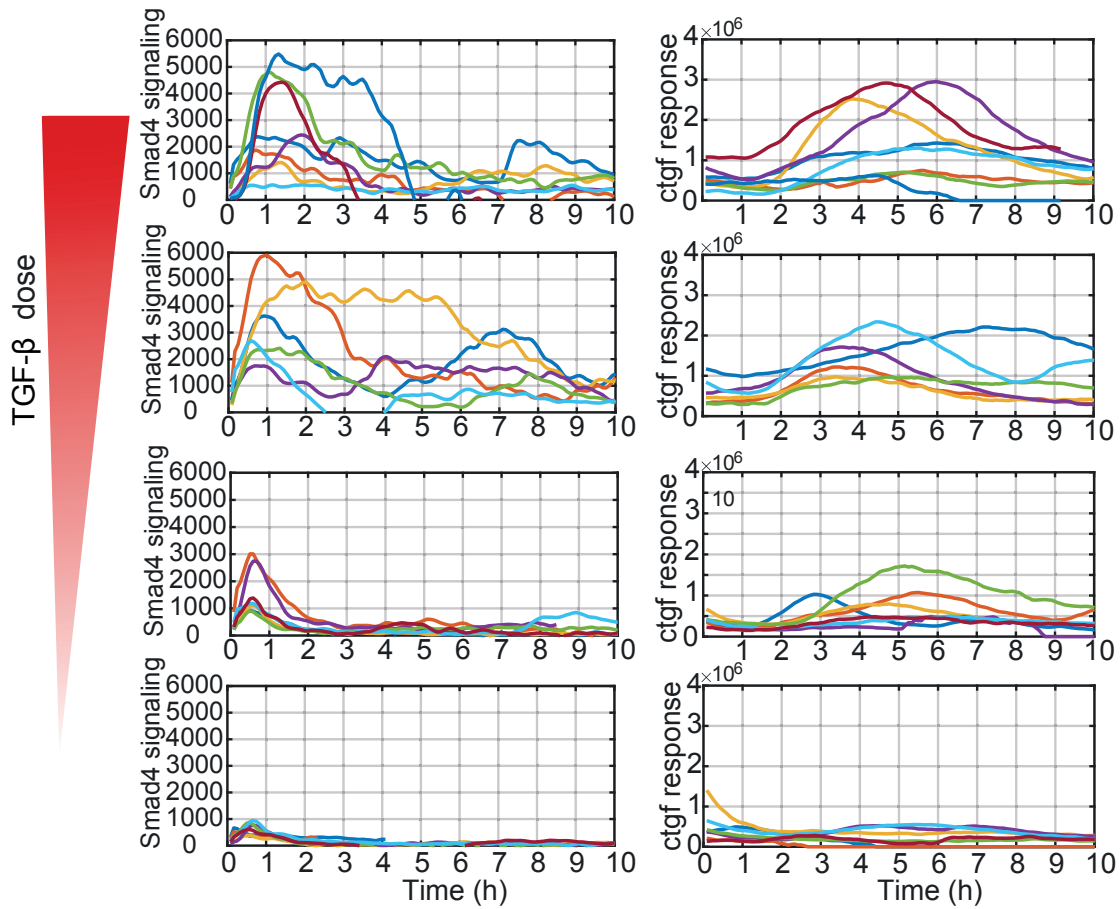


Figure 3.28 – Ligand dose response of TGF- β -induced SMAD4 signalling and the ctgf response in single cells. At time $t=0$, cells were treated with the indicated concentrations of TGF- β and representative traces for active single cells are shown. Left panels show single-cell SMAD4 nuclear to cytoplasmic signal difference quantification and right panels show ctgf responses in the corresponding single cells color-coded.

3.3. TGF- β dose modulation on both SMAD signaling and *ctgf* response

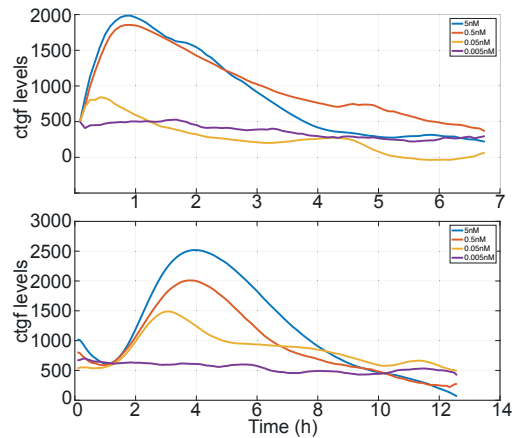


Figure 3.29 – Population averaged ligand dose–response of TGF- β -induced SMAD4 signalling and the *ctgf* expression. At time $t=0$, cells were treated with different doses of TGF- β represented in the plot legends. Upper panel shows the overlaid average of single-cell SMAD4 nuclear to cytoplasmic signal difference quantification. Plots were shown until 7hours to emphasize the region in which the transient response in SMAD4 occurs. Lower panel shows *ctgf* responses upon stimulation, plotted up to 12hours after stimulation.

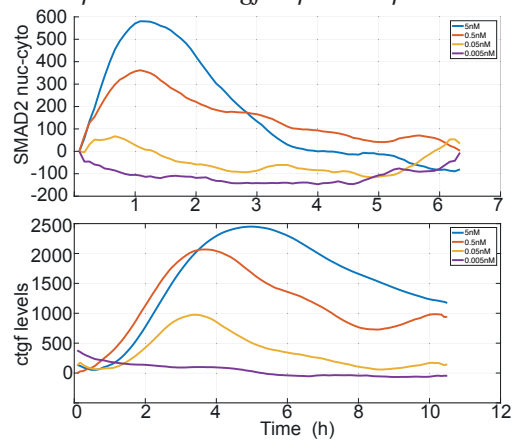


Figure 3.30 – Population averaged ligand dose–response of TGF- β -induced SMAD2 signalling and the *ctgf* expression. At time $t=0$, cells were treated with different doses of TGF- β represented in the plot legends. Upper panel shows the overlaid average of single-cell SMAD2 nuclear to cytoplasmic signal difference quantification. Plots were shown until 7 hours to emphasize the region in which the transient response in SMAD2 occurs. Lower panel shows *ctgf* responses upon stimulation, plotted up to 10 hours after stimulation

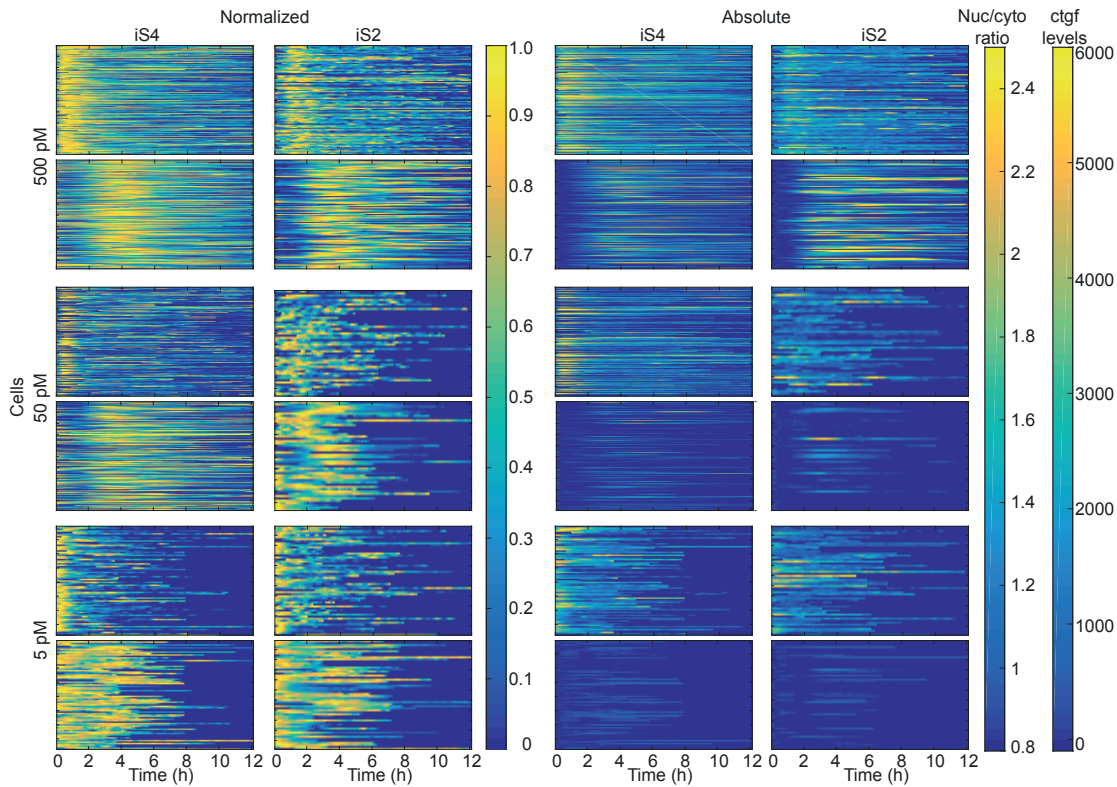


Figure 3.31 – Heat maps of stimulation dose response of TGF- β -induced SMAD4 signalling and ctgf response in samples treated with a range of ligand dose. Heat maps of stimulation dose response of TGF- β -induced SMAD signalling and ctgf response in single cells for *Nluc-SMAD4* and *Nluc-SMAD2* cell lines. At time $t=0$, cells were treated with different doses of TGF- β (500-50-5 pM) and response traces for all cells are shown both in normalized (left) and absolute (right) terms. Each experiment is represented by two heat maps representing SMAD signaling (upper) and ctgf response (lower). Absolute response values are plotted using the same scale: ctgf levels (0-6000), SMAD nuclear/cytoplasmic ratio (0.8-2.4).

SMAD2 and SMAD4 signaling dose response profiles differed slightly regarding their saturation concentrations and response alterations between different doses. SMAD signaling did not always scaled linearly with ctgf response levels while both signals revealed the analog encoding of TGF- β dose information (Figure 3.32).

Similarly, temporal profiles of responses displayed trends influenced by ligand concentrations (Figure 3.33). In single cells, SMAD4 response peak times showed significant increase in transition from 5 to 50 pM and reaching a saturation starting from 500 pM. SMAD2 signaling peak times, on the other hand, revealed a sharp transition of peak times between 50 and 500 pM. Reduced levels of ligand treatments resulted in smaller response transients of SMAD translocation compared to higher ligand doses. Interestingly, such alteration of SMAD translocation temporal peak did not reflect into ctgf dynamics since ctgf peak timing did not much vary in a dose dependent manner. This might suggest that single cells, once activated by sufficiently high SMAD signaling, respond with the same temporal dynamics tailored by other components of the pathway.

3.4. Ligand-dependent translocation and the target gene response

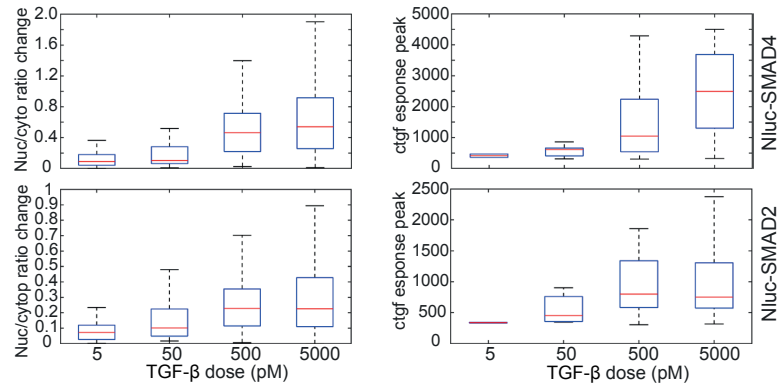


Figure 3.32 – Analog encoding of TGF- dose information by dose-dependent SMAD and ctgf responses. Data from cells (responsive ones) treated with different dose of ligand was analyzed to determine SMAD translocation response, quantified as nuclear/cytoplasmic ratio change from stimulation to the translocation peak and ctgf response, quantified as expression peak level change from stimulation to the ctgf peak.

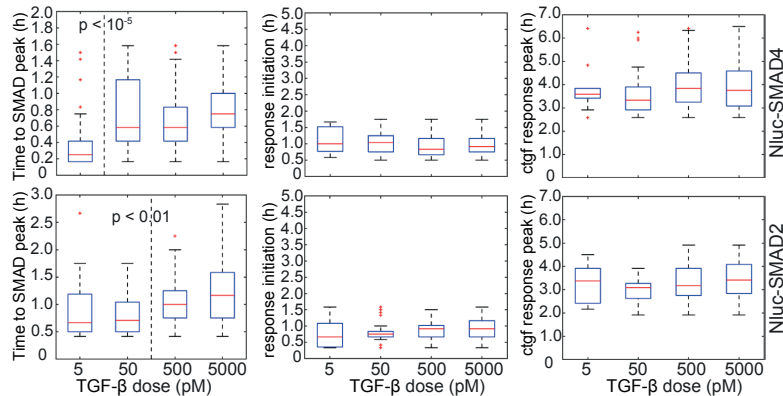


Figure 3.33 – Encoding of TGF- β concentration information in temporal profiles of SMAD signaling and ctgf responses. Data from the same cells treated with different dose of ligand were also temporally analyzed to determine the time taken to initiate ctgf response (middle), to reach the first peak of nuclear translocation of SMAD (left) and ctgf (right) (t-test for significance).

3.4 Ligand-dependent translocation and the target gene response

Cells receive a variety of signals from their environment and process these signals through shared signaling components. Response to both TGF- β and BMP signaling, for instance, involves parallel activation of two signal transduction pathways converging on SMAD4 to further regulate downstream events. In this section, how a different upstream activation mechanism, BMP4 stimulation, causes SMAD4 translocation and consequent ctgf response is investigated.

Similar to TGF- β signaling, BMP4 initiates the signal transduction cascade by binding to cell surface receptors and forming a heterotetrameric complex consisting of two type I and two type II serine/threonine kinase receptors. Upon formation of the complex, type II receptor transphosphorylates the type I receptor which leads to phosphorylation of receptor-regulated

Chapter 3. Results

SMADs (SMAD 1/5/8) and their association with SMAD4 to further activate downstream gene expression. Therefore, this pathway represents a similar cascade to TGF- β pathway causing SMAD4 to translocate and regulate gene expression. Therefore, it may provide us a useful platform to investigate SMAD4 mediated mechanisms of transcriptional regulation. To do that, I stimulated cells with BMP4 and measured both SMAD4 translocation and *ctgf* response which led to a similar response with a transient response in SMAD4 translocation and its consequent transient expression activation of *ctgf*. This stimulation exhibited a narrower transient on SMAD4 translocation compared to TGF- β stimulation, reaching a peak in half an hour despite the same translocation amplitude. Interestingly, this same amount of SMAD4 translocation resulted in much lower gene expression response which may suggest the activator type specificity of the gene expression response. In case of of BMP stimulation, the complex SMAD1/5/8-SMAD4 may activate the gene with different efficiency than the SMAD2/3-SMAD4 heteromers in response to TGF- β (Figure 3.34).

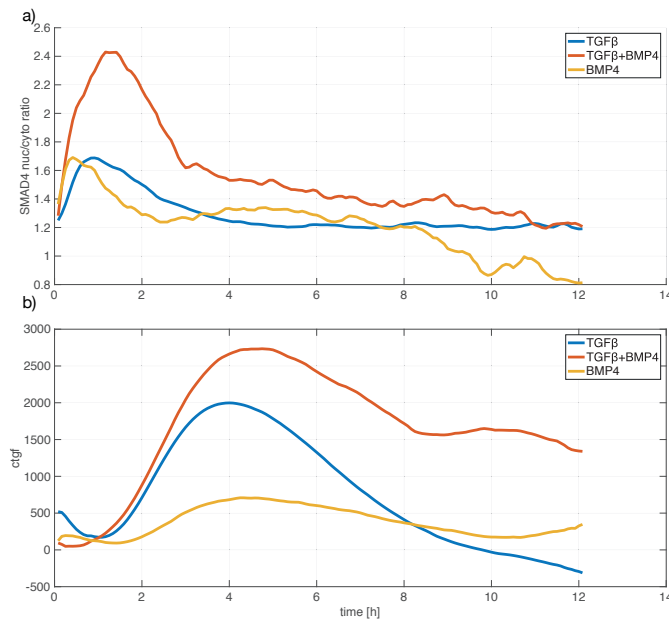


Figure 3.34 – Experimental observation of integrative processing of simultaneous TGF- β and BMP4 signals. At time $t=0$, cells were treated with either TGF- β (blue) or BMP4 (yellow) or both (red). Transient SMAD4 translocations characterized by nuclear to cytoplasmic ratio are shown up to 12 hours after stimulation (upper panel). *Ctgf* responses differed in amplitude and in temporal profiles between different treatments (lower panel)

To understand how cells process multiple signals in this context, a scenario involving co-stimulation with BMP4 and TGF- β was also examined. In this case, cells showed a prolonged and altered SMAD4 translocation as well as prolonged and stronger expression response. The net response caused by simultaneous two stimuli was the sum of the responses that would have been caused by each stimulus individually. These results provide experimental evidence of integrative processing of multiple simultaneous input signals in which cells display dynamics that represent the additive contribution of the two pathways. This may suggest that previously characterized dose responsive activation of the pathway enables cells

to respond in an integrating manner when subjected to different stimuli. To further elucidate the dependency of these two signalling branch, an experiment in which cells were first stimulated with a high-dose of TGF- β followed by a BMP4 stimulation was performed. Upon first TGF- β stimulation, cells normally move into a refractory phase in which cells do not respond until TGF- β is depleted or potent receptors are returned back to cell surface. Yet, the second stimulation with BMP4 led to both SMAD4 translocation and *ctgf* response which would not be possible otherwise with a second TGF- β stimulation (Figure 3.35). Overall, these observations suggest an integrative information processing of the pathway which sums up contributions originated from two independent stimuli of different strength to activate the same gene.

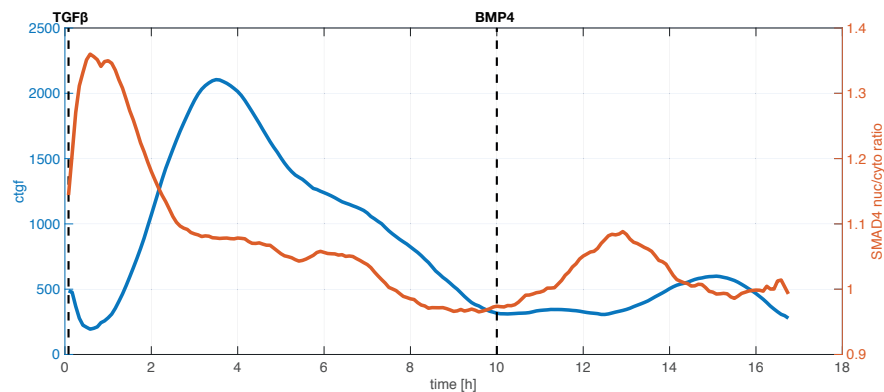


Figure 3.35 – Consecutive stimulation experiment with TGF- β (first) and BMP4 (second) stimulation. Response to stimuli is characterized by SMAD4 nuclear to cytoplasmic ratio (red) and *ctgf* expression level (blue). Stimulations are represented by dashed black arrows.

3.5 Single-cell correlations

Our dual-luminescence system technique allowed us to accurately measure both the levels and dynamics of transcription factor (SMAD2/4) reporter and its endogenous target gene (*ctgf*) with a temporal resolution of 5min. We applied this system to investigate SMAD2/4 translocations and its possible modulation on the target gene response in response to stimuli, mainly using TGF- β as the ligand. In principle, the modulation of transcriptional response could exhibit a broad array of behaviours depending on stimulation profile, promoter architecture and the topology of the pathway. Here, we attempted to dissect the contribution of translocation profile including features both in time and amplitude on the response features with single-cell resolution (Figure 3.36).

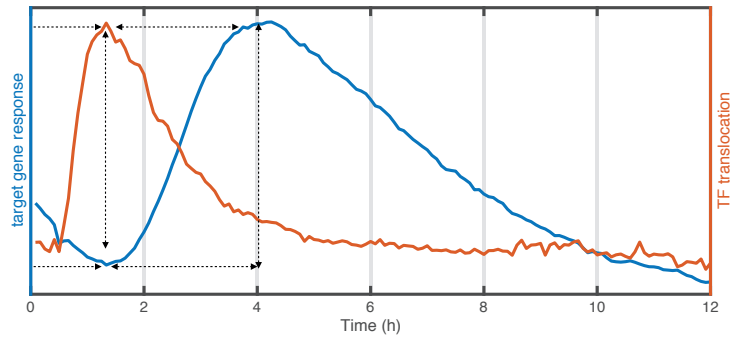


Figure 3.36 – Single-cell features of SMAD translocation and *ctgf* response dynamics in response to TGF- β stimulation. Both the amplitude and timing of response peaks were utilized as a measure of the signaling and the gene expression response. Horizontal black arrows represent temporal features while vertical black arrows denote amplitude features.

3.5.1 Temporal correlation of signaling activity and the expression response

One decent and straightforward hypothesis would be that SMAD translocation correlates with the expression temporally due to their mechanistic causality in the pathway. This would reflect into temporal correlations in single cells which can be characterized by peaking time points on both TF translocation and the target gene expression. Due to short-lived nature of the target gene reporter, temporal profile of the expression response also gives clues about the transcriptional response kinetics. Therefore, I attempted to check whether such correlation exist in individual cells (Figure 3.37).

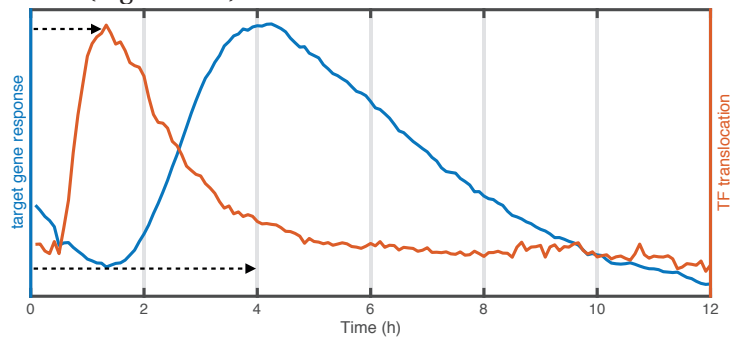


Figure 3.37 – Schematic representation for temporal correlation of signaling activity and the expression response.

To determine whether temporal profile of *ctgf* response, particularly the peaking time point, is influenced by SMAD4 translocation timing, I sorted cells either according to their SMAD4 peaking time point or according to the *ctgf* expression response peak time point. However, we did not observe any obvious correlation between SMAD4 translocation peaking and *ctgf* expression response (Figure 3.38 and Figure 3.39).

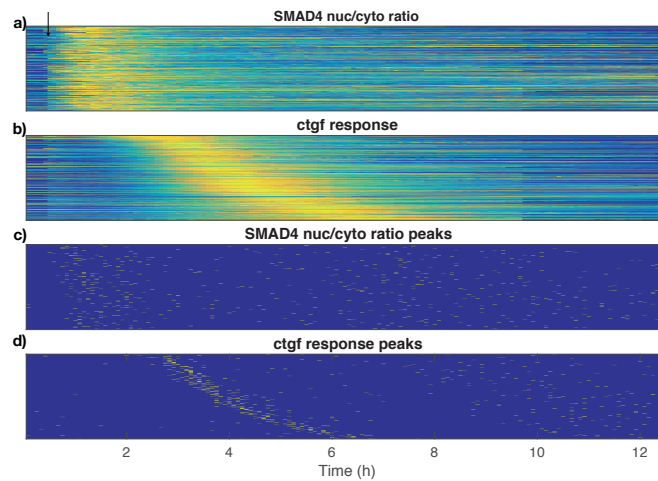


Figure 3.38 – Temporal correlation of signaling activity and the expression response based on *ctgf* peaks. a) Heat map of SMAD4 translocation in individual cells over time in response to 5nM TGF- β 1. Each horizontal line represents a single-cell signal obtained with min-max scaling. b) Heat map of *ctgf* expression response for each single-cell corresponding to SMAD4 heat map is shown again with min-max scaling. c) Distribution of SMAD4 nuc/cyto ratio peaks obtained with automated peak detection algorithm. d) Distribution of *ctgf* expression response peaks obtained with automated peak detection algorithm. Black arrow represents the stimulation timing.

Heat maps for both SMAD4 and *ctgf* expression response were normalized using min-max scaling to visualize better the temporal profile of peaking in both signals. To ensure a robust detection of peak points, both automated and manual annotations were performed which reveal a high agreement. Heat maps were sorted according to manual annotations of peaks while binary peak maps showed the result of automatic peak detection algorithm.

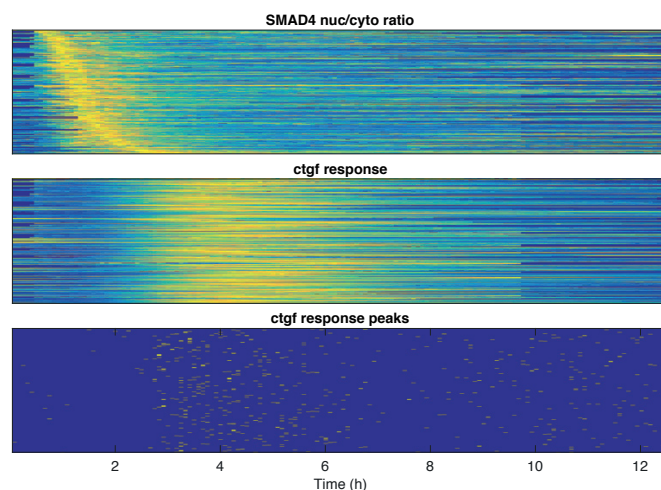


Figure 3.39 – Temporal correlation of signaling activity and the expression response based on SMAD4 translocation peaks. a) Heat map of SMAD4 translocation in individual cells over time in response to 5nM TGF- β 1 ordered by their peak timing. Each horizontal line represents a single-cell signal obtained with min-max scaling. b) Heat map of *ctgf* expression response for each single-cell corresponding to SMAD4 heat map is shown again with min-max scaling. c) Distribution of *ctgf* expression response peaks obtained with automated peak detection algorithm.

To further validate this observation, we applied the same kind of analysis to SMAD2 experiments which also did not reveal any obvious correlation between translocation and expression response timings. Such lack of correlation between SMAD translocation peaking and the response peaking may suggest a mechanism that SMADs trigger the transcriptional response upon its translocation while other components in the pathway shape the response further determining the duration of the expression response.

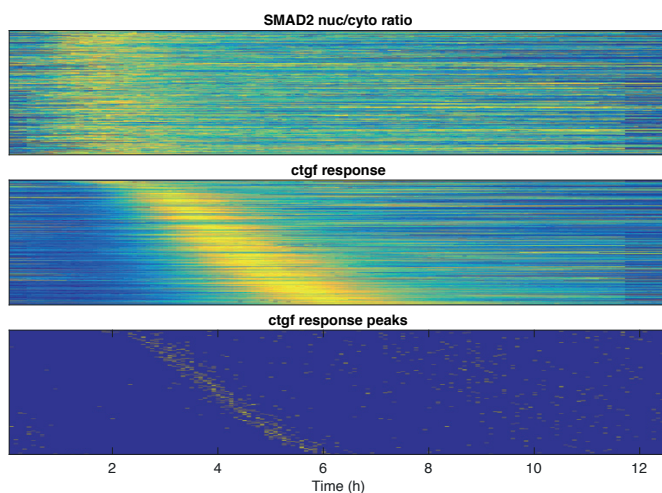


Figure 3.40 – Temporal correlation of signaling activity and the expression response based on SMAD2 translocation peaks. a) Heat map of SMAD2 translocation in individual cells over time in response to 5 nM TGF- β 1 stimulation. Each horizontal line represents a single-cell signal obtained with min-max scaling. b) Heat map of *ctgf* expression response for each single-cell corresponding to SMAD2 heatmap is shown again with min-max scaling, ordered based on peak timing for each cell. c) Distribution of *ctgf* expression response peaks obtained with automated peak detection algorithm.

3.5.2 The expression response amplitude scales with the expression response duration

Despite the lack of temporal correlations between SMAD and the expression response, one feature emerged during our investigations regarding the *ctgf* expression response itself. When sorted according to their response maximum time points, the *ctgf* expression response amplitude compared to their basal levels exhibited an increasing profile with the increasing response peaking time. Cells that reach their response peaks earlier exhibited weaker response amplitudes while late peaking individuals showed stronger responses (Figure 3.41). Since negative feedbacks have the strongest potential to cease the activation and parameters such as TGF- β concentration, SMAD levels were not varied in this set of examples, the variability of activation duration in individual cells can be attributed to the variability in negative feedback components.

To further validate this observation, cell traces were grouped into three equally sized bins based on their peak timing, and labelled as early, intermediate and late. Such grouping also supported the previous observation that early peaking individuals showed weaker and

narrower responses while late peaking ones exhibited stronger and wider responses (Figure 3.41).

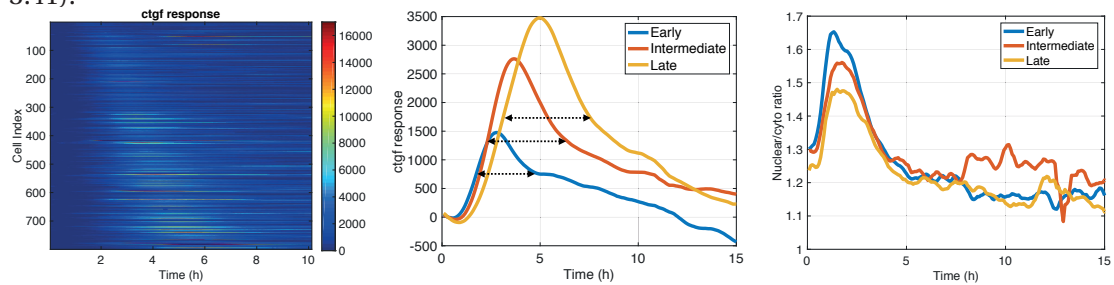


Figure 3.41 – Heat map of CtGF expression responses in single cells to TGF- β stimulation (at $t=0$), sorted according to their response peak timing, shows an increasing trend also in the response amplitude (Left). Color bar represents the signal amplitude. Grouping of CtGF expression responses into three categories based on their response peaking time (Middle). Groups are labeled as early, intermediate and late. Dashed arrows represent full width at half maximum (FWHM) as a measure of response duration. Groups did not show significant difference in SMAD translocation profiles (Right).

The previously defined three categories of cells did not show an obvious trend in their SMAD translocation profiles (Figure 3.41). This observation similarly suggests an initiation role of SMADs on the gene expression in response to stimulation while negative feedbacks [Warmflash et al., 2012, Zi et al., 2011, Wegner et al., 2012] or secondary mechanisms of SMADs recruiting activators and suppressors shape the transcriptional response further at later instances [Coda et al., 2017].

3.6 Improvements of the experimental pipeline

Dual-reporter system integrated with microscopy system and image analysis pipeline has still room for further improvements. One important modification that would alter the precision of measurements is to account for substrate instability leads to a slight drop in Nluc signal during long-term experiments. Since Nanoluciferase is a newly engineered protein being in its infancy, its substrate stability performance gets better with new commercially available variants. To evaluate its current performance for long-term imaging, a cell line expressing a fusion protein of Nanoluciferase and firefly luciferase was developed. This calibration cell line, owing to 1:1 stoichiometry, allowed us to measure the slow luminescence decay of nanoluciferase by cross-correlating signal between two channels (Figure 3.42). While the substrate of firefly luciferase is stable for several days in our experimental conditions, it can be utilized to correct for slow decaying of Nanoluciferase signal. In future experiments, incorporation of this Nluc-Fluc cell line into experiments with Nluc-SMAD2/4 cell line would be a useful modification for the robustness of measurements.

Chapter 3. Results

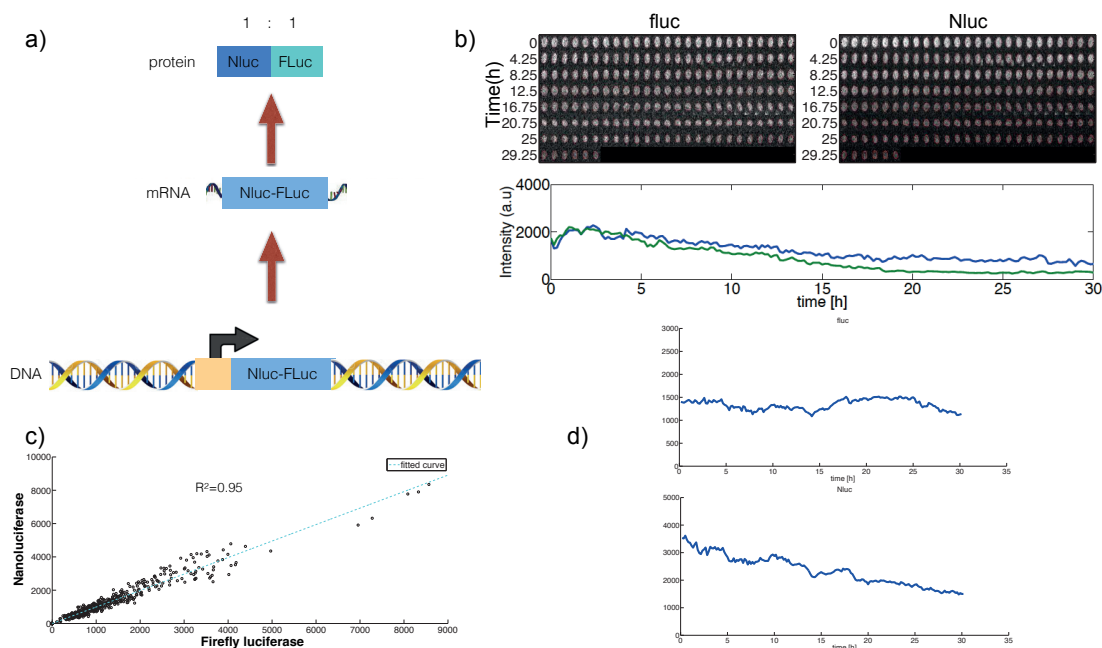


Figure 3.42 – Nluc-Fluc cell line for signal correction. a) Schematic representing expression of Nluc-Fluc fusion protein under doxycycline inducible promoter with 1:1 stoichiometry. b) A sample single cell trace showing coupled expression of Fluc and Nluc for extended periods of time. c) The linear correlation over a wide range between signals quantified in both channels. d) Population mean of both channels showing the slow decay in Nluc signal compared to stable Fluc signal.

In the current system, the quantification of SMAD translocation is characterized by a method quantifying nuclear-to-cytoplasmic ratio of the Nluc signal in which a ring-shaped annulus around the nucleus defined the cytoplasmic region. This aspect of the system further can be improved by including cytoplasmic markers that can allow us to clearly segment the cytoplasmic compartment. Another modification of the experimental system can be the involvement of microfluidic cell culture systems to apply complex time courses of stimulation. To this end, it was once attempted to incorporate a 2x2 microfluidic chamber system, altogether fitting into the field of view, which would, in principle, allow us to perform 4 experiments in parallel subjected to different stimuli (Figure 3.43). Such parallel experiments performed in the same conditions might reduce sources of noise due to batch effect and provide the opportunity to apply complex forms of ligand input to samples.

3.6. Improvements of the experimental pipeline

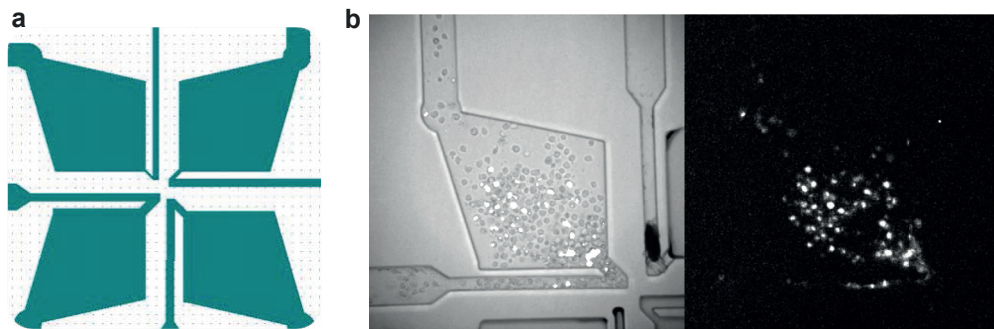


Figure 3.43 – Initial implementations of microfluidic chambers. *a)* The design of 4 microfluidic chambers fitting into the field of view. *b)* One successful run of the system with cells loaded into chambers (left) and luminescence imaging of cells in the chamber (right).

4 Discussion and Outlook

In this chapter, the main points of the study are summarized and the biological significance of results are elaborated.

4.1 Dual-luminescence imaging system allows highly sensitive monitoring of protein levels over extended time periods

Despite their widespread usage in monitoring dynamic protein levels in single cells, there are certain aspects that limit the use of fluorescent reporters. First, the sensitivity of fluorescence imaging is limited by autofluorescence which hinders measurements of protein abundances lower than a few thousands [Filipczyk et al., 2015]. Second, phototoxicity and photobleaching put a constraint on the use of fluorescent reporter systems in long-term imaging measurements. Finally, certain fast protein expression dynamics can not be captured with fluorescence due to slow maturation processes and variable folding efficiency [Miyawaki et al., 2003, Shaner et al., 2005]. Even the fast folding variants such as sfGFP (superfolder green fluorescent protein) require on average 6 mins to become fluorescent which may still be slow to capture some fast dynamics [Khmelniskii et al., 2012]. Therefore, luminescent reporter counterparts provide a unique alternative to fluorescent reporters. Owing to its higher sensitivity which can go down to hundreds of protein molecules, luminescent reporters have been used in a broad array of studies for dynamical transcriptional and gene expression measurements [Mazo-Vargas et al., 2014]. In addition to higher sensitivity, duration of imaging is not limited due to any cellular perturbation from the luminescent imaging system.

As a novelty in this study, we used NanoLuc (Nluc), which is a newly engineered small luciferase, in combination with firefly luciferase. NanoLuc (Nluc) has been shown to provide several advantages over other luminescence systems. Its smaller size, enhanced stability and significantly brighter profile compared to the traditional luminescence counterparts make it a great choice as a fusion tag to monitor proteins in long-term imaging experiments [Hall et al., 2012]. Therefore, we utilized Nluc reporters in combination with previously developed firefly luciferase reporters in the same cell line owing to their differing emission spectra [Mandic

Chapter 4. Discussion and Outlook

et al., 2017]. In two of those cell lines, Nluc is fused to two transcriptions factors, SMAD2 and SMAD4, which are main downstream activators of many target genes responding to TGF- β family ligands (Section 3.1.1). In addition, cells expressing Nluc-Fluc fusion proteins were also utilized for evaluation of furimazine substrate stability (Section 3.6). Expression of these three fusion proteins was driven by a doxycycline inducible promoter, TRE3G, whose expression spans a broad range depending on doxycycline levels. In our experiments with SMADs, we exploited this tunability of the promoter and varied expression levels by more than two orders of magnitude. In this wide range of expression, Nluc was able to operate linearly reporting protein levels varying from subendogenous to overexpressed levels. Owing to its bright luminescence, we were able to monitor low levels of SMAD proteins that were not detectable by Western blotting.

Despite its aforementioned advantages, luminescence imaging has several limitations that need to be addressed. It requires certain experimental parameters to be optimized before experiments depending on the expression level of the protein being monitored. For example, the time resolution of the imaging needs to be adjusted to allow for a temporal window to collect enough luminescence signal. Therefore, long exposure times are needed for lowly expressed proteins which makes it difficult to image multiple positions without sacrificing temporal resolution. Long exposure times above several minutes lead to poorer spatial resolution due to cellular/subcellular motions in comparison to fluorescence imaging which require much shorter exposure times for a single frame. This low spatial resolution due to minutes scale exposure time may have caused small deviations in our SMAD localization measurements compared to reported values in the literature obtained using fluorescence imaging [Warmflash et al., 2012, Strasen et al., 2018]. However, we believe these potential deviations did not affect our conclusions significantly since we typically seeded cells with high confluency which should limit the excess movement of cells. Moreover, the use of different cell lines and expression levels of our reporter may have also contributed to such deviation.

Similar to exposure times, substrate concentrations need to be optimized depending on the protein expression levels. Therefore, marked changes of SMAD levels due to different doses of doxycycline required adjustments of substrate levels used. Another issue that concern our dual imaging system was the stability of Nanoluc substrate used. Although the Fluc substrate, luciferin, exhibits very high stability up to several days, the most common substrate of Nluc, furimazine, suffers from instability in cell culture medium (half life around 70min). Commercially available alternatives of furimazine, however, display much higher stability since these pro-substrates are converted to their active form only in presence of cells. The Nluc substrate utilized in this study (MT Cell Viability Assay Substrate, Promega) relies on the reduction potential of cells to be activated. Therefore, upon addition of this pro-substrate, the activated form concentration shows a slow increase over time reaching a plateau phase followed by a slow decay phase. To characterize and account for these temporal changes in activated substrate concentration, the possibility to utilize the cell line expressing Nluc-Fluc fusion protein allowing us to cross-correlate the Nluc signal to the stable Fluc signal was demonstrated (Section 3.6). The cell line expressing Nluc-Fluc fusion protein revealed this

4.2. Simultaneous monitoring of SMAD mediated signaling and target gene activation in response to TGF- β revealed high variability in single cells

slow decay in Nluc prosubstrate luminescent activity. These slow changes appeared to make a difference only in time scales longer than 24 hours while SMAD translocation dynamics upon stimulation occur at much shorter time scales, typically 2-3 hours. At these short time scales, the dynamics of substrate stability are negligible.

4.2 Simultaneous monitoring of SMAD mediated signaling and target gene activation in response to TGF- β revealed high variability in single cells

Despite great efforts to understand pathways and their implications in different cellular contexts, complex interactions between upstream regulators and their downstream target is still not completely understood. Despite a detailed picture of each step, it still remains unanswered how information is transferred between each level. This is partly due to the lack of methods which can causally relate a single pathway to its target gene responses. Therefore, methods that can allow to measure multiple steps in signaling pathways is needed to disentangle regulation in different stages of the central dogma. Some recent studies monitoring response pathways at different nodes have contributed to our understanding of signaling, transcription and translation [Aymoz et al., 2016, Wilson et al., 2017, Xu et al., 2015, Schulz et al., 2018].

Deciphering the information flow between consecutive steps is challenging due to complex networks of upstream signaling and downstream targets. Receptor level induction usually activates multiple signaling pathways in parallel with different dynamics converging on target gene responses. These pathways are typically regulated at different steps which makes it difficult to causally relate any signaling pathway to their target gene responses. To make the investigation easier, one can isolate a single pathway to prevent the impact of other regulators. Similarly, exerting a fine-tuned control on the activation levels of the pathway may help to causally relate input-output relations. Lastly but most importantly, developing an experimental system to measure different nodes during the response process is crucial. In this study, we embarked on this task to dissect regulation of the target gene (ctgf) response in relation to their downstream activators, SMADs, in response to TGF- β treatment.

Previous studies characterized the TGF- β pathway by dissecting regulation at different steps of the cascade. Some studies focused on the pathway at the receptor level revealing the importance of TGF- β availability and depletion as an important mechanism [Vizán et al., 2013]. Other experimental and theoretical studies investigated the mechanisms controlling the temporal dynamics of SMAD signaling [Warmflash et al., 2012, Strasen et al., 2018]. It has been further shown that negative feedback mechanisms acting on the pathway at multiple levels determine the dynamics of signal termination [Itoh and ten Dijke, 2007, Wegner et al., 2012]. Despite investigations at multiple steps of TGF- β pathway, a quantitative relationship between upstream activators (SMADs) and the downstream target gene transcriptional activity is still not clear mostly due to gene specific nature of such relation and unavailability of experimental systems to investigate this regulation. To this end, we established two Tet-On inducible stable

cell lines each expressing a fusion protein of a luminescence (Nanoluc) reporter to either SMAD4 or SMAD2, the two main transcription factors downstream of the TGF- β signaling pathway (Section 3.1.1). These cell lines also contain a short-lived firefly luciferase reporter for the expression of the target endogenous *ctgf* gene, which allowed us to quantitatively link nuclear accumulation of TFs upon TGF- β stimulation to the target gene expression in real-time single cell measurements. Our experimental approach allowed us to monitor the translocation of SMADs (SMAD2 and SMAD4) from cytoplasm to nucleus together with *ctgf* response in TGF- β stimulation experiments at the single cell level. We characterized transient SMAD translocations into the nucleus concurrently with *ctgf* response initiation in hundreds of single cells.

Single-cell data obtained from stimulation experiments revealed agreement with the previously reports while differing in some respects. Average SMAD2 and SMAD4 translocation dynamics upon stimulation agreed with the biochemical measurements in previous studies [Vizán et al., 2013] regarding SMAD accumulation in the nucleus peaking at 1h due to formation of activated R-SMAD-SMAD4 complexes. This increase in nuclear localization is attributed to the increased interaction with DNA and reduced export levels [Hill, 2009]. Next, R-SMAD dephosphorylation results in dissociation of SMAD4 from complexes and export from the nucleus to the cytoplasm which reflected into reduced nuclear localization signals in our measurements in both SMAD2 and SMAD4 case. Yet, in some studies, probably due to used cell types, SMAD2 showed sustained nuclear localization which was not the case in our measurements [Warmflash et al., 2012].

Regarding basal localization, both SMAD4 and SMAD2 localization before induction was shown to be homogeneous in cells which may slightly deviate from the previous reports claiming homogeneous SMAD4 and predominantly cytoplasmic SMAD2 signals in the absence of stimulation [Strasen et al., 2018, Warmflash et al., 2012]. Such deviation, however, could be explained with differences between studies including expression levels of SMAD reporters, cell types used, single-cell quantification strategy and the type of microscopy.

SMAD translocation and *ctgf* response showed substantial heterogeneity among individual cells while the gene response showed much higher variability both temporally and in amplitude, agreeing with previous observations [Warmflash et al., 2012, Strasen et al., 2018]. When response peak time points of both SMAD and *ctgf* were measured, translocation of SMADs into the nucleus showed a much narrower distribution compared to the widespread temporal distribution of the gene expression (from 2h to 6h). Increased variability both temporally and in amplitude of the gene response compared to the signaling may suggest the propagation and accumulation of noise throughout the pathway (Figure 4.1).

4.2. Simultaneous monitoring of SMAD mediated signaling and target gene activation in response to TGF- β revealed high variability in single cells

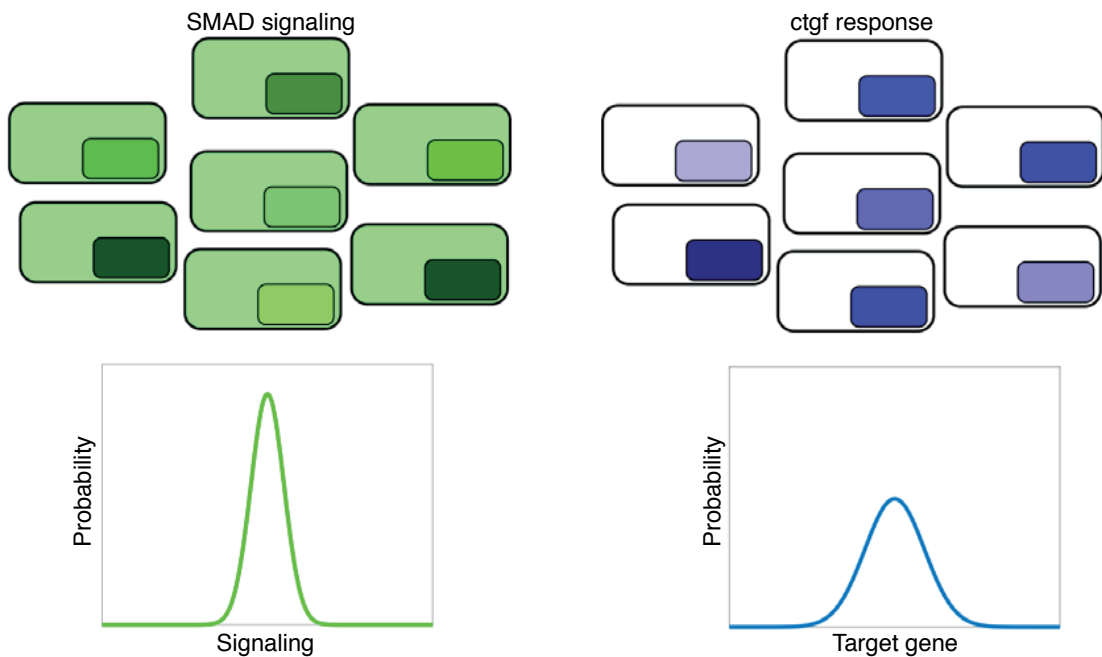


Figure 4.1 – Target gene expression response, both temporally and in amplitude, showed a higher cell-to-cell variability compared to SMAD signaling. Color shades represent the response intensity.

When assessing the heterogeneity both in basal and induced condition of the pathway, it is crucial to note that both SMAD signaling and ctgf itself have been recently reported as a target of certain proteins, YAP/TAZ, which themselves are regulated by Hippo signaling pathway. Ctgf, for instance, is identified as a direct target gene regulated by activated YAP-TEAD complexes [Zhao et al., 2008]. In the nucleus, TAZ/YAP act as transcriptional cofactors regulating the activity of several transcription factors including TGF- β -regulated SMADs [Varelas et al., 2008]. The Hippo pathway senses cell density information to control tissue growth by regulating the localization of TAZ/YAP which in return dictate the localization of active SMAD complexes in response to cell density information [Varelas et al., 2010]. In high-density cell culture, the Hippo pathway drives the cytoplasmic localization of YAP/TAZ resulting in sequestration of phosphorylated SMADs too in the cytoplasm and overall suppression of the TGF- β signaling. While these observations might partially explain the cellular variability of ctgf expression observed in the basal state, the contribution of these mechanism was tried to be kept minimal using high and the same confluency of cell cultures throughout experiments.

Our dual reporter system, in principle, might allow us further to investigate the contribution of different factors that influence pathway at different stages which eventually lead to modulation of the response. Owing to the destabilized nature of the luciferase reporter, the system has the potential to infer underlying mechanisms and deduce dynamics of the transcription indirectly [Suter et al., 2011, Molina et al., 2013]. Similarly, the tunability of SMAD reporter provided us the plasticity to operate in different regimes of SMAD expression.

4.2.1 Higher SMAD4, but not SMAD2, levels favor sustained and oscillatory *ctgf* responses

In principle, TGF- β induced SMAD signaling has been shown to be predominantly transient while consecutive pulse stimulations may result in sustained activation [Warmflash et al., 2012, Sorre et al., 2014]. In our experiments, we observed both transient and sustained *ctgf* expression responses co-occurring in individual cells, especially to higher doses of TGF- β stimulation. Cells that respond in a transient manner returned back to their basal levels approximately in 8 hours while the sustained population remained higher over an extended period of time. Despite the small fraction of cells showing sustained profile, we attempted to dissect these qualitatively different responses since we thought that biological implications of such variability could be important. Of the several other biological processes involved, higher *ctgf* expression induced by TGF- β has been shown to be highly associated with pathological fibrosis and fibroblasts are direct targets of such high expression levels [Holmes et al., 2001].

To entangle such variability of responses in hundreds of cells, we utilized k-means clustering which quantified differences and similarities in dynamic patterns of single cell traces. Since our main interest in this classification was to define classes with different dynamic profiles rather than the expression amplitude itself, we utilized correlation as the dissimilarity metric. We also made a heuristic choice of two clusters since this can summarize what we observe in time courses. However, we do not exclude the possibility to involve more number of classes which can yield to more fine-tuned description of the whole data. Our clustering approach still exhibited good performance evaluated with silhouette scores. In principle, such variability of transient and sustained responses would have been attributed to negative feedback mechanisms based on previous studies. However, exploiting the inducible nature of our TRE3G promoters expressing SMAD2/4, we attempted to assess whether SMAD levels have any impact on this clustered classes. Varying SMAD levels has revealed an interesting trend such that the portion of sustained subpopulation increased with altered levels of SMAD4 while SMAD2 levels did not have any significant effect. This difference can be attributed to the fact that when in heteromeric complex of SMAD2/SMAD4, only SMAD4 can bind DNA through its SMAD binding elements (SBE) and among R-SMADs, SMAD3 plays a more significant role for the induction of TGF- β target genes including *ctgf*.

When expressed subendogenous levels, SMAD4 did not show any alteration on the gene expression. Starting from ectopic expression that are comparable to endogenous levels, samples with high SMAD4 expression showed much bigger portion of sustained responses. In the extreme cases in which SMAD4 expression reach several fold higher than the endogenous levels, a large portion of cells, reaching up to half of the population, exhibited sustained and even oscillatory *ctgf* responses. This observation interestingly related sustained responses to SMAD4 levels which had been generally attributed to negative feedbacks in the pathway. Population level measurements also supported this SMAD-type-specific promotion of sustained responses. However, only the single-cell approach coupled with measurements both on transcription factor and the target gene expression levels allowed us to determine the extent

4.2. Simultaneous monitoring of SMAD mediated signaling and target gene activation in response to TGF- β revealed high variability in single cells

of which SMAD4 levels can promote such sustained responses. This observation also defined an expression level below which sustained responses have the same incidence rate which may suggest a thresholding mechanism that filters out small fluctuations in SMAD4 levels and prevent undesired cellular fate decisions such as fibrosis in case of *ctgf*. Despite its influence on the response transientness, SMAD4 levels did not exhibit any significant effect on the first response amplitude in the target gene in high-dose TGF- β stimulation experiments. However, such lack of correlation can be due to saturation in the response amplitude. Therefore, the effect of SMAD abundance on the target gene response amplitude in low doses of stimulation needs further investigation.

4.2.2 Analog encoding of TGF- β dose information in both signaling and transcriptional response

While current explanations about response variability of members of the TGF- β superfamily of ligands focus mainly on the downstream signal transduction cascades, the concentration of TGF- β in the surrounding environment and its depletion emerge as the main determinants of the signaling response [Vizán et al., 2013]. Recent studies revealed the dose dependent responses of SMAD signaling [Strasen et al., 2018]. However, the question about how cells encode and transmit this ligand concentration dependent signaling information into downstream genes is still unanswered (Figure 4.2). Therefore, using our dual-reporter system, how different doses of TGF- β influence SMAD signalling and its consequent gene expression response were analyzed (Section 3.3).

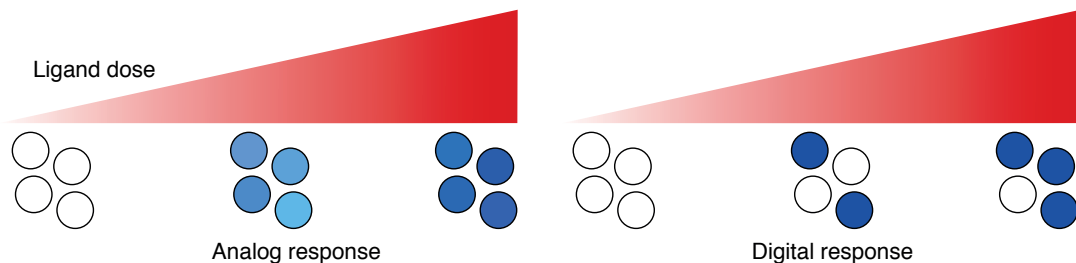


Figure 4.2 – Schematics of the two different scenarios for single-cell responses to stimuli. White filled circles represent the unresponsive single-cells while blue shades display the response strength.

In principle, the increase in the ligand concentration alters the probability for a receptor to bind the ligand which would lead to continuous increase in the response. Yet, due to ligand induced degradation of receptors, TGF- β responses were shown to be eventually attenuated [Vilar et al., 2006]. In our ligand dose-response experiments, using a wide range of stimulation doses, a span of three orders of magnitude (5pM–5nM), different regimes of activation were characterized in signaling and the transcriptional response (Section 3.3). Such a range of ligand concentrations was utilized covering the full range of activation as previously reported and further validated with our experiments. Of note, molar concentration is used throughout the investigation to express ligand dose levels, despite some studies using density, to comply

Chapter 4. Discussion and Outlook

with previous studies regarding ligand doses [Molina et al., 2013]. Yet, it is important to state that our ligand dose regime (5 pM to 5 nM) corresponds to ligand density range of 0.125-125 ng/ml. The upper bound of the range was intentionally chosen high to ensure that the pathway was activated to the full extent. Moreover, some studies claimed TGF- β molecules per cell as the main parameter to assess the activation levels [Zi et al., 2011]. When converted to this unit, our doses cover a range spanning three orders of magnitude, varying from 1.2×10^4 to 1.2×10^7 TGF- β molecules per cell.

5 pM (0.125 ng/ml) was determined, in measurements of this study, as the lower threshold for TGF- β stimulation which induced virtually no or very slight response in both SMAD signaling and the target gene. Previous studies have reported that SMAD4 signaling amplitude but not the kinetics is dependent on the dose based on the observation meaning that low ligand treatments (0.1 ng/ml) just altered the amplitude of nuclear SMAD4 peak but not the duration of nuclear localization [Warmflash et al., 2012]. Yet, in our observations, sufficiently low doses of ligands reduced the SMAD4 nuclear localization time window as well as the translocation amplitude. Such low dose stimulations resulted in very short transients in signaling and relatively weaker *ctgf* responses. This suggests that the time that SMAD complexes stayed nuclear impacts the transcriptional response to TGF- β in a dose-dependent manner.

Above such low doses, a sensitive regime of stimulation dose (around 50 pM) revealed small signaling responses leading to pronounced target gene response. Overall, the activation was identified as an analog process at the single-cell level. Consistent with an analog response, nearly all of the cells displayed a response to all of the doses of TGF- β if not insufficiently low (5 pM - 0.125ng/ml). The response amplitude was progressively altered until both SMAD signaling and the target gene response reached a saturation regime in which further higher doses of stimulation did alter neither SMAD translocation nor the target expression levels. Previous studies also reported the dose response profiles of the pathway, but only focusing on signaling dynamics [Sorre et al., 2014]. Here, it was shown that SMAD signaling scale similarly, but not exactly in the same manner, with *ctgf* response levels in a dose-dependent manner.

In our observations, SMAD2 and SMAD4 signaling dose response profiles differed only slightly regarding their saturation concentrations and response alterations between different doses which may indicate an existing upstream mechanism. Indeed, the availability of TGF- β in the medium for instance, governed by cellular intake and consequent depletion, was shown to play a key role in shaping the signaling and the transcriptional response, particularly in long-term experiments [Vilar et al., 2006, Vizán et al., 2013]. Supporting that notion, initial low doses of ligand treatments in our observations allowed restimulation of cells after a certain time while high doses kept cells in a refractory state for extended periods of time. TGF- β dose was identified as one of main determinants of cellular response to stimulation which needs to be taken into account assessing the cellular response dynamics both in short and long timescales. Previous analysis of response dynamics using clustering approach, for instance, revealed the existence of sustained subpopulation whose activity extends beyond initial transients (Section 2.4). High doses of ligands used in that set of experiments might

4.2. Simultaneous monitoring of SMAD mediated signaling and target gene activation in response to TGF- β revealed high variability in single cells

have allowed certain individuals to respond in a sustained manner. Therefore, it is crucial to consider ligand profiles together with other important components of the pathways, SMADs in our case, when assessing the variability of the pathway responses.

4.2.3 Integrative processing of independent inputs in signaling and transcriptional response

Cells receive a broad array of signaling molecules from their environment which makes it crucial to process concurrent signals to direct downstream cellular processes. Although studies investigate signaling pathways in isolation, it is still elusive how cells process concurrent signals. In a scenario of multiple simultaneous stimulation, some pathways exhibited a non-integrative processing in which cells respond to only one of the two inputs depending on variability of certain cellular components. In the nuclear factor- $\kappa\beta$ (NF- $\kappa\beta$) system, for instance, tumor necrosis factor (TNF) and lipopolysaccharide (LPS) induce distinct responses due to different dynamics of I- $\kappa\beta$ kinase and NF- $\kappa\beta$. Yet, in case of multiple simultaneous stimulation, the pathway exhibited a non-integrative processing in which cells respond to only one of the two inputs depending on variability of certain cellular components [Kellogg et al., 2017].

This study predominantly characterized the ligand TGF- β response of cells focusing on SMAD signaling and *ctgf* expression. Yet, at the molecular level *in vivo* conditions, there is a complex signal transduction network combining 42 identified different ligands of the TGF- β superfamily depending on the cellular context [Massagué, 2012]. Both ligands and receptors form an intricate interaction network in which different ligands share receptors and downstream signal transducers, probably coupling their signaling.

To investigate how cells handle simultaneous stimulation and how these signals are integrated to transcriptional responses, BMP4 was used as the second ligand in a small set of experiments. TGF- β and BMP pathways have similar topology that also interact with each other through SMAD4 signal transduction. Like TGF- β , BMP stimulation initiates a SMAD signaling cascade in which receptor regulated SMADs, SMAD 1/5/8, get phosphorylated on the membrane through serine/threonine kinase activity of type I BMP receptor (BMPRI). However, both TGF- β and BMP uses SMAD4, the common mediating partner for all R-SMADs, to further activate downstream gene responses. Signals are expected to be coupled if SMAD4 is saturated with ligand activation and uncoupled if ligand concentrations do not saturate SMAD4.

Using our cell line having reporter for both SMAD4 and its target *ctgf*, BMP4 stimulation was observed to display slightly distinct response profile compared to TGF- β activation, characterized by quicker translocation of SMAD4 and weaker, but still detectable, gene expression response. Knowing response profiles of each ligand separately, a combination of TGF- β -BMP4 ligands were provided and single-cell SMAD4 and *ctgf* dynamics were measured using live cell imaging. Such simultaneous treatment of ligands revealed a response combining features of both stimuli in an additive manner. On SMAD4 level, the average translocation profile revealed

an earlier translocation compared to only TGF- β scenario, reaching to a higher translocation peak level in a more sustained manner.

The gene expression response exhibited stronger and more sustained profile than the case in which ligands were applied individually. This observation may point towards an integrative information processing which may also be linked to previously reported graded (analog) nature of the pathway. However, such observation currently depends on a limited number of experiments and this aspect of the pathway needs further investigation with larger single cell data sets including combination of ligands of varying doses. Such efforts applying conditions that are closer representation of *in vivo* conditions where multiple ligands of the TGF- β superfamily are present at the same time interacting with each other, will provide valuable insight regarding information processing of the pathway.

4.3 Negative feedback and oscillations in TGF- β signaling

Feedback loops are mechanisms connecting the output signal back to its input. Negative feedbacks are main mechanisms that provide robustness for signaling pathways against sustained ligand stimulation. Feedback mechanisms generally control information processing by determining strength and duration of the signal. Both experimental and modeling analysis revealed the significance of feedbacks in different pathways which contributes to formation of transient, sustained or oscillating signaling profiles. Many biological processes such as circadian rhythm, cell cycle, hormone secretion and certain signalling pathways exhibit oscillatory behaviours based on certain feedback topologies. In principle, oscillatory systems need to meet four requirements: negative feedback, time delay, nonlinearity of reaction kinetics and the proximity of antagonistic reaction timescales. Certain pathways exhibit characteristic oscillations due to the complex opposing interplay between activation and the negative feedbacks [Lahav et al., 2004, Tay et al., 2010].

Although TGF- β signaling was predominantly associated with transient responses in response to stimulations, sustained or oscillatory responses, in principle, are also possible due to complex interplay between activation and negative feedback branches [Zi et al., 2012]. Among several negative feedback mechanisms identified, SMAD7 has been well characterized and shown to antagonize TGF- β signaling through a variety of mechanisms. As an early response gene to TGF- β stimulation, SMAD7 is transcriptionally activated to ensure a stable transient activation of the pathway [Kavsak et al., 2000]. In addition, other feedback regulators (Arkadia, Smurf1, Smurf2, SnoN and Ski) contribute further the stabilization of the pathway [Ebisawa et al., 2001, Deheuninck and Luo, 2009]. Recent modeling studies attempted to evaluate the strength of these negative feedback mechanisms, together with SMAD signaling, in terminating signaling at different levels of the pathway. Some studies claimed the existence of oscillating signaling responses also in TGF- β pathway by sampling of an extensive parameter space related to negative feedbacks [Wegner et al., 2012]. They determined the impact of the different feedback components on the potential oscillatory dynamics which revealed SMAD7

4.3. Negative feedback and oscillations in TGF- β signaling

and Smurf2 as the most important candidates responsible for oscillations. However, this expectation of oscillations in the pathway has not clearly been validated since such negative feedback mechanisms may not be strong enough at the endogenous level. Moreover, existing population level measurements were likely to obscure such oscillatory behaviours even if certain individual cells exhibited such profiles. Therefore, it is crucial to investigate responses with single-cell resolution in different cellular culture and stimulation conditions.

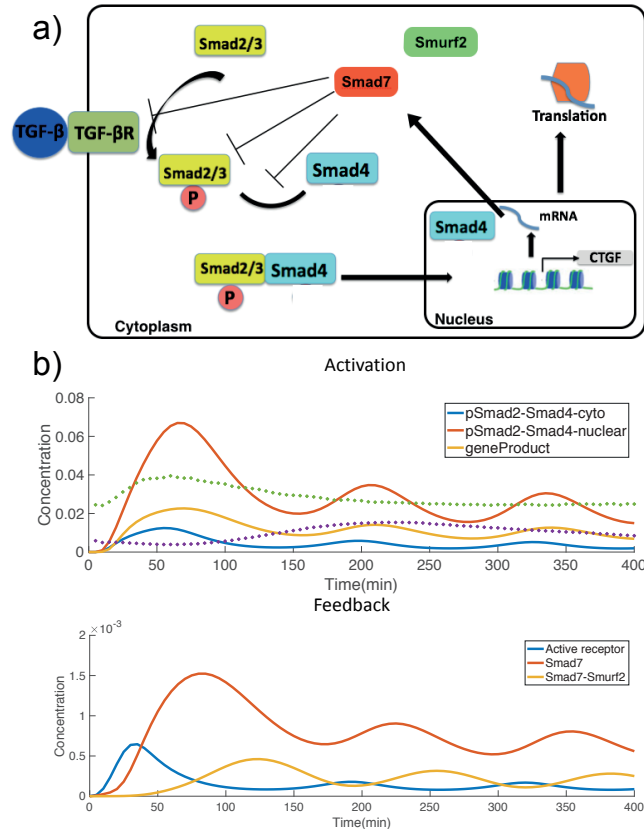


Figure 4.3 – Negative feedback and oscillations in TGF- β signaling. a) Schematic representing both activation and negative feedback branches of TGF- β signaling pathway. b) Time course simulation of the pathway using a previous model encompassing activation, transcription and feedbacks of the pathway in response to TGF- β stimulation at $t=0$ [Wegner et al., 2012]. Dotted lines represent our experimental data of SMAD4 nuclear/cytoplasmic ratio (green) and *ctgf* response (purple). Units are arbitrary both in simulation and experimental data.

Our observation that high SMAD4 levels favor for sustained and even oscillatory responses may contribute to this investigation. Interestingly, previous studies identified only negative feedback components as the determinants of oscillations. However, in our samples with high SMAD4 abundance, almost half of single cells exhibited sustained and oscillatory profile on the target gene expression level. Even the average profile of single cells revealed such oscillations which suggest an underlying robust mechanism. One possible mechanism for high SMAD4 levels resulting in oscillations can be through suppression of SMAD7 expression with nuclear SMAD4-SnoN complexes [Liu et al., 2001]. These complexes have been shown to bind and repress SMAD7 gene promoter which can explain the predominant oscillatory

profiles in SMAD4 abundant cells.

4.4 SMAD modulation of transcriptional regulation and bursting parameters

Suter et al. previously used short-lived reporters to infer the dynamics of endogenous gene from time courses which revealed the bursty nature of transcription in *ctgf* gene, among many others [Suter et al., 2011]. In the follow-up study, Molina et al. investigated the transcriptional profile of *ctgf* gene in response to two stimuli, serum and TGF- β 1 [Molina et al., 2013]. Using telegraph model of gene expression, the transcriptional state of the gene was inferred from protein level traces thanks to destabilized nature of the reporter. In steady state, cells displayed asynchronous bursts of activity while stimuli modulated transcriptional kinetics in certain ways. Both serum and TGF- β 1 stimulation resulted in a rapid and acute increase in burst size which was mediated by slightly different mechanisms. Serum stimulation caused a pronounced first transcriptional burst followed by a refractory period while TGF- β 1 revealed prolonged transcriptional activation due to an increase in transcription rate for extended periods of time. Such prolonged alterations in transcriptional rate may explain subpopulations with sustained *ctgf* responses in our experiments. Cells with higher SMAD4 levels that could overcome the negative feedback barrier may have resulted in sustained expression responses exploiting such prolonged high levels of transcription rate. Moreover, one of the conclusions of the previous study was the dose dependent increase of the transcription rate. In our experiments, similar ligand dose response profiles showed gradual increase in translocation strength. This may suggest that transcription factor modulation on transcriptional bursting occurs through transcription rate rather than burst frequency which ultimately affects the burst size (Figure 4.4, left scenario). BMP4 stimulation experiments may also provide an interesting case since in these experiments, the same amounts of SMAD translocations as with TGF- β 1 stimulation resulted in much lower transcriptional response. This may suggest that transcription factor complexes, SMAD2/3-SMAD4 (TGF- β 1) versus SMAD1/5/8-SMAD4 (BMP4) in this case, distinctly modulate the amplitude of transcriptional response without altering their temporal profiles.

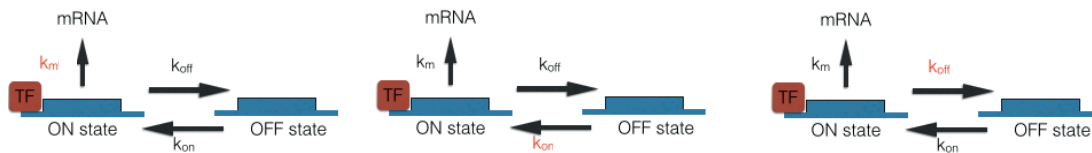


Figure 4.4 – Potential scenarios for SMAD mediated modulation of transcriptional bursting. Temporal profiles of transcription factors can be incorporated into random telegraph model using time-dependent transcriptional bursting parameters. Three scenarios in which transcription factors can modulate transcription rate (left), on rate (middle), off rate (right).

4.5 Implications for *ctgf* function

TGF- β plays crucial roles in different cellular processes such as cell proliferation, differentiation, regeneration and tissue homeostasis. Depending on the cellular context, the number of target genes it can modulate range from a few (pluripotent ES cells) to hundreds (differentiated cells). It can have both activating or suppressing effect on different genes. In principle, TGF- β functions as a potent growth inhibitor leading to cell cycle arrest or apoptosis in normal epithelial cells or at early tumor stages. However, in certain pathologic contexts, cells selectively lose their growth inhibitory responses to TGF- β which converts it from a tumor suppressor into a tumor promotor.

Among other targets, *ctgf* has been reported to respond to TGF- β primarily at the level of transcription. TGF- β stimulation leads SMAD2/3-SMAD4 complex to SMAD4 binding regions of *ctgf* promoter. Together with TGF- β , *ctgf* participates in wound healing process reconstituting a properly arranged connective tissue. However, uncontrolled activity of the gene is generally associated with pathological forms of fibrosis characterized by uncontrolled scarring. Therefore, *ctgf* is only lowly expressed in normal tissue while fibrotic or some cancer tissues exhibit high basal levels of *ctgf*.

In MDA-MB-231 human breast cancer cell line, for instance, *ctgf* has been shown as highly overexpressed promoting bone metastasis in vivo. Due to its angiogenic functionality, high basal levels of *ctgf* provide favorable environments for metastasis when further induced by TGF- β . It has been further shown that canonical TGF- β /SMAD pathway is essential for such metastatic activity characterized by high *ctgf* levels. Depletion of SMAD4 in these cells caused a substantial reduction on metastatic potential while restoration or exogenous expressions of SMAD4 restored metastatic activity [Kang et al., 2003, 2005b]. These observations may emphasize the importance of our observations revealing that higher SMAD4 levels shift *ctgf* responses from a transient to more sustained profile. Higher levels of SMAD4 in some subpopulations of cells may lead to overexpression of *ctgf* which may further provide selective advantage favoring growth of tumor cell population. Taken together, the canonical TGF- β /SMAD pathway may emerge as one of the therapeutic targets to be investigated in the context of metastasis.

Also in normal adult fibroblasts, TGF- β controls the expression of *ctgf* which induces fibroblast proliferation and production of extracellular matrix. In contrast to normal fibroblasts, fibrotic tissue samples revealed high expression levels of *ctgf*. In case of scleroderma (systemic sclerosis), a fibrotic disorder, *ctgf* has been shown to be expressed constitutively while this basal expression levels were SMAD-independent [Holmes et al., 2001]. This SMAD-independent mechanism observed in sclerotic lesions suggest anti-*ctgf* treatments as a more plausible treatment strategy in this particular case. However, in another study of benign biliary stricture, a positive correlation between SMAD4, *ctgf* and TGF- β was observed suggesting the importance of TGF- β /SMAD/ *ctgf* signaling pathway [Geng et al., 2008]. And finally, in the context of renal diseases, studies suggest that early TGF- β 1 fibrotic responses such as induc-

tion of *ctgf* expression and E-cadherin down-regulation are crucially dependent on SMAD3 abundance [Phanish et al., 2006]. Taken together, these studies revealed the complex and context dependent nature of TGF- β -induced, SMAD-mediated and *ctgf* related pathological cases which complicates the task of making generalizations. However, previous studies typically utilized traditional biochemical methods to analyze only populations of cells rather than individual cells and at very few time points of an experiment rather than continuously. Therefore, our experimental approach providing time-resolved measurements of different nodes in this signalling pathway with single-cell resolution presents a powerful tool for future investigations.

4.6 Conclusion and Outlook

In this thesis, a novel method allowing to measure simultaneously transcription factor translocation and target gene response dynamics in single cells using a dual-luminescence reporter system was presented. Using this system, the quantitative relationship between transcriptional activators (SMAD2 and SMAD4) and the target gene *ctgf* response induced by the ligand TGF- β was uncovered. Thanks to destabilized nature of firefly luciferase reporter expressed under *ctgf* gene, underlying mechanisms of the transcriptional regulation in this signal transduction pathway was assessed. Combined with statistical and clustering approaches, single-cell time courses were analyzed to determine the extent how much the transcription factor profile is encoded in the transcriptional response.

First, a substantial heterogeneity in single cells regarding both the translocation and the response profiles was quantified while the expression response showed more widespread distributions across the population both temporally and in amplitude, suggesting a possible accumulation of noise along the path. Second, a quantitative clustering framework was demonstrated simplifying the observed variability in target gene responses down to two categories: transient and sustained. Further, these two categories were linked to SMAD abundance which suggested that initial high SMAD4, but not SMAD2, levels favor for sustained and even oscillatory responses in single cells despite the predominant transient nature of the pathway. Third, effects of several variations of the pathway components such as ligand type, ligand dose and consecutive stimulations on both the translocation and the response profile were explored. This provided useful insight into the information processing mechanisms of this particular signal transduction pathway. Finally, mutual correlations of different single cell features both in translocation and the response were checked. Due to several biological and technical factors imposing additional noise on single cell measurements, these single cell features and their correlations still remains to be further investigated. It would be interesting to check these single cell features in different experimental conditions varying ligand dose, ligand type, etc.

Taken together, the experimental approach presented in this study allowed us to link a panoply of single cell features in translocation and target gene response profiles. However,

both on SMAD and *ctgf* level, single-cell time courses were subjected to biological noise in addition to technical noises explained before. On one hand, *ctgf*, which is a highly inducible gene responsive to serum concentrations and shear stress, required careful handling to minimize variability while its wide dynamic range of response provided us a convenient platform to investigate transcriptional regulation. On the other hand, expression levels of SMAD2/4 were fine-tuned to span a wide range owing to its doxycycline-inducible promoter which also required careful optimization of the substrate concentration and experimental conditions. In future experiments, expressing SMAD reporters under constitutively active promoters would be a good alternative for stable signal levels. Yet, our observations of SMAD4 levels modulating response dynamics revealed SMAD4 levels as an important determinant of a mechanism in which upstream activators modulate the target gene response. In a recent study, Strasen et al., using live-cell imaging combined with clustering approach, grouped SMAD signaling time-courses into distinct classes. Therefore, it would be expected that the heterogeneity in SMAD signaling reflects into transcriptional responses as determined in our investigations. Such heterogenous responses were previously reported in some other pathways [Tay et al., 2010, Ryu et al., 2015].

Negative feedback was shown to determine the heterogeneity in SMAD signaling [Strasen et al., 2018]. Therefore in our system, it would be interesting to experimentally test the role of feedback using either siRNA-mediated knockdown or Cas9-mediated gene knock-out of certain negative feedback components such as SMAD7. Such perturbations on feedback mechanism would modulate SMAD signaling and consequently the target gene response. Moreover, approaches such as ChIP-seq to assess SMAD-induced chromatin remodeling and investigations of proteins involved in the pathway, other than SMAD4/2, contributing to gene activity at later times would be useful future research directions to gain a better understanding of the pathway [Coda et al., 2017]. Such measurements coupled with modeling approaches would provide new insights that can contribute to both TGF- β and transcriptional regulation research.

A Appendix

A.1 Supplementary Figures

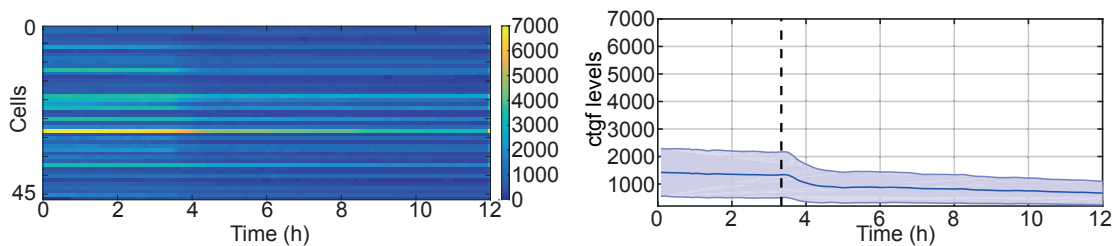


Figure A.1 – Single-cell *ctgf* expression levels in samples treated with TGF- β -receptor agonist SB-431542 (dashed black line) shown in heatmap (left) and as population average (right). The solid line and the shaded area indicate the population average of single cells and the standard deviation, respectively.

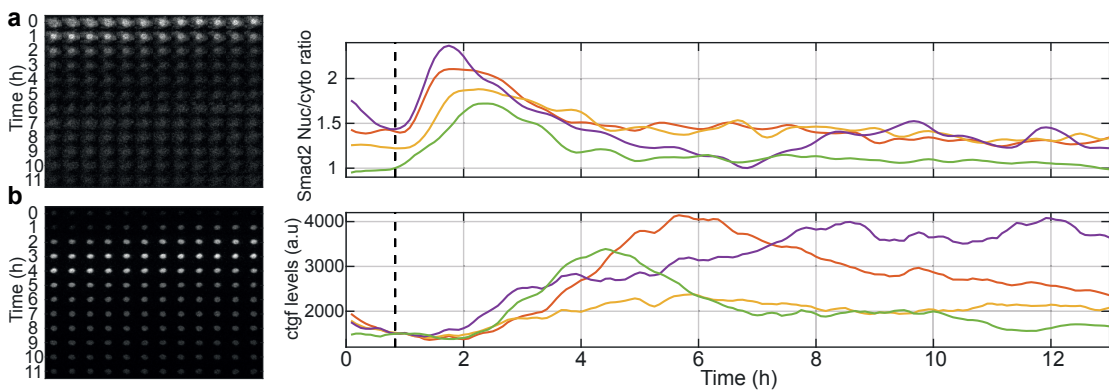


Figure A.2 – Single-cell quantification of *Nluc-SMAD2* cell line in luminescence imaging. Time series images of a tracked single cell of *iS2* cell line on both *Nluc* (left) and *fluc* (right) channel in the luminescence microscopy. Time goes from top left to bottom right, each row represents 1 hour. Images are taken every 5 minutes. Single-cell quantification of SMAD2 translocation (a) and *ctgf* expression level (b) profiles in four individual cells stimulated with TGF- β (5nM, black dashed lines).

Appendix A. Appendix

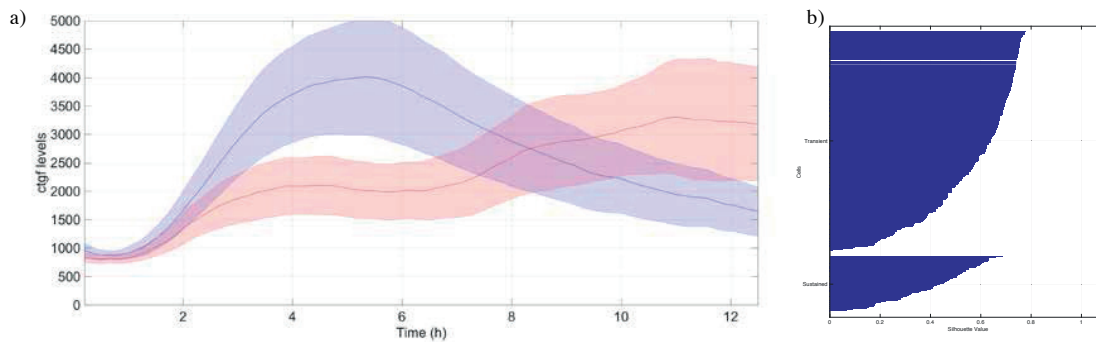


Figure A.3 – ctgf response dynamics decompose into two distinct classes in Nluc-SMAD2 cell line. a) Single-cell traces from TGF- β (5 nM) stimulation experiments (Nluc-SMAD2 cell line, as in Figure 3.16) were categorized into two classes using k-means clustering (N=342 cells). Time courses of the identified response classes showed qualitatively distinct behavior. The solid lines indicate the mean of the cell population while the shaded area represents the standard deviation of the mean. b) Silhouette plot of cells sorted according to ctgf expression dynamics. Plots represent how similar each cell to its own cluster (in our case, transient and sustained) compared to other clusters. Positive silhouette scores indicate that ctgf responses are more similar to their own group, while negative scores suggest that the corresponding trace is closer to the other group.

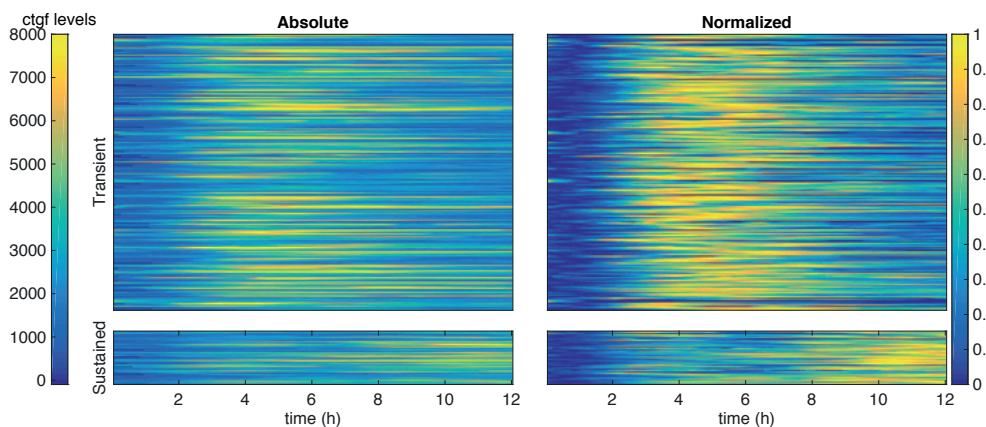


Figure A.4 – Decomposition of cell-specific ctgf responses into transient and sustained classes in Nluc-SMAD2 cell line. Heatmaps for ctgf traces belonging to transient (upper panels) and sustained (lower panels) are shown, quantified as both absolute (left) and normalized (right) levels (min-max scaled, 0 to 1). TGF- β (5 nM) is applied at $t=0$ and ctgf responses in Nluc-SMAD2 cell line is quantified (Dox concentration, 2ng/ml)

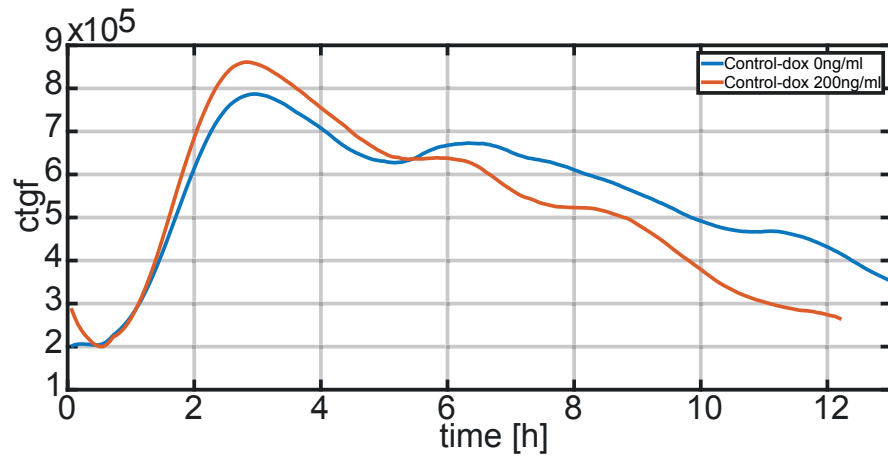


Figure A.5 – Control experiments with non-inducible doxycycline. TGF- β stimulation of 5 nM was applied at $t=0$ to the cell line, pretreated with either no (0 ng/ml) or high dose (200 ng/ml) of doxycycline. Mean ctgf expression responses from two experiments are shown.

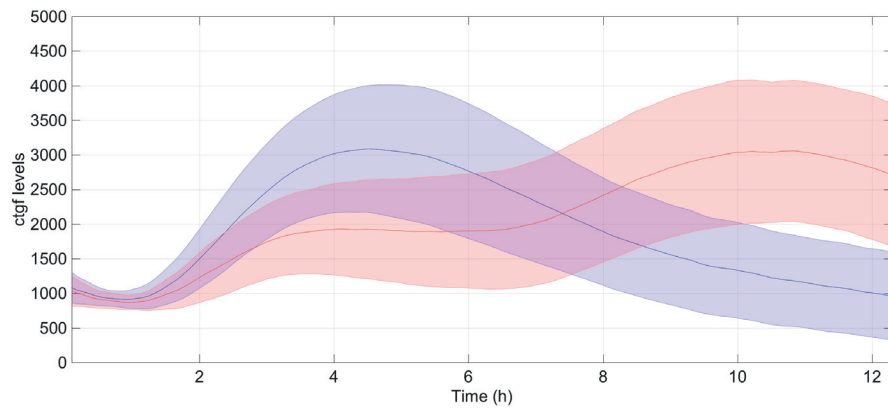


Figure A.6 – ctgf response dynamics decompose into distinct classes in experiments with varying doses of doxycycline stimulation. Single-cell traces from TGF- β stimulation experiments (Nluc-SMAD2 and Nluc-SMAD4 cell line treated with a range of doxycycline concentrations (0-200 ng/ml) were pooled and categorized into two classes using k-means clustering. Average time courses of the identified response classes were shown. The solid lines indicate the mean of the cell population while the shaded area represents the standard deviation of the mean.

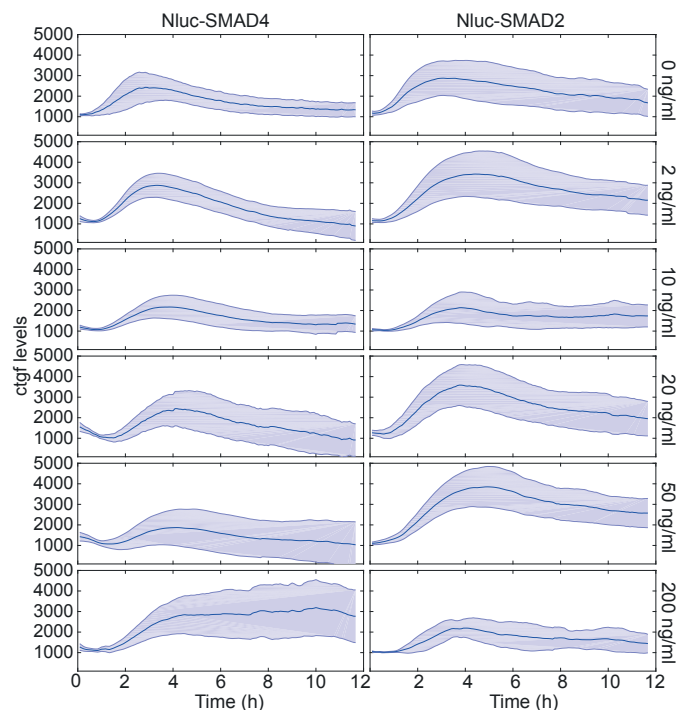


Figure A.7 – Population time courses for ctgf responses upon TGF- β stimulation in samples treated with varying doses of doxycycline (0-2-10-20-50-200 ng/ml) are shown. TGF- β stimulation is at $t=0$. Nluc-SMAD4 cell line showed distinct long-term ctgf dynamics depending on SMAD4 levels (left panels). In Nluc-SMAD2 cell line, no difference was found between samples regarding long-term dynamics displaying predominantly transient profiles (right panels).

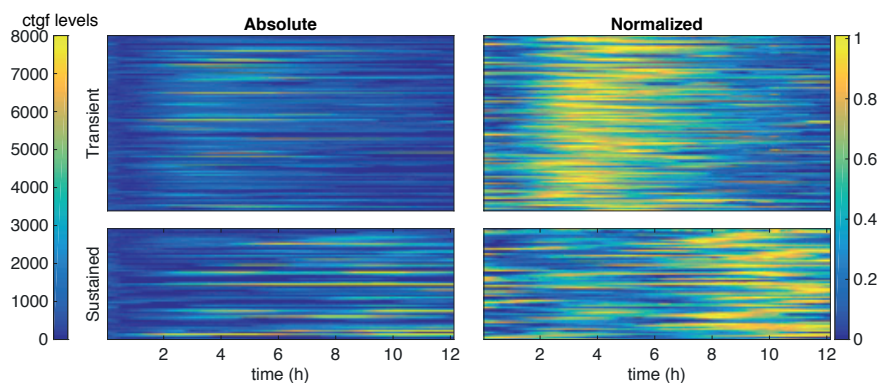
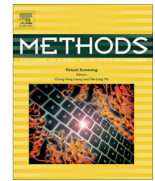


Figure A.8 – Decomposition of cell-specific ctgf responses into transient and sustained classes in case of high SMAD4 abundance when samples treated with 50 ng/ml doxycycline. Heat maps for ctgf traces belonging to transient (upper panels) and sustained (lower panels) are shown, quantified as both absolute (left) and normalized (right) levels (min-max scaled, 0 to 1). TGF- β (5 nM) is applied at $t=0$ and ctgf responses in Nluc-SMAD4 cell line is quantified

A.2 CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings



CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings



Simon Blanchoud¹, Damien Nicolas, Benjamin Zoller, Onur Tidin, Félix Naef^{*}

The Institute of Bioengineering (IBI), School of Life Sciences, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland

ARTICLE INFO

Article history:

Received 16 February 2015
Received in revised form 14 April 2015
Accepted 21 April 2015
Available online 28 April 2015

Keywords:

Time-lapse microscopy
Image analysis
Single-cell analysis
Quantitative transcription

ABSTRACT

Fluorescence and bioluminescence time-lapse imaging allows to investigate a vast range of cellular processes at single-cell or even subcellular resolution. In particular, time-lapse imaging can provide uniquely detailed information on the fine kinetics of transcription, as well as on biological oscillations such as the circadian and cell cycles. However, we face a paucity of automated methods to quantify time-lapse imaging data with single-cell precision, notably throughout multiple cell cycles. We developed CAST (Cell Automated Segmentation and Tracking platform) to automatically and robustly detect the position and size of cells or nuclei, quantify the corresponding light signals, while taking into account both cell divisions (lineage tracking) and migration events. We present here how CAST analyzes bioluminescence data from a short-lived transcriptional luciferase reporter. However, our flexible and modular implementation makes it easily adaptable to a wide variety of time-lapse recordings. We exemplify how CAST efficiently quantifies single-cell gene expression over multiple cell cycles using mouse NIH3T3 culture cells with a luminescence expression driven by the *Bmal1* promoter, a central gene of the circadian oscillator. We further illustrate how such data can be used to quantify transcriptional bursting in conditions of lengthened circadian period, revealing thereby remarkably similar bursting signature compared to the endogenous circadian condition despite marked period lengthening. In summary, we establish CAST as novel tool for the efficient segmentation, signal quantification, and tracking of time-lapse images from mammalian cell culture.

© 2015 Published by Elsevier Inc.

1. Introduction

Improvements in microscopy techniques has enabled rapid progress in quantitative cell biology [1–7]. Notably in mammalian cells, time-lapse imaging has provided unprecedented insights into complex problems such as transcriptional bursting [8–11], cell cycle transitions [12,13] and the circadian oscillator in individual cells [14–17]. However, a current bottleneck is that accurate quantification of time-lapse imaging data is typically slow, repetitive, and requires a significant amount of human intervention (reviewed in [18]). Therefore, versatile automated solutions to robustly quantify temporal signals reflecting noisy biological processes are needed. In particular, single-cell analysis typically requires quantifying hundreds of individual expression traces to

overcome inherent stochasticity and gain statistical significance [15,16,19,20]. However, both the segmentation, i.e. the recognition of object in single frames, as well as the tracking, namely the linking of objects corresponding to the same instance over consecutive frames, are challenging computational tasks.

An intuitive and widespread method for localizing objects is to determine a threshold value in the intensity of the image to separate the signal from the background [21–23]. However, this method is particularly sensitive to fluctuations often observed in biological data, such as cell-to-cell variation, low signal images or uneven illumination. Therefore, segmentation methods based on thresholding are often used as an initial step that is refined by other algorithms such as active contours [24], watersheds [25], morphological operations [26–28], or machine learning methods [29]. In addition, segmentation algorithms that use alternative detection strategies were developed, including multi-scale wavelets [30], cross-correlations [31] or likelihood criteria [32]. However, the performances of each detection method is strongly dependent on the type of data to be segmented, in particular for images with low signal-to-noise ratio (SNR) or objects of varying sizes (reviewed in [33]).

Abbreviations: CAST, Cell Automated Segmentation and Tracking platform; OME-TIFF, Open Microscopy Environment Tagged Image File Format; MAD, Median Absolute Deviation; SNR, signal-to-noise ratio.

***** Corresponding author at: AAB 040, Station 15, CH-1015 Lausanne, Switzerland.

E-mail address: felix.naef@epfl.ch (F. Naef).

¹ Present address: Developmental Biology Laboratory, Department of Anatomy, University of Otago, 9054 Dunedin, New Zealand.

<http://dx.doi.org/10.1016/j.ymeth.2015.04.023>

1046-2023/© 2015 Published by Elsevier Inc.

After objects are localized in a sequence of time-lapse frames, tracking them across frames is also a very challenging task, which is often complicated by the motion, disappearance, fusion or splitting of the studied objects. Moreover, tracking suffers from a problem of combinatorial explosion in the number of potential assignments that renders the identification of the global optimum practically unfeasible for biological datasets. Different strategies were developed to circumvent these problems such as fitting Gaussian curves to object intensity values [34], using cross-correlation of consecutive frames [35] or calculating the probability of each set of assignment [36]. Alternatively, a number of greedy algorithms avoid the aforementioned issues by solving the frame-to-frame correspondence problem in a spatially global fashion [37–40]. However, objective comparisons between these methods concluded that these often need to be tailored to a particular problem, and typically require significant manual intervention and curation [18,41–43]. An additional challenge occurs when the signal intensity of an object drops below detectability during certain time intervals (e.g. during the cell cycle [44] or the circadian cycle [14,15]), in which case the algorithm needs to recognize and stitch the track over the resulting gaps. However, while this caveat can be alleviated by using constitutively expressed reporters in multi-channel fluorescent imaging (e.g. a nuclear or membrane marker), this is not easily possible in bioluminescence imaging, which typically permits the acquisition of only one channel.

In our ongoing work on single-cell transcription [8,9], we experienced that none of the available software tools providing implementations of such segmentation and tracking algorithms [45–56] combine the level of accuracy and automation that we are aiming for our bioluminescence data. Consequently, we set out to develop CAST (Cell Automated Segmentation and Tracking platform), a robust and automated image analysis algorithm for the segmentation, tracking and quantification of time-lapse recordings. To illustrate the potential of CAST, we analyze single-cell bioluminescence reporters engineered to study transcriptional bursting, and demonstrate how transcriptional fluctuations can be deciphered using mathematical modeling to dissect the underlying transcription process. Regardless of this particular application, CAST was conceived in a highly modular fashion, so that its implementation is versatile enough to be readily fitted to many recording configurations, including fluorescence imaging.

2. Material and methods

CAST was developed as a set of custom MATLAB functions accessed through user-friendly graphical interfaces (see Appendix A). Importantly, while we here detail its applicability on bioluminescence data, its modular implementation can be easily adapted to the analysis of a wide variety of time-lapse data. CAST consists of four main steps (described below) that are configurable, editable and, once the corresponding parameters are manually tuned to the specificity of the analyzed recording, fully automatized. Together, these provide a robust way to track the reporter signals in mammalian culture cells using a single emission channel. All code is freely available via the Github service (<https://git.epfl.ch/repo/cast.git>).

2.1. Preprocessing of images

For maximal compatibility between the numerous existing types of biological data, CAST starts by converting the time-lapse recording to be analyzed to a standardized format (unsigned 16-bits OME-TIFF stacks, [57]). Three filtering steps can then be applied to the resulting images. First, a background correction

can compensate for non-uniform illumination of the images. The background is obtained, for each image of the recording independently, by morphologically opening the raw image, smoothing it using a Gaussian kernel, and least-squares fitting of a 2D quadratic surface on the resulting intensities. Of note, the fitted background, that is then subtracted to the raw image, is centered (i.e. has a mean of zero) to remove trends from the image without altering the range of values. Second, CAST can remove saturated pixels from cosmic rays (useful for bioluminescence when long exposure times are used) with a local histogram approach [58], using robust statistics for the estimation of the location and scale (i.e. using the median instead of the mean and the spread, instead of the standard deviation). Thirdly, CAST can remove defective pixels by detecting them as a signal varying more from the mean of the whole image than a threshold defined times the standard deviation. Finally, the entire stack of images can be normalized (i.e. scaled) to compensate for variation between recordings. The result of all these operations is then stored in a new image stack that is used for all the subsequent steps of the analysis.

2.2. Image segmentation

The steps described below are optimized for detection of spherical objects as typically encountered in bioluminescence imaging (e.g. cells, nuclei or other organelles depending on the type of reporters and image resolution), however, these steps can straightforwardly be adapted to segment different types of objects (e.g. non-spherical). As the first step towards segmentation, a background subtraction (as described in 2.1) can be applied, here affecting only this step of the analysis. Second, default de-noising is performed by applying a Gaussian filter with a radius of 0.6 px [59] and by subtracting the estimated mean of the uniform Gaussian white noise, replacing the homogeneity analyzer proposed in [60,61], which is utilized to identify the empty portions of the image, by the Absolute Difference Mask (ADM) edge detector [62]. Candidate object locations are then detected using the “à trous” wavelet transform [30], followed by a filtering step in which only single-pixel local maxima detections are kept. The bioluminescence signal corresponding to each of the detected position is then estimated by iterated least-square fitting of a 2D symmetric Gaussian function (based on [63]). Spurious detections are then filtered out using size and intensity thresholds (all thresholds come with default values but can be changed by the user). Finally, overlapping cells are fused together. The position of the new cell substituting the fused objects is estimated using a weighted average of their respective positions, each weight being proportional to the integral of the corresponding signal intensity. The parameters of the signal are then similarly computed by a weighted average of the fused signals, using a Gaussian weighting kernel of their respective distance to the averaged position.

2.3. Cell tracking

Tracking of cells is performed in CAST using a custom, more memory efficient, implementation of an algorithm that solves both the frame-to-frame linking of objects and the problem of assigning trajectories globally using the Hungarian algorithm [40]. Our implementation uses sparse matrices for the previously proposed cost matrices to decrease the memory load of the algorithm, a major bottleneck in the standard implementation, hence permitting the efficient handling of a very large number of trajectories (i.e. tested on more than 5000 trajectories). Thus, CAST can efficiently handle the gap closing (i.e. linking similar objects that were not segmented in some frames), merging and splitting steps

described in [40], and thus provides a powerful framework for accurately tracking cell lineages in mammalian culture cells. Importantly, the splitting step grants an automated approach for tracking both daughter cells of a division, thus providing a full lineaging of the cells present in the recording. Finally, to remove spurious detections of cells, we implemented an additional optional step to filter out short trajectories before the gap closing, merging and splitting steps, thus preventing these glitches from being linked together into a spurious trajectory.

2.4. Path filtering

The obtained cell trajectories are then filtered by three facultative steps. First, short trajectories can be discarded as they typically result from spuriously detected objects. Second, trajectories that split over some frames before merging back together are zipped into a unique average trajectory. Such splitting-merging typically results from numerous detections of the same cell that were not properly fused during the segmentation step described in 2.2. Third, the position and signal of cells missed in some frames (e.g. as a consequence of biological fluctuations in the signal such as transcriptional bursting or circadian oscillations) is interpolated linearly using the information obtained in the gap-closing step described in 2.3. Finally, the signal of all new cell positions (i.e. either zipped or interpolated) is re-estimated on the raw data. When that new estimate of the signal meets the size-based thresholds described in 2.2, it will then be used to replace the previously interpolated values. Otherwise, the intensity alone is estimated on the raw data for a cell with the interpolated size characteristics to replace the interpolated value. Importantly, no intensity threshold is applied during the re-estimation step to permit quantification of low intensity cells.

2.5. Cell culture

Time-lapse luminescence microscopy was performed on two previously described strains of NIH3T3 mouse fibroblasts [8]. The *Ctgf* fibroblast clone (Figs. 1 and 2A) consists in a gene trap insertion of the short-lived luciferase into the 5th exon of the *Ctgf* gene (connective tissue growth factor, [8,9]), whereas the *Bmal1* strain (Fig. 2B) was generated by transduction of a lentiviral vector carrying the –1014/129 region of the mouse *Bmal1* promoter driving the expression of the short-lived luciferase reporter [8]. To illustrate applicability of CAST to other cell types, two bioluminescent mouse ES cell lines (E14 cells) were engineered using gene trap as previously described [8] (Fig. 2C and D).

Luminescent cells were diluted with non-luminescent NIH3T3 cells 24 h prior to recording at 1:10 and 1:100 ratios for *Ctgf* and *Bmal1* strains, respectively. Cells were plated in a 2.2 cm glass bottom dish (WillCo) to a total of 9×10^5 cells per dish. Prior to luminescence recording, *Bmal1* cells were synchronized for 2 h at 37 °C in 50% horse-serum (BioConcept). Right before recording, both cell strains were washed with PBS, and placed in white DMEM supplemented with 10% serum and D-luciferin (Lifetechnologies) to a final concentration of 1 mM. In lengthened circadian period condition, longdaysin (Sigma) was added to the medium at a concentration of 4 μ M. In control conditions, this volume was compensated with dimethyl sulfoxide (DMSO, Thermo Scientific).

2.6. Bioluminescence time-lapse imaging

Luminescence monitoring was performed in a LuminoView LV200 microscope (Olympus) equipped with a cooled C9100-13 EM-CCD camera (Hamamatsu). During the recording, cells were maintained at 37 °C in a humid environment with 5% CO₂. Since

Ctgf cells displayed elevated signal [8,9], they were monitored using a 40 \times objective, binning 1 and an exposure time and time resolution of either 2 min (Fig. 2A) or 5 min to reach a higher SNR (Fig. 1). The *Bmal1* cells, for which the luminescence signal is typically lower [8], were monitored with an exposure time and time resolution of 5 min, binning 4 and a 20 \times objective (Fig. 2B). Since ES cells have smaller size, these were imaged with an exposure time and time resolution of 5 min, binning 1 and using either a 40 \times (Fig. 2C) or a 60 \times (Fig. 2D) objective. All cell lines were imaged using an EM gain of 150 and 5 \times photon imaging mode (Hamamatsu camera setting). Together, the different imaging conditions provided a good test set for CAST's versatile applicability.

3. Theory

To highlight the applicability and the benefits of CAST regarding single-cell bioluminescence time-lapse imaging, we performed quantitative analysis of transcription kinetics based on stochastic mathematical modeling of short-lived luciferase reporters driven by an individual promoter in mouse fibroblasts.

Following our previous studies (details in [8,9]), we aimed to decipher the kinetics of the processes underlying the temporal fluctuations of luminescence signals, using a three-layer stochastic gene expression model, accounting for translation, transcription and gene activation (Fig. 4). Briefly, the gene is assumed to switch between an active ("on") and an inactive ("off") state, potentially leading to transcriptional bursting, i.e. short and episodic events of transcription further amplified by translation resulting in large fluctuations at the protein level. The model is parameterized by six kinetic rates (k_p , γ_p , k_m , γ_m , k_{on} , k_{off}) assumed constant in time and identical among the cell population. Three kinetic parameters (k_p , γ_p , γ_m) were previously measured [8], while the others three describing the transcriptional kinetics (k_m , k_{on} , k_{off}) were to be inferred from the time-lapse imaging.

Each luminescence time-traces S was modeled as a Hidden-Markov Model (HMM, [64]) whose hidden states were encoded by pmg , namely the protein copy number (p), the transcript copy number (m) and the gene state (g , "on"/"off") respectively. The probabilities of the measured light intensity s given p protein copies, $P_e(s|p)$ followed from the protein-light calibration and the microscope noise model, while the transition probabilities $P_t(pmg|p'm'g')$ were dictated by the master equation of the stochastic gene expression model (see [8,9]). The likelihood of a single luciferase time-traces consisting of L time frames, $S = (s_1, \dots, s_L)$, is then given by

$$L(S|\Theta) = \sum_{\{pmg\}} \prod_{i=1}^{L-1} P_e(s_i|p_i) P_t(p_i m_i g_i | p_{i-1} m_{i-1} g_{i-1}; \Theta) P(s_0 p_0 m_0 g_0)$$

where the sum runs over all hidden trajectories $\{pmg\}$. The likelihood for the cell population is then simply given by the product of each single-cell likelihood. We inferred the unknown kinetic parameters of the population $\Theta = (k_m, k_{on}, k_{off})$ adopting a Bayesian approach by estimating the posterior distribution of the parameters:

$$P(\Theta|S) = \frac{L(S|\Theta)P(\Theta)}{\int L(S|\Theta)P(\Theta)d\Theta}$$

We sampled the posterior $P(\Theta|S)$ according to a Markov-Chain Monte Carlo algorithm (MCMC, [65]). Inferred parameters were estimated as the mean of the posterior distribution. As in our previous work [8,9], we used the optimal model as a prior to deconvolve the bioluminescence traces, by means of Gibbs sampling [65]. This allowed us to reconstruct most likely states for the protein, transcript and gene activity all along the traces.

4. Results

We first illustrate the different steps performed by CAST to automatically segment, track and quantify objects in time-lapse recordings using an unstable bioluminescence reporter in NIH3T3 culture cells [8]. Secondly, we demonstrate the robustness of our approach to a variety of imaging conditions. Thirdly, we exemplify how CAST can follow cells over various cell cycles. Finally, we study transcriptional bursting driven by the *Bmal1* promoter using data extracted by CAST. In particular we compare transcriptional bursting in normal and longdaysin treated cells, a conditions that lengthens the period of the circadian oscillator [66].

4.1. Quantifying bioluminescence time-lapse imaging using CAST

To exemplify the steps described in Section 2, CAST was utilized to analyze a time-lapse recording of bioluminescent NIH3T3 culture cells in which an unstable, nuclear, transcriptional luciferase reporter was driven by the *Ctgf* promoter (Fig. 1A, Section 2). As detailed above, CAST first preprocesses the recording (Fig. 1B) by converting the original movie into a standardized format and applying facultative image filters, such as corrections for non-uniform background correction, cosmic rays and defective pixels. Second, the image was segmented (Fig. 1C, D and D') and the signal and sizes corresponding to individual objects (i.e. individual nuclei) were estimated. Next, the algorithm proceeded to the tracking step (Fig. 1E, E' and F), during which it computed the trajectories of individual cells, taking into account cellular divisions, and assembled objects detected in different frames into cell traces. CAST also closed recording gaps by interpolating the trajectory of cells that transiently disappeared due to transcriptional discontinuities [8,9]. Finally, during the path filtering step, the

software discarded short traces resulting from sporadic detections, and merged trajectories arising from accidental over-segmentation of a single object. Importantly, each of these steps is fully automated, making CAST a platform to straightforwardly follow cells throughout a recording and thus thoroughly quantify the corresponding expression signals.

4.2. CAST can be applied to a range of cell-lines and imaging conditions

Time-lapse imaging is a broadly used technic across many research fields. Consequently, resulting movies can widely differ, notably in terms of image quality and resolution. Also, the biological purpose of a recording may require specific culturing conditions (e.g. confluence cells, presence or absence of cell divisions), as well as distinct requirements regarding downstream analyses (e.g. positions or sizes of the tracked objects). CAST was designed as a highly flexible tool with tunable parameters, hence adaptable to a broad range of requirements. Here, we illustrate the versatility of our platform by analyzing time-laps recording from different experimental and imaging conditions, reporter strengths and cell types.

As a first example, we used CAST to analyze recordings of NIH3T3 cells expressing a short-lived luciferase driven by the strong endogenous *Ctgf* promoter [8,9]. To illustrate CAST capacities in tracking motile objects, cells were plated at medium confluency, allowing them to freely migrate during the 2 days of recording while dividing only very rarely (Section 2). CAST efficiently segmented and followed each individual cell throughout the recording (Fig. 2A).

Second, we analyzed recordings of another NIH3T3 clone expressing a luminescence reporter driven by a *Bmal1* promoter [8]. Cells carrying this weaker promoter typically display lower SNR and thus required higher camera binning (Section 2).

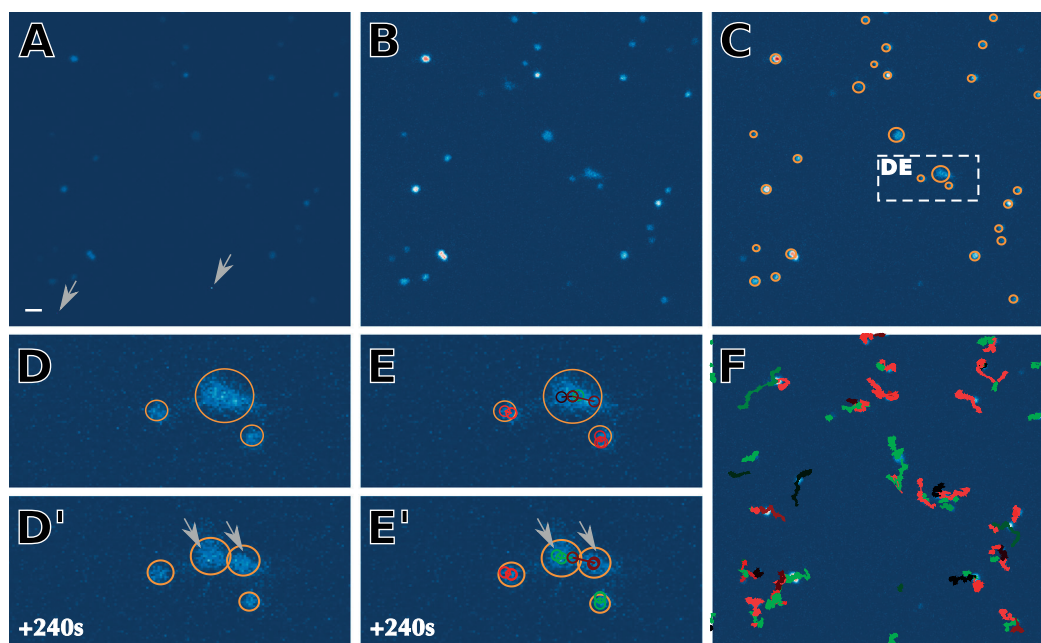


Fig. 1. Sequential analysis of a time-lapse recording using CAST. (A) Raw image (40 \times) of NIH3T3 cells expressing a luminescence reporter from a *Ctgf* promoter. The white scale bar represents 10 μ m. The gray arrows point at cosmic rays. (B) The processing step removes, among other filters, cosmic rays. (C) CAST detects signal-emitting cells, the width of which is represented using orange circles. The white dashed box highlights the portion of the image that is magnified in panels D and E. (D) Both the position as well as the signal of the emitting cells is properly determined by CAST. (E) CAST tracks cells over consecutive frames. (D'–E') Same steps as D and E but applied to the next frame of the recording, thus exemplifying the detection of a dividing cell (gray arrows show daughter cells). (F) Full tracks of all identified cells throughout the recording. Each individual track is color-coded using a randomly chosen color in a red–black–green color-code.

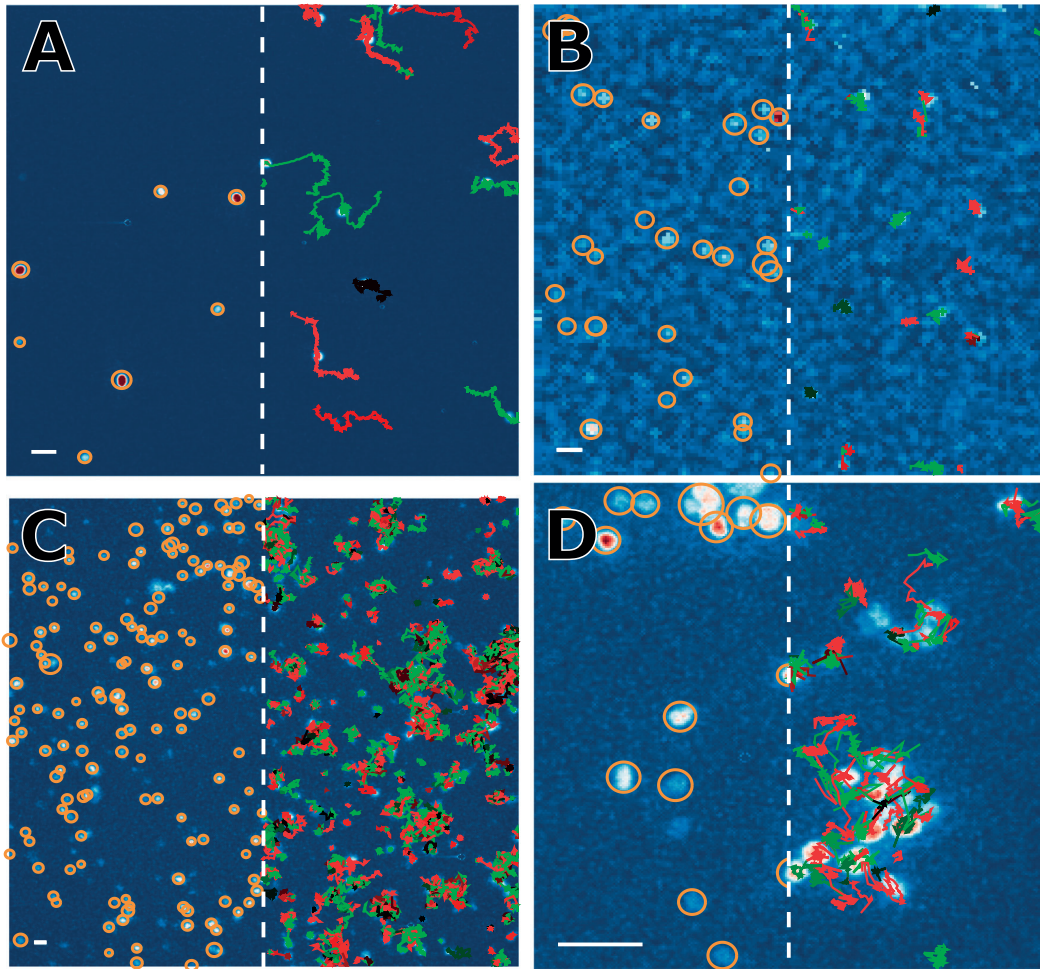


Fig. 2. CAST can track cells in a variety of imaging conditions and cell types. In each panel, the left side represents a frame after the segmentation step (detected objects highlighted in orange), while the right side corresponds to the same frame after tracking (single-cell trajectories color-coded using a randomly chosen color in a red–black–green color-code). All white bars represent 10 μm . (A) NIH3T3 *Ctgf* clones presenting high mobility and high SNR. (B) NIH3T3 *Bmal1* clones displaying low SNR and lower resolution. (C) High density of ES cells monitored in low spatial resolution (10 \times objective). (D) High resolution recording of ES cells (60 \times objective). Note that the cells deviate from circular (Gaussian-like) shapes at this resolution but are still detected correctly.

Nevertheless, CAST thoroughly detects the weaker and lower resolution signals in this recording (Fig. 2B).

Third, CAST analysis was applied on high-density time-lapse images monitoring mouse ES cells expressing a luciferase reporter (Section 2). Despite this higher density, a large majority of cells were successfully segmented properly, even in cases of partial overlaps (Fig. 2C).

Finally, we monitored the same ES cells using a 60 \times objective. At this higher resolution, the irregular profiles of the luminescent objects (i.e. the nuclei) are very apparent, hence making them deviate from circular (i.e. Gaussian-like) shapes. Notwithstanding, CAST correctly detects the vast majority of cells (Fig. 2D).

4.3. CAST robustly quantifies emission signals over several cell cycles

Keeping track of dividing cells through multiple days recordings is crucial for many applications involving time-lapse microscopy [12–15]. An important objective of our platform was thus to efficiently and robustly follow cell lineages.

To illustrate CAST capacities in tracking dividing cells, we recorded *Ctgf* clones in proliferating conditions (Section 2). During the nearly 70 h of recording, individual cells underwent up to two division events, resulting in the presence of four daughter cells by the end of the movie (Fig. 3A–C). CAST efficiently detected divisions and thus reconstructed the corresponding cell lineage throughout the recording (Fig. 3D). The provided output features facilitated the graphical representation of dividing cells, such as their positions over time (Fig. 3D), or the signal intensity of each daughter cell (Fig. 3E). Importantly, CAST's robust tracking capacities (i.e. performing gap closure) permit to overcome the loss of signal during mitosis, which is caused by the dilution of the nuclear reporter into the cytoplasm upon nuclear envelope breakdown (Fig. 3E).

4.4. The bursting kinetics of *Bmal1* promoter in normal vs. longdaysin-treated culture cells are similar

In the past years, time-laps microscopy has become a method of choice to monitor transcriptional bursting (reviewed in [67,68]).

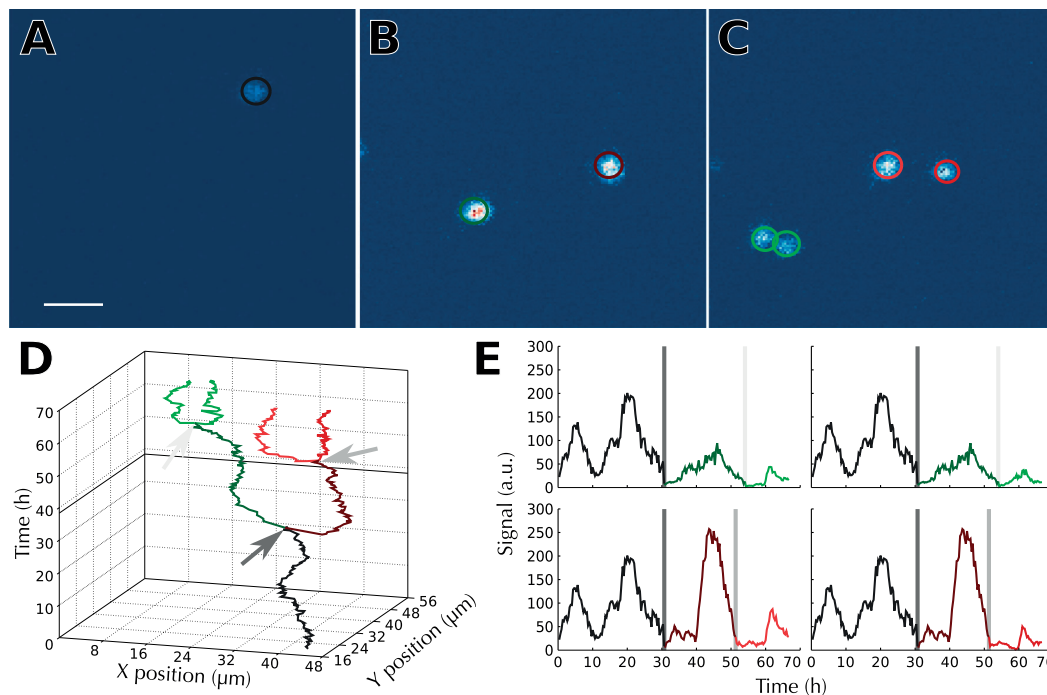


Fig. 3. CAST tracks cells over several cellular divisions. (A) Magnification of the first frame of the recording used in Fig. 1 showing a parental cell that will undergo two successive divisions, overlaid with its estimated size color-coded in black. The white bar represents 10 μm . (B) Intermediate and (C) final frames of the same region of the recording as in A, depicting the various cells originating from the single ancestor. Each detected cell is overlaid with its estimated size, color-coded in tones of red or green. (D) Diagram of the X–Y movements of the dividing cells throughout the recording. The 2D trajectory of each cell is color-coded as in A–C. Cellular division events are indicated by gray arrows. The full black line, visible slightly before 40 h on the time axis, indicates the position of the frame depicted in B. (E) Quantification of the bioluminescence signal emitted by the cells throughout the recording. Note that because of the shared ancestors, the darker portions of the signal are duplicated between traces. Cellular divisions are indicated by vertical gray bars, color-coded as in D.

This biological phenomenon arises from the sequential series of molecular events inherent to gene expression, resulting in the stochastic synthesis of downstream gene products (reviewed in [69]). Indeed, transcription of individual mammalian genes frequently occurs discontinuously in the form of short active periods, termed bursts, followed by windows of transcriptional inactivity.

In previous studies, we used single-cell level real-time measurements of a short-lived nuclear luciferase from a collection of mammalian promoters to monitor gene transcription [8,9]. Combined with a mathematical model [70,71], these data allowed us to directly estimate the transcriptional kinetics underlying the transcriptional bursting pattern of endogenous and synthetic genes. We notably described the signature activity of the *Bmal1* promoter, a master regulator of the mammalian circadian clock.

We now use CAST to segment and track new recordings, thus avoiding the tedious work of manual tracking. In particular, we aimed at studying how the endogenous transcriptional bursting behavior of the *Bmal1* promoter is altered upon longdaysin treatment. As this compound lengthens the circadian period in a dose-dependent manner, by delaying the degradation of the PER proteins through CK1 δ inhibition [66], it is of interest to understand how this stretching is reflected at the level of transcriptional parameters, such as the durations of “on” and “off” times, the transcription rate, or the burst sizes.

To validate the segmentation and tracking capacities of CAST on such data, we first compared traces of individual cells with their manually tracked equivalents (Fig. 4A). The similarity between traces obtained with CAST and manual tracking confirmed that

the platform could recover data of similar quality as obtained manually, as well as much faster (typically 20 min for a whole recording containing 50 traces compared to 5 min for a single manual trace). Moreover, CAST allowed us to increase the number of cell traces extracted from a recording (compared to manual tracking), notably due to the close proximity of some objects making it too delicate to follow each one of them by eye. In fact, automated tracking increased by more than 20% the amount of exploitable data per recording.

As a validation, the population average of individual traces resulting from CAST processing both wild type and longdaysin treated cells successfully recapitulated the phenotypes previously described for cell populations (Fig. 4B, [66]). Indeed, we clearly observed a stretch in the circadian period upon longdaysin treatment, increasing it from 25 h in endogenous circadian conditions to 32 h after treating cells with 4 μM of the compound.

We then analyzed the collection of single-cell traces obtained both in control and longdaysin conditions using our mathematical modeling approach (Theory, [8,9]). This method consists in inferring the amount of luciferase proteins, transcripts and *Bmal1* promoter activity that most appropriately reflect the luminescence signal observed in the recording (Fig. 4C). By analyzing over a hundred single-cells, we could deduce the transcriptional kinetics defining the bursting signature of the *Bmal1* gene in both conditions (Fig. 4D). Interestingly, we did not observe major differences between the bursting kinetics of cells cultured in an endogenous circadian context and those enduring lengthened period, suggesting that longdaysin treatment does not affect the transcriptional bursting characteristics of *Bmal1*.

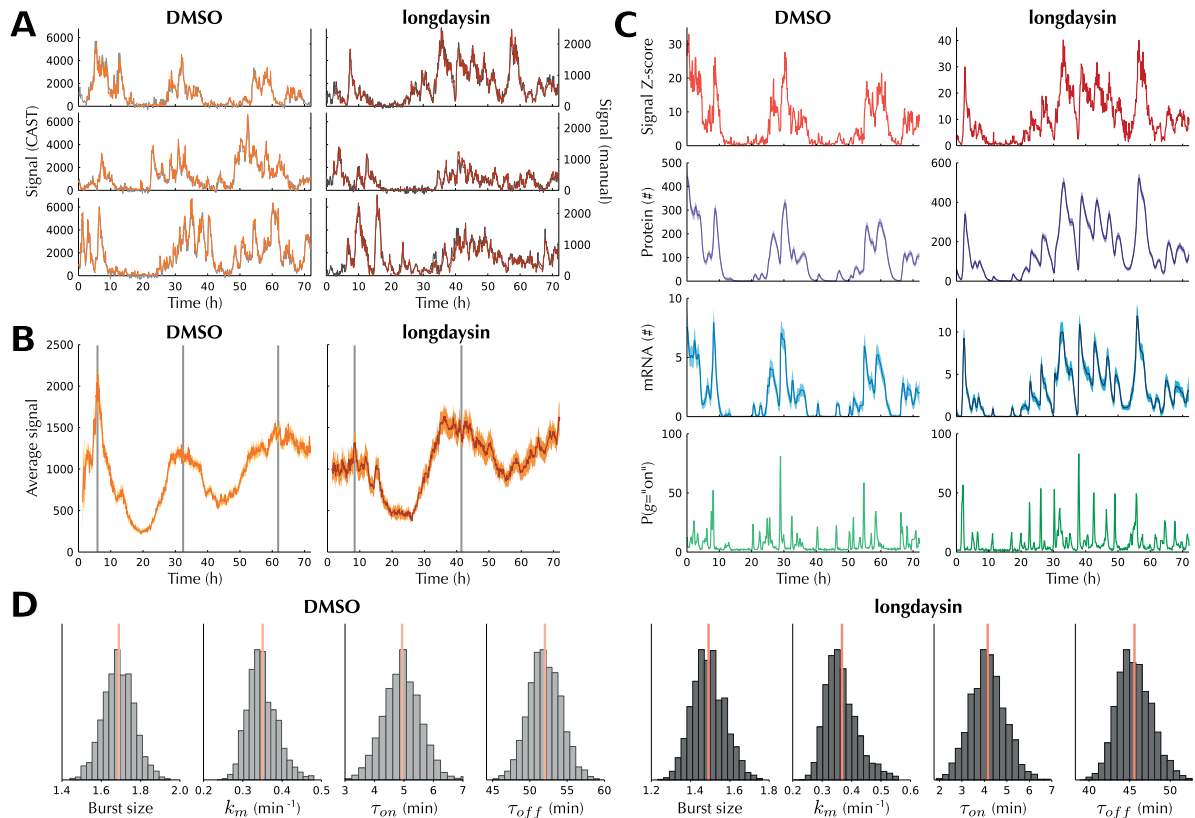


Fig. 4. Comparison between the transcriptional bursting pattern of *Bmal1* in control and longdaysin conditions (A) superposition of three typical *Bmal1* single-cell traces obtained by CAST (brown lines, left axis) and manual tracking (gray lines, right axis) in both endogenous (DMSO, left column) and lengthened circadian period (red, right column). (B) Average *Bmal1* signal displayed by individual cells tracked with CAST, both in control condition (orange, $n = 127$) and in $4 \mu\text{M}$ longdaysin condition (brown, $n = 108$), recapitulating the phenotype of a population of cells. Orange areas delimit the standard error and vertical grey bars estimate the position of expression peaks (C) Example of deconvolved single-cell traces obtained from our mathematical modeling approach (control condition on the lighter left column, and longdaysin $4 \mu\text{M}$ on the darker right one). From the raw luminescence signal (first row in panel A, corresponding Z-score in red), we infer the amount of luciferase proteins (purple) and transcripts (blue), as well as the probability of gene activity (green). (D) Transcriptional bursting kinetics of *Bmal1* in endogenous (gray) and lengthened (black) circadian conditions. Displayed kinetics are the mean number of transcripts produced per burst (burst size), the transcription rate (k_m), as well as the mean time spent in active (τ_{on}) and inactive states (τ_{off}).

5. Discussion

5.1. CAST can quantify and track cells in a broad range of time-lapse imaging data

Biological research, and in particular transcription regulation, has greatly benefitted from quantitative approaches and mathematical modeling [8,68,72,11,9]. However, the extraction of data with sufficient precision as required for quantitative analysis is typically highly time consuming, mostly because of the significant degree of manual intervention required [18]. To circumvent this bottleneck, we have developed CAST, the automated analytical platform presented here (Fig. 1), which quantifies robustly the expression signal of culture cells without the need of manual curation. Moreover, we showed that CAST could adapt to a variety of cell types and imaging conditions, hence providing a general analytical framework for study of bioluminescence time-lapse data (Fig. 2). Furthermore, we demonstrate CAST performances in tracking cell divisions, which makes it a powerful tool for long-term experiments covering several cell cycles (Fig. 3). Notably, given that cell divisions typically introduce a perturbation in the expression trace (Fig. 3E), detecting these events is of importance to discriminate the corresponding alterations from genuine variations in gene expression (e.g. in stimulation experiments, [9]).

While the robustness to variations in SNR and cell size is currently provided at the segmentation step for Gaussian-like objects [30], the modular architecture of CAST allows one to easily implement alternative segmentation algorithms, thus permitting the analysis of different types of data such as fluorescence imaging. Similarly, while the provided tracking algorithm [40] is highly versatile and can be applied to a large variety of biological data, it could also straightforwardly be extended higher dimensional datasets (e.g. 3D + time [3]). Importantly, many fields of research could directly benefit from the segmentation and tracking capabilities of CAST. Overall, CAST is a highly versatile algorithm with an intuitive user-friendly graphical interface that can be used to solve a large number of segmentation, tracking and quantification problems.

5.2. Transcriptional bursting of *Bmal1* upon longdaysin or TSA treatment is similar

To exemplify CAST capacities, we used it to monitor the transcriptional bursting pattern of *Bmal1* (Fig. 4). The software segmented, tracked and quantified the expression of bioluminescent reporter, thus providing single-cell traces suitable for mathematical modeling. This analysis allowed us to observe two remarkable effects associated with the alterations in *Bmal1* expression upon longdaysin treatment. First, and in agreement with previous

studies [15,66], longdaysin stretched the rhythmic expression pattern of *Bmal1* (Fig. 4B). Second, the compound did not affect the transcriptional bursting signature of the gene (Fig. 4D).

These results are strikingly similar to the ones previously obtained by treating similar cells with Trichostatin A (TSA), an histone deacetylase inhibitor that promotes a transcriptionally favorable chromatin state [8]. Indeed, while TSA treatment abolished the periods where *Bmal1* is repressed, the fine kinetics of transcriptional bursting remained virtually unchanged compared to the control condition. Together, these results imply that the bursting signature of *Bmal1* is dominated by the impact its transcriptional activators, and not that of repressive mechanisms that are directly linked to the activity of histone deacetylases, which are inhibited by TSA [73]. Consequently, modifying the circadian phase at which *Bmal1* activators operate (e.g. using TSA or longdaysin), is not sufficient to observe major changes in the bursting signature of the gene. Thus, to further understand the role of molecular factors in shaping the bursting signature of *Bmal1*, it would be interesting to directly modulate the activity of *Bmal1* transcriptional activators.

5.3. Advantages and limitations of CAST

Biological time-lapse imaging data display great variability in terms of reporter type, object shape, image quality and signal dynamics (both in time and space). Consequently, it is often problematic to find an analytic platform able to suitably segment and track the signal, while providing a quantitative output directly applicable for downstream analyses. Indeed, many tools are available either as standalone applications or integrated in generic platforms [45–56], but although they excel in the analysis of some imaging data, they may not represent a perfectly suitable solution for other, less conventional, types of recordings such as bioluminescence imaging.

Notably, we attempted to analyze our *Bmal1* promoter bioluminescence real-time data with popular tracking methods, though with limited success [45,46]. One reason was that many methods lack flexibility as they contain hard-coded limitations that prevented us from obtaining the type of data required for our subsequent analyses. Among other important features, the following three properties were not available simultaneously in any single platform: (1) a complete and automated pipeline (2) proper quantification of signal intensities during intervals of low signals (3) lineage of both daughter cells over multi cell cycles. Indeed, recordings such as the *Bmal1* bioluminescence traces pose significant challenges for popular algorithms, due to the weak signals, the limited resolution of the images (e.g. high binning), or the absence of signal over prolonged periods owing to circadian regulation (in some cases up to 12 h, or almost 150 frames). Consequently, to overcome this limitations, and to avoid the tedious and time-consuming manual tracking used so far [8,9], we developed CAST as a versatile and extensible solution.

CAST was specifically developed to deal with challenging recordings, and comprises several advantages. First, it consists in a self-contained platform that combines all the steps for cell tracking and quantification (i.e. image pre-processing, segmentation, tracking, path filtering and signal quantification). Second, CAST adapted robust and published approaches for each of these steps. Notably, CAST's pre-processing steps treat cosmic rays [58], the segmentation permits the detection of variable-size objects in noisy data [30], its tracking algorithm efficiently links related objects over large gaps or divisions [40] and CAST interpolates the trajectory of objects below detection levels, thus providing reliable estimation of the signal intensity. Third, each step contains tunable parameters, allowing the platform to adapt to a large variety of experiment types. In addition, a graphical interface allows the real-time visualization of CAST performances throughout the

whole procedure, and on any frame of the recording, hence greatly facilitating the tuning of parameters during each step of the analysis. Finally, CAST was developed as a modular Matlab package, thus providing a convenient platform for further improvements to fit more specific requirements.

However, CAST may show limitations depending on the type of subsequent analysis performed. Typically, the large amount of tunable parameters (for which default parameters are provided) may sometimes represent a drawback, notably with more conventional recordings, for which existing tracking are sufficiently powerful. Indeed, selecting the most appropriate parameters may require significant prior knowledge, and be time-consuming. Thus, the use of CAST is mainly recommended in cases where data are delicate to parse, or if the same tracking parameters can be applied to large collections of independent time-lapse recordings. Additionally, because of its development platform, CAST can be slower than other available solutions, and computation time may represent a limitation when treating high-temporal resolution recordings. Finally, although we have successfully used CAST on fluorescence data, the current pipeline was primarily developed to deal with luminescence traces. Consequently, the robust analysis of fluorescence data will most likely require the development of additional modules.

6. Conclusion

We developed CAST as a new versatile tool for the analysis of single-cell imaging data. Moreover, we presented its robust capacity to quantify bioluminescence signals from transcriptional reporters in single, and showed how such data can be combined with mathematical modeling to analyze the fine kinetics of *Bmal1* transcription, a key regulator of the circadian rhythm. Consequently, we are confident that CAST will be widely applicable to many outstanding problems in quantitative single-cell biology.

Acknowledgments

We thank Aleksandra Mandic and David Suter for providing recordings on the mouse ES cells. The Naef laboratory was supported by the Swiss National Science Foundation (SNSF grant 31-130714), the European Research Council (ERC-2010-StG-260667) and the École Polytechnique Fédérale de Lausanne (EPFL). F.N. also received funding from StoNets, a grant from the Swiss SystemsX.ch (www.systemsx.ch) initiative evaluated by the Swiss National Science Foundation (SNSF). O.T. is funded by a SystemsX.ch IPhD grant 51PH_0125979 to David Suter and F.N. The computations were performed at the Vital-IT (<http://www.vital-it.ch>) Center for high-performance computing of the Swiss Institute of Bioinformatics.

Appendix A. CAST graphical user interface

CAST was developed in Matlab and can be utilized through a user-friendly graphical interface (Fig. S1). This interface allows the user to control all the steps of the algorithm detailed in Section 2, and has extensive possibilities to visualize the status of the different steps in the analysis. CAST thus provides a flexible tool to monitor and adapt each step of analysis. In addition, this interface allows users to modify all parameters in the analysis pipeline, which can also be saved and shared between experiments. Finally results of the analysis can be exported to common formats (i.e. spreadsheet and/or movie files).

Time-lapse recordings as well as their respective CAST configuration files (corresponding to Figs. 1 and 2) are freely available via the Github service (<https://git.epfl.ch/repo/cast-supplements.git>).

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ymeth.2015.04.023>.

References

- [1] Z. Bao, J.I. Murray, T.J. Boyle, S.L. Ooi, M.J. Sandel, R.H. Waterston, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 2707–2712, <http://dx.doi.org/10.1073/pnas.0511111103>.
- [2] S. Blanchoud, Y. Budirahardja, F. Naef, P. Gönczy, *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 239 (2010) 3285–3296, <http://dx.doi.org/10.1002/dvdy.22486>.
- [3] P.J. Keller, A.D. Schmidt, A. Santella, K. Khairy, Z. Bao, J. Wittbrodt, et al., *Nat. Methods* 7 (2010) 637–642, <http://dx.doi.org/10.1038/nmeth.1476>.
- [4] N. Olivier, M.A. Luengo-Oroz, L. Duloquin, E. Faure, T. Savy, I. Veilleux, et al., *Science* 329 (2010) 967–971, <http://dx.doi.org/10.1126/science.1189428>.
- [5] N. Déneraud, J. Becker, R. Delgado-Gonzalo, P. Damay, A.S. Rajkumar, M. Unser, et al., *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 15842–15847, <http://dx.doi.org/10.1073/pnas.1308265110>.
- [6] D. Keller, M. Orpinell, N. Olivier, M. Wachsmuth, R. Mahen, R. Wyss, et al., *J. Cell Biol.* 204 (2014) 697–712, <http://dx.doi.org/10.1083/jcb.201307049>.
- [7] S. Blanchoud, C. Busso, F. Naef, P. Gönczy, *Biophys. J.* 108 (2015) 799–809, <http://dx.doi.org/10.1016/j.bpj.2014.12.022>.
- [8] D.M. Suter, N. Molina, D. Gatzfeld, K. Schneider, U. Schibler, F. Naef, *Science* 332 (2011) 472–474, <http://dx.doi.org/10.1126/science.1198817>.
- [9] N. Molina, D.M. Suter, R. Cannavo, B. Zoller, I. Gotic, F. Naef, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 20563–20568, <http://dx.doi.org/10.1073/pnas.1312310110>.
- [10] A.M. Corrigan, J.R. Chubb, *Curr. Biol.* 24 (2014) 205–211, <http://dx.doi.org/10.1016/j.cub.2013.12.011>.
- [11] D.R. Larson, C. Fritzsche, L. Sun, X. Meng, D.S. Lawrence, R.H. Singer, *eLife* 2 (2013), <http://dx.doi.org/10.7554/eLife.00750>.
- [12] D. Gerlich, B. Koch, F. Dupeux, J.-M. Peters, J. Ellenberg, *Curr. Biol.* 16 (2006) 1571–1578, <http://dx.doi.org/10.1016/j.cub.2006.06.068>.
- [13] B. Neumann, T. Walter, J.-K. Heriche, J. Bulkescher, H. Erfle, C. Conrad, et al., *Nature* 464 (2010) 721–727.
- [14] E. Nagoshi, C. Saini, C. Bauer, T. Laroche, F. Naef, U. Schibler, *Cell* 119 (2004) 693–705, <http://dx.doi.org/10.1016/j.cell.2004.11.015>.
- [15] J. Bieler, R. Cannavo, K. Gustafson, C. Gobet, D. Gatzfeld, F. Naef, *Mol. Syst. Biol.* 10 (2014) 739, <http://dx.doi.org/10.15252/msb.20145218>.
- [16] C. Feillet, P. Krusche, F. Tamani, R.C. Janssens, M.J. Downey, P. Martin, et al., *Proc. Natl. Acad. Sci. U.S.A.* (2014) 1320474111, <http://dx.doi.org/10.1073/pnas.1320474111>.
- [17] P. Pulimeno, T. Mannic, D. Sage, L. Giovannoni, P. Salmon, S. Lemeille, et al., *Diabetologia* 56 (2013) 497–507, <http://dx.doi.org/10.1007/s00125-012-2779-7>.
- [18] J.F. Dorn, G. Danuser, G. Yang, *Methods Cell Biol.* 85 (2008) 497–538, [http://dx.doi.org/10.1016/S0091-679X\(08\)85022-4](http://dx.doi.org/10.1016/S0091-679X(08)85022-4).
- [19] J. Yu, J. Xiao, X. Ren, K. Lao, X.S. Xie, *Science* 311 (2006) 1600–1603, <http://dx.doi.org/10.1126/science.1119623>.
- [20] P. Paszek, S. Ryan, L. Ashall, K. Sillitoe, C.V. Harper, D.G. Spiller, et al., *Proc. Natl. Acad. Sci.* 107 (2010) 11644–11649, <http://dx.doi.org/10.1073/pnas.0913798107>.
- [21] K. Wu, D. Gauthier, M.D. Levine, *IEEE Trans. Biomed. Eng.* 42 (1995) 1–12, <http://dx.doi.org/10.1109/10.362924>.
- [22] Y.J. Zhang, *Pattern Recognit.* 29 (1996) 1335–1346, [http://dx.doi.org/10.1016/0031-3203\(95\)00169-7](http://dx.doi.org/10.1016/0031-3203(95)00169-7).
- [23] N. Otsu, *IEEE Trans. Syst. Man Cybern.* 9 (1979) 62–66, <http://dx.doi.org/10.1109/TSMC.1979.4310076>.
- [24] I. Seroussi, D. Veikherman, N. Ofer, S. Yehudai-Resheff, K. Keren, *J. Microsc.* 247 (2012) 137–146, <http://dx.doi.org/10.1111/j.1365-2818.2012.03624.x>.
- [25] F. Meyer, S. Beucher, *J. Vis. Commun. Image Represent.* 1 (1990) 21–46, [http://dx.doi.org/10.1016/1047-3203\(90\)90014-M](http://dx.doi.org/10.1016/1047-3203(90)90014-M).
- [26] N. Malpica, C.O. De Solórzano, J.J. Vaquero, A. Santos, I. Vallcorba, J.M. García-Sagredo, et al., *Cytometry* 28 (1997) 289–297, [http://dx.doi.org/10.1002/\(SICI\)1097-0320\(19970801\)28:4<289::AID-CYTO3>3.0.CO;2-7](http://dx.doi.org/10.1002/(SICI)1097-0320(19970801)28:4<289::AID-CYTO3>3.0.CO;2-7).
- [27] D. Thomann, D.R. Rines, P.K. Sorger, G. Danuser, *J. Microsc.* 208 (2002) 49–64, <http://dx.doi.org/10.1046/j.1365-2818.2002.01066.x>.
- [28] D. Sage, F.R. Neumann, F. Hediger, S.M. Gasser, M. Unser, *IEEE Trans. Image Process.* 14 (2005) 1372–1383, <http://dx.doi.org/10.1109/TIP.2005.852787>.
- [29] S. Jiang, X. Zhou, T. Kirchhausen, S.T.C. Wong, *Cytometry A* 71 (2007) 563–575, <http://dx.doi.org/10.1002/cyto.a.20404>.
- [30] J.-C. Olivo-Marin, *Pattern Recognit.* 35 (2002) 1989–1996, [http://dx.doi.org/10.1016/S0031-3203\(01\)00127-3](http://dx.doi.org/10.1016/S0031-3203(01)00127-3).
- [31] S. Bonneau, M. Dahan, L.D. Cohen, *IEEE Trans. Image Process.* 14 (2005) 1384–1395, <http://dx.doi.org/10.1109/TIP.2005.852794>.
- [32] J. Boulanger, C. Kervran, P. Boutheymy, *Med. Image Anal.* 13 (2009) 132–142, <http://dx.doi.org/10.1016/j.media.2008.06.017>.
- [33] I. Smal, M. Loog, W. Niessen, E. Meijering, *IEEE Trans. Med. Imaging* 29 (2010) 282–301, <http://dx.doi.org/10.1109/TMI.2009.2025127>.
- [34] G.J. Schütz, H. Schindler, T. Schmidt, *Biophys. J.* 73 (1997) 1073–1080, [http://dx.doi.org/10.1016/S0006-3495\(97\)78139-6](http://dx.doi.org/10.1016/S0006-3495(97)78139-6).
- [35] A. Kusumi, Y. Sako, M. Yamamoto, *Biophys. J.* 65 (1993) 2021–2040, [http://dx.doi.org/10.1016/S0006-3495\(93\)81253-0](http://dx.doi.org/10.1016/S0006-3495(93)81253-0).
- [36] D. Reid, *IEEE Trans. Autom. Control* 24 (1979) 843–854, <http://dx.doi.org/10.1109/TAC.1979.1102177>.
- [37] K. Shafiq, M. Shah, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (2005) 51–65, <http://dx.doi.org/10.1109/TPAMI.2005.1>.
- [38] I.F. Sbalzarini, P. Koumoutsakos, *J. Struct. Biol.* 151 (2005) 182–195, <http://dx.doi.org/10.1016/j.jsb.2005.06.002>.
- [39] S. Jiang, X. Zhou, T. Kirchhausen, S.T.C. Wong, *Cytometry A* 71 (2007) 576–584, <http://dx.doi.org/10.1002/cyto.a.20411>.
- [40] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S.L. Schmid, et al., *Nat. Methods* 5 (2008) 695–702, <http://dx.doi.org/10.1038/nmeth.1237>.
- [41] E. Meijering, O. Dzyubachyk, I. Smal, *Methods Enzymol.* 504 (2012) 183–200, <http://dx.doi.org/10.1016/B978-0-12-391857-4.00009-4>.
- [42] N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I.F. Sbalzarini, Y. Gong, et al., *Nat. Methods* 11 (2014) 281–289, <http://dx.doi.org/10.1038/nmeth.2808>.
- [43] M.K. Cheezum, W.F. Walker, W.H. Guilford, *Biophys. J.* 81 (2001) 2378–2388.
- [44] A. Sakaue-Sawano, H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, et al., *Cell* 132 (2008) 487–498, <http://dx.doi.org/10.1016/j.cell.2007.12.033>.
- [45] A.E. Carpenter, T.R. Jones, M.R. Lamprecht, C. Clarke, I.H. Kang, O. Friman, et al., *Genome Biol.* 7 (2006) R100, <http://dx.doi.org/10.1186/gb-2006-7-10-r100>.
- [46] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, et al., *Nat. Methods* 9 (2012) 676–682, <http://dx.doi.org/10.1038/nmeth.2019>.
- [47] H. Shen, G. Nelson, D.E. Nelson, S. Kennedy, D.G. Spiller, T. Griffiths, et al., *J. R. Soc. Interface* 3 (2006) 787–794, <http://dx.doi.org/10.1098/rsif.2006.0137>.
- [48] A. Sacan, N. Ferhatosmanoglu, H. Coskun, *Bioinformatics* 24 (2008) 1647–1649, <http://dx.doi.org/10.1093/bioinformatics/btn247>.
- [49] F. Li, X. Zhou, J. Ma, S.T.C. Wong, *IEEE Trans. Med. Imaging* 29 (2010) 96–105, <http://dx.doi.org/10.1109/TMI.2009.2027813>.
- [50] A. Sergé, N. Bertaux, H. Rigneault, D. Marguet, *Nat. Methods* 5 (2008) 687–694, <http://dx.doi.org/10.1038/nmeth.1233>.
- [51] S.S. Rogers, T.A. Waigh, X. Zhao, J.R. Lu, *Phys. Biol.* 4 (2007) 220–227, <http://dx.doi.org/10.1088/1478-3975/4/3/008>.
- [52] J. Huth, M. Buchholz, J.M. Kraus, K. Møllhave, C. Gradinaru, G.V. Wichert, et al., *Comput. Methods Programs Biomed.* 104 (2011) 227–234, <http://dx.doi.org/10.1016/j.compbi.2011.06.002>.
- [53] L. Bosgraaf, P.J.M. Van Haastert, T. Bretschneider, *Cell Motil. Cytoskeleton* 66 (2009) 156–165, <http://dx.doi.org/10.1002/cm.20338>.
- [54] C.P. Bacher, M. Reichenzeller, C. Athale, H. Herrmann, R. Eils, *BMC Cell Biol.* 5 (2004) 45, <http://dx.doi.org/10.1186/1471-2121-5-45>.
- [55] S. Rajaram, B. Pavie, L.F. Wu, S.J. Altschuler, *Nat. Methods* 9 (2012) 635–637, <http://dx.doi.org/10.1038/nmeth.2097>.
- [56] D. Sage, M. Unser, P. Salmon, C. Dibner, *Cell Div.* 5 (2010), <http://dx.doi.org/10.1186/1747-1028-5-17>.
- [57] M. Linkert, C.T. Rueden, C. Allan, J.-M.M. Burel, W. Moore, A. Patterson, et al., *J. Cell Biol.* 189 (2010) 777–782, <http://dx.doi.org/10.1083/jcb.201004104>.
- [58] W. Pych, *Publ. Astron. Soc. Pac.* 116 (2004) 148–153, <http://dx.doi.org/10.1086/381786>.
- [59] S. Jaensch, M. Decker, A.A. Hyman, E.W. Myers, *Bioinforma. Oxf. Engl.* 26 (2010) i13–20, <http://dx.doi.org/10.1093/bioinformatics/btq190>.
- [60] A. Amer, E. Dubois, *IEEE Trans. Circuits Syst. Video Technol.* 15 (2005) 113–118, <http://dx.doi.org/10.1109/TCSVT.2004.837017>.
- [61] P. Paul, H. Duessmann, T. Bernas, H. Huber, D. Kalamatianos, *Comput. Med. Imaging Graph.* 34 (2010) 426–434, <http://dx.doi.org/10.1016/j.compmedimag.2010.04.001>.
- [62] F.M.M. Alzahrani, T. Chen, *Real-Time Imaging* 3 (1997) 363–378, <http://dx.doi.org/10.1006/rtim.1996.0071>.
- [63] H. Guo, *IEEE Signal Process. Mag.* 28 (2011) 134–137, <http://dx.doi.org/10.1109/MSP.2011.941846>.
- [64] R. Durbin, S. Eddy, A. Krogh, G. Mitchison, *Hidden Markov models*, in: *Biological Sequential Analysis*, C.U. Press, 1998, pp. 52–62.
- [65] C. Andrieu, N. De Freitas, A. Doucet, M.I. Jordan, *Mach. Learn.* 50 (2003) 5–43, <http://dx.doi.org/10.1023/A:1020281327116>.
- [66] T. Hirota, J.W. Lee, W.G. Lewis, E.E. Zhang, G. Breton, X. Liu, et al., *PLoS Biol.* 8 (2010), <http://dx.doi.org/10.1371/journal.pbio.1000559>.
- [67] A. Raj, A. van Oudenaarden, *Annu. Rev. Biophys.* 38 (2009) 255–270, <http://dx.doi.org/10.1146/annurev.biophys.37.032807.125928>.
- [68] D.R. Larson, *Curr. Opin. Genet. Dev.* 21 (2011) 591–599, <http://dx.doi.org/10.1016/j.gde.2011.07.010>.
- [69] D.M. Suter, N. Molina, F. Naef, U. Schibler, *Curr. Opin. Cell Biol.* 23 (2011) 657–662, <http://dx.doi.org/10.1016/j.cob.2011.09.004>.
- [70] J.M. Pedraza, J. Paulsson, *Science* 319 (2008) 339–343, <http://dx.doi.org/10.1126/science.1144331>.
- [71] A. Raj, C.S. Peskin, D. Tranchina, D.Y. Vargas, S. Tyagi, *PLoS Biol.* 4 (2006) 1707–1719, <http://dx.doi.org/10.1371/journal.pbio.0040309>.
- [72] C.V. Harper, B. Finkenstädt, D.J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, et al., *PLoS Biol.* 9 (2011), <http://dx.doi.org/10.1371/journal.pbio.1000607>.
- [73] L. Yin, M.A. Lazar, *Mol. Endocrinol. Baltim. Md.* 19 (2005) 1452–1459, <http://dx.doi.org/10.1210/me.2005-0057>.

A.3 Revisions

Here I list the critiques of the experts and explain how they were addressed in the revised manuscript.

1. Some figures lack statistical analysis and information on statistical significance. e.g. 2.29 and 2.30. These two figures should also be merged into one.

Mentioned figures were combined into Figure 3.41. In relevant parts of the thesis, statistical tests and single cell data were provided explicitly to support the claim (e.g. Figure 3.15, 3.17, 3.21).

2. Some figures are not cited in the text, while others needs a further description of the results and conclusions in the main text.

All figures were carefully checked, cited and explained in the text.

3. The entire text was scanned and reviewed carefully for the detection and the correction of typographical and grammatical errors. Referencing style to bibliography is now fixed across the thesis stating author and publication year for each reference (e.g., [Larson et al., 2009]).

4. Parts of the thesis that need modification were extended and supported with new figures whose explanations were provided here:

Section A (Appendix) was inserted to display the supplementary figures and data.

Section 1.2.7 presenting the state of the art in single-cell studies regarding TGF- β signaling was added. This section explains a number of studies about SMAD signaling dynamics, the single-cell heterogeneity of the pathway and its consequent responses. It also emphasizes the motivation of this thesis and the gap this study attempted to fill in transcriptional regulation and TGF- β signaling field.

Throughout the thesis, there are two versions of single cell signal traces. The first version is the min-max scaled (between 0 to 1) single cell quantification which requires no color scale (Figure 3.14 and 3.16). This version is sometimes used to have a better visualization of the response, to visualize the responding percentage for instance. However, absolute signal quantification is utilized when assessing the heterogeneity among individual cells, therefore, Figure 3.15 and 3.17 were inserted showing absolute signal quantification of SMAD and ctgf signals revealing the true heterogeneity in the response.

Paragraph 1.2.1 was extended to explain how the specificity of responses to different members of TGF- β superfamily of ligands is achieved. In addition, the crosstalk between canonical and non-canonical pathways was emphasized to indicate the plasticity of the pathway depending on combinations of ligands and receptor interactions. Though the focus in this work is the canonical SMAD pathway mediated TGF- β signaling, the literature described in this section emphasizes the complex regulation of SMAD signaling.

References about transcriptional control by SMAD signaling, missing before, were inserted in section 1.2.3.

The new figure 3.28 now displays the stimulation dose response of TGF- β SMAD4 signalling and the target gene *ctgf* response in single cells. It reveals single-cell heterogeneity and transitions in the response, from almost no response to saturation profile with the increasing doses of the ligand.

Figure 1.4 was inserted. This figure displays a comparative summary of existing methods to measure transcription and gene expression emphasizing the motivation to utilize single-cell dynamic measurements for our investigations.

Figure 3.7b was updated. Western blot experiments indicating the dose-response of the SMAD4 expression of the reporter as a function of doxycycline concentration were repeated which led to a cleaner observation compared to previous version.

Figure 3.3a was updated. Now, schematics illustrating the constructs used to generate cell lines display the complete details the constructs.

Section 1.2.5 about fibrosis is inserted. This section now explains previous studies describing the involvement of *ctgf* in cellular processes such as the wound healing and the pathological case, fibrosis. The section also presents both TGF- β -dependent and independent regulation of the target gene *ctgf* used in the study in different cell types and cellular contexts. Such a detailed investigation about *ctgf* is important for a few reasons: Firstly, use of NIH 3T3 fibroblast cells is shown to be a good choice due to previously shown TGF- β -dependent SMAD-mediated regulation of *ctgf* in those cells. Secondly, *ctgf* expression, shown to be minimal in the absence of stimulation in normal fibroblasts, appeared to have elevated levels of expression in the pathological case of fibrosis which may further motivate our investigations. Moreover, the section explains previously reported SMAD regulation of *ctgf* promoter and therefore why our single cell dynamic measurements would be helpful for therapeutic targeting of the TGF- β pathway, as discussed in section 4.5.

All discussion chapters now have the literature cited.

Section 4.2 in Discussion explains the contribution of YAP/TAZ pathway on SMAD mediated TGF pathway.

The previous problem of white space were sourced mostly from automatic page formattings from latex which was now corrected by using proper packages and also reviewing individual figures and their integration into the related text. Figure captions now have italic fonts to distinguish from the normal text.

5. Data on method development that are presently shown in the Methods section (and one data figure in the Discussion) should be at the beginning of the Results section.

Previously mislocated sections in Discussion part including development of the calibration cell line expressing Nluc-fluc fusion protein and initial implementation of simple microfluidics setup have now been transferred to Results part (Section 3.6).

Similarly, the figure, previously mislocated in Methods section, displaying the silhouette score of k-means clustering was transferred to Results section (Figure 3.19). Also, the order of the Methods and Results section is interchanged.

6. Text in the Results section needs to be heavily revised by adhering to a more conventional writing style: For every figure, the question asked should be stated at the beginning of a new paragraph (not at the end of the preceding paragraph), and each description of the data should be followed by a conclusion statement, defining clearly what is claimed based on this data.

Results section was updated based on the given recommendation about the flow of ideas for each paragraph. Updates and insertion of new sections and figures are discussed throughout this revision file.

7. Importantly, there is a clear style mismatch between the introduction, which is very clearly written, and the result section which is much less clear. It took me a lot of time to extract the essence of the result section. It would be good if this was clarified a bit.

Each section in Results were expanded, both figures and their explanation texts were modified.

8. Fig.2.17 legend is continuous with the text. This is confusing.

This is corrected. Figure captions now have italic fonts and they are separated from the normal text with certain spacing.

9. This paper contains trajectories of signaling or transcriptional activity of single cells, but often information about the data is missing. It would be nice to mention how many cells have been analyzed. It would be important to mention if the experiments that are shown have been performed multiple times, and if they are representative. Heatmaps would benefit from a scale to judge dynamic range of the signal. For many figures, population averages are shown without any information about stdev. A homogeneous figure style would enhance the presentation of the data.

Now, some figure captions are provided with cell numbers analyzed. Otherwise, heatmaps of those experiments indicated how many cells were used, on y-axis.

10. Some experiments that would capture experimental noise would help to understand the data. What is the fluctuation of the signal in conditions in which SMAD signaling is inhibited using drugs or genetic perturbations. This would enable to get an estimation of what is real signal or not.

Section 3.1.2 and 3.1.3 were expanded including complete performance characterization of the dual-reporter system. Figure 3.8 additionally covers control experiments including those with no ligand stimulation which reveals baseline signals and variability associated with those. In addition, Figure 3.12 displays the comparative profiles of stimulation and no stimulation conditions giving a sense of signal-to-noise performance of

both SMAD and ctgf reporters. Furthermore, updated Figure 3.11, including immunofluorescence and western blot analysis, revealed no reporter-dependent perturbation of the endogenous system but agreement of reporter and endogenous dynamics and levels. Finally, the performance of translocation reporters is evaluated with signal-to-noise ratio assessment (Figure 3.13).

11. Especially pertinent to the last point, are the response peaks observed detected by automated peak detection meaningful in Fig 3.38 and 3.39 biologically relevant ? or do they arise from the "bumpiness" of the tracks ? I did not understand what information is conveyed in these traces.

In Figures 3.38 and 3.39, heatmaps reveal general behaviour of traces; however, in heatmaps, one can not resolve, by eye, the exact peak points for each single-cell. Therefore, automated peaks detected with certain thresholds were provided to guide the reader. As noticed by the reviewer, both SMAD and ctgf traces have some other peak points in the long-term since the peak detection criteria may be satisfied also by those smaller bumps. However, the relevant peaks for our investigation are those with certain amplitudes occurring at typical response timescales.

12. Fig.3.20: decomposition of different cell trajectories by clustering. Transient versus sustained trajectories are identified. Sustained trajectories seems to be oscillating not in phase. Yet a population average of the latter cellular state is shown ? Is this population average representative ?

Fig.3.20 displays the existence of two different behaviours. I agree with the comment that the sustained portion can also be interpreted as oscillating while the stability of those oscillations in longer terms is not clear. Such behaviour can be interpreted as due to interplay between negative feedbacks and the activation [Warmflash et al., 2012, Zi et al., 2011] or secondary mechanisms of SMADs such as recruiting activators and suppressors shaping the transcriptional response further at later time points[Coda et al., 2017]. Heat maps regarding two clusters now display both versions of the signal, absolute and normalized, min-max scaled (between 0 to 1). Yet, population averages provided are based on the quantification of the absolute signals.

Bibliography

- D. J. Abraham, X. Shiwen, C. M. Black, S. Sa, Y. Xu, and A. Leask. Tumor Necrosis Factor α Suppresses the Induction of Connective Tissue Growth Factor by Transforming Growth Factor- β in Normal and Scleroderma Fibroblasts. *Journal of Biological Chemistry*, 2000.
- J. G. Abreu, N. I. Ketpura, B. Reversade, and E. M. De Robertis. Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF- β . *Nature Cell Biology*, 2002.
- K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nature Reviews Genetics*, 2012.
- F. H. Andrews, B. D. Strahl, and T. G. Kutateladze. Insights into newly discovered marks and readers of epigenetic information. *Nature Chemical Biology*, 2016.
- J. P. Annes, J. S. Munger, and D. B. Rifkin. Making sense of latent TGF β activation. *Journal of Cell Science*, 2003.
- E. Aragón, N. Goerner, Q. Xi, T. Gomes, S. Gao, J. Massagué, and M. J. Macias. Structural Basis for the Versatile Interactions of Smad7 with Regulator WW Domains in TGF- β Pathways. *Structure*, 2012.
- D. Arthur and S. Vassilvitskii. K-means++: The advantages of careful seeding. In *In Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007.
- L. Attisano and J. L. Wrana. Signal Transduction by the TGF- β Superfamily. *Science*, 2002.
- D. Aymoz, V. Wosika, E. Durandau, and S. Pelet. Real-time quantification of protein expression at the single-cell level via dynamic protein synthesis translocation reporters. *Nature Communications*, 2016.
- A. M. Azzaz, M. W. Vitalini, A. S. Thomas, J. P. Price, M. J. Blacketer, D. E. Cryderman, L. N. Zirbel, C. L. Woodcock, A. H. Elcock, L. L. Wallrath, and M. A. Shogren-Knaak. Human Heterochromatin Protein 1 Promotes Nucleosome Associations That Drive Chromatin Condensation. *Journal of Biological Chemistry*, 2014.
- N. BabuRajendran, P. Palasingam, K. Narasimhan, W. Sun, S. Prabhakar, R. Jauch, and P. R. Kolatkar. Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF- β effectors. *Nucleic Acids Research*, 2010.

Bibliography

- K. Bahar Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg, and S. Itzkovitz. Bursty gene expression in the intact mammalian liver. *Molecular Cell*, 2015.
- N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial Persistence as a Phenotypic Switch. *Science*, 2004.
- C. R. Bartman, S. C. Hsu, C. C. S. Hsiung, A. Raj, and G. A. Blobel. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell*, 2016.
- N. Battich, T. Stoeger, and L. Pelkmans. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nature Methods*, 2013.
- N. Battich, T. Stoeger, and L. Pelkmans. Control of Transcript Variability in Single Mammalian Cells. *Cell*, 2015.
- J. Batut, B. Schmierer, J. Cao, L. A. Raftery, C. S. Hill, and M. Howell. Two highly related regulatory subunits of PP2A exert opposite effects on TGF- β /Activin/Nodal signalling. *Development*, 2008.
- J. M. Bean, E. D. Siggia, and F. R. Cross. Coherence and Timing of Cell Cycle Start Examined at Single-Cell Resolution. *Molecular Cell*, 2006.
- M. Bengtsson, A. Ståhlberg, P. Rorsman, and M. Kubista. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research*, 2005.
- D. Bennett and L. Alphey. PP1 binds Sara and negatively regulates Dpp signaling in *Drosophila melanogaster*. *Nature Genetics*, 2002.
- E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long. Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell*, 1998.
- J. Bieler, R. Cannavo, K. Gustafson, C. Gobet, D. Gatfield, and F. Naef. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular Systems Biology*, 2014.
- W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins. Noise in eukaryotic gene expression. *Nature*, 2003.
- S. Blanchoud, D. Nicolas, B. Zoller, O. Tidin, and F. Naef. CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings. *Methods*, 2015.
- G. C. Blobe, W. P. Schieman, and H. F. Lodish. Role of Transforming Growth Factor- β in Human Disease. *New England Journal of Medicine*, 2000.

- M. D. Boeck and P. T. Dijke. Key role for ubiquitin protein modification in TGF β signal transduction. *Upsala Journal of Medical Sciences*, 2012.
- S. Boireau, P. Maiuri, E. Basyuk, M. de la Mata, A. Knezevich, B. Pradet-Balade, V. Bäcker, A. Kornblihtt, A. Marcello, and E. Bertrand. The transcriptional cycle of HIV-1 in real-time and live cells. *The Journal of Cell Biology*, 2007.
- J. P. Bothma, H. G. Garcia, E. Esposito, G. Schlissel, T. Gregor, and M. Levine. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living *Drosophila* embryos. *Proceedings of the National Academy of Sciences*, 2014.
- E. P. Böttinger, J. J. Letterio, and A. B. Roberts. Biology of TGF- β in knockout and transgenic mouse models. *Kidney International*, 1997.
- S. Brabletz and T. Brabletz. The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer. *EMBO reports*, 2010.
- Y. Brody, N. Neufeld, N. Bieberstein, S. Z. Causse, E.-M. Böhnlein, K. M. Neugebauer, X. Darzacq, and Y. Shav-Tal. The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing. *PLOS Biology*, 2011.
- L. Bronstein, C. Zechner, and H. Koeppl. Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population. *Methods*, 2015.
- C. R. Brown and H. Boeger. Nucleosomal promoter variation generates gene expression noise. *Proceedings of the National Academy of Sciences*, 2014.
- C. R. Brown, C. Mao, E. Falkovskaia, M. S. Jurica, and H. Boeger. Linking Stochastic Fluctuations in Chromatin Structure and Gene Expression. *PLOS Biology*, 2013.
- A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 2006.
- M. J. Carrozza, B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W.-J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*, 2005.
- S. S.-K. Chan and M. Kyba. What is a Master Regulator? *Journal of stem cell research & therapy*, 2013.
- B. Chaqour and M. Goppelt-Struebe. Mechanical regulation of the Cyr61/CCN1 and CTGF/CCN2 proteins. *The FEBS Journal*, 2006.
- K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 2015.

Bibliography

- X. Chen, M. J. Rubock, and M. Whitman. A transcriptional partner for MAD proteins in TGF- β signalling. *Nature*, 1996.
- Y. Chen, I. E. Blom, S. Sa, R. Goldschmeding, D. J. Abraham, and A. Leask. CTGF expression in mesangial cells: Involvement of SMADs, MAP kinase, and PKC. *Kidney International*, 2002.
- Y.-G. Chen. Endocytic regulation of TGF- β signaling. *Cell Research*, 2009.
- G. S. Chin, W. Liu, Z. Peled, T. Y. Lee, D. S. Steinbrech, M. Hsu, and M. T. Longaker. Differential expression of transforming growth factor-beta receptors I and II and activation of Smad 3 in keloid fibroblasts. *Plastic and reconstructive surgery*, 2001.
- S. Chong, C. Chen, H. Ge, and X. S. Xie. Mechanism of transcriptional bursting in bacteria. *Cell*, 2014.
- J. R. Chubb, T. Trcek, S. M. Shenoy, and R. H. Singer. Transcriptional pulsing of a developmental gene. *Current biology*, 2006.
- I. Cicha and M. Goppelt-Struebe. Connective tissue growth factor: Context-dependent functions and mechanisms of regulation. *BioFactors*, 2009.
- I. Cicha, K. Beronov, E. L. Ramirez, K. Osterode, M. Goppelt-Struebe, D. Raaz, A. Yilmaz, W. G. Daniel, and C. D. Garlich. Shear stress preconditioning modulates endothelial susceptibility to circulating TNF- α and monocytic cell recruitment in a simplified model of arterial bifurcations. *Atherosclerosis*, 2009.
- D. C. Clarke, M. D. Betterton, and X. Liu. Systems theory of Smad signalling. *Systems Biology*, 2006.
- D. M. Coda, T. Gaarenstroom, P. East, H. Patel, D. S. J. Miller, A. Lobley, N. Matthews, A. Stewart, and C. S. Hill. Distinct modes of SMAD2 chromatin binding and remodeling shape the transcriptional response to NODAL/Activin signaling. *eLife*, 2017.
- D. B. Constam. Regulation of TGF β and related signals by precursor processing. *Seminars in Cell & Developmental Biology*, 2014.
- M. F. Cordeiro, A. Mead, R. R. Ali, R. A. Alexander, S. Murray, C. Chen, C. York-Defalco, N. M. Dean, G. S. Schultz, and P. T. Khaw. Novel antisense oligonucleotides targeting TGF- β inhibit in vivo scarring and improve surgical outcome. *Gene Therapy*, 2003.
- M. P. Creighton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 2010.
- F. Crick. Central Dogma of Molecular Biology. *Nature*, 1970.

- E. Dacheux, N. Malys, X. Meng, V. Ramachandran, P. Mendes, and J. E. McCarthy. Translation initiation events on structured eukaryotic mRNAs generate gene expression noise. *Nucleic Acids Research*, 2017.
- R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 2012.
- R. D. Dar, S. M. Shaffer, A. Singh, B. S. Razooky, M. L. Simpson, A. Raj, and L. S. Weinberger. Transcriptional Bursting Explains the Noise–Versus–Mean Relationship in mRNA and Protein Levels. *PLOS ONE*, 2016.
- X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer. *In Vivo* dynamics of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, 2007.
- J. Deheuninck and K. Luo. Ski and SnoN, potent negative regulators of TGF- β signaling. *Cell Research*, 2009.
- C. Deluz, E. T. Friman, D. Streibinger, A. Benke, M. Raccaud, A. Callegari, M. Leleu, S. Manley, and D. M. Suter. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes & Development*, 2016.
- S. Dennler, S. Itoh, D. Vivien, P. ten Dijke, S. Huet, and J.-M. Gauthier. Direct binding of Smad3 and Smad4 to critical TGF β -inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. *The EMBO Journal*, 1998.
- R. Derynck and Y. E. Zhang. Smad-dependent and Smad-independent pathways in TGF- β family signalling. *Nature*, 2003.
- S. S. Dey, J. E. Foley, P. Limsirichai, D. V. Schaffer, and A. P. Arkin. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Molecular Systems Biology*, 2015.
- L. Di Croce and K. Helin. Transcriptional regulation by Polycomb group proteins. *Nature Structural & Molecular Biology*, 2013.
- G. M. Di Guglielmo, C. Le Roy, A. F. Goodfellow, and J. L. Wrana. Distinct endocytic pathways regulate TGF- β receptor signalling and turnover. *Nature Cell Biology*, 2003.
- S. Dupont, A. Mamidi, M. Cordenonsi, M. Montagner, L. Zacchigna, M. Adorno, G. Martello, M. J. Stinchfield, S. Soligo, L. Morsut, M. Inui, S. Moro, N. Modena, F. Argenton, S. J. Newfeld, and S. Piccolo. FAM/USP9x, a Deubiquitinating Enzyme Essential for TGF β Signaling, Controls Smad4 Monoubiquitination. *Cell*, 2009.
- S. Dupont, L. Morsut, M. Aragona, E. Enzo, S. Giulitti, M. Cordenonsi, F. Zanconato, J. L. Digabel, M. Forcato, S. Bicciato, N. Elvassore, and S. Piccolo. Role of YAP/TAZ in mechanotransduction. *Nature*, 2011.

Bibliography

- T. Ebisawa, M. Fukuchi, G. Murakami, T. Chiba, K. Tanaka, T. Imamura, and K. Miyazono. Smurf1 Interacts with Transforming Growth Factor- β Type I Receptor through Smad7 and Induces Receptor Degradation. *Journal of Biological Chemistry*, 2001.
- M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic Gene Expression in a Single Cell. *Science*, 2002.
- C.-H. L. Eng, S. Shah, J. Thomassie, and L. Cai. Profiling the transcriptome with RNA SPOTs. *Nature Methods*, 2017.
- C. G. England, E. B. Ehlerding, and W. Cai. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjugate Chemistry*, 2016.
- D. Ezer, V. Moignard, B. Göttgens, and B. Adryan. Determining Physical Mechanisms of Gene Expression Regulation from Single Cell Gene Expression Data. *PLOS Computational Biology*, 2016.
- A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer. Visualization of Single RNA Transcripts in Situ. *Science*, 1998.
- X.-H. Feng and R. Derynck. Specificity and Versatility in TGF- β Signaling Through Smads. *Annual Review of Cell and Developmental Biology*, 2005.
- X.-H. Feng, Y. Zhang, R.-Y. Wu, and R. Derynck. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for Smad3 in TGF- β -induced transcriptional activation. *Genes & Development*, 1998.
- Z. Feng, Z. Zi, and X. Liu. Measuring TGF- β Ligand Dynamics in Culture Medium. In *TGF- β Signaling*, Methods in Molecular Biology. Humana Press, New York, NY, 2016. ISBN 978-1-4939-2965-8 978-1-4939-2966-5.
- G. J. Fillion, J. G. van Bommel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. van Steensel. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell*, 2010.
- A. Filipczyk, C. Marr, S. Hastreiter, J. Feigelman, M. Schwarzfischer, P. S. Hoppe, D. Loeffler, K. D. Kokkaliaris, M. Endeke, B. Schauburger, O. Hilsenbeck, S. Skylaki, J. Hasenauer, K. Anastasiadis, F. J. Theis, and T. Schroeder. Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nature Cell Biology*, 2015.
- T. Fowler, R. Sen, and A. L. Roy. Regulation of Primary Response Genes. *Molecular Cell*, 2011.
- C. L. Frick, C. Yarka, H. Nunns, and L. Goentoro. Sensing relative signal in the Tgf- β /Smad pathway. *Proceedings of the National Academy of Sciences*, 2017.
- G. Friedrich and P. Soriano. Promoter traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. *Genes & Development*, 1991.

- N. J. Fuda, M. B. Ardehali, and J. T. Lis. Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature*, 2009.
- T. Fukaya, B. Lim, and M. Levine. Enhancer Control of Transcriptional Bursting. *Cell*, 2016.
- G. Gabbiani. The myofibroblast in wound healing and fibrocontractive diseases. *The Journal of Pathology*, 2003.
- Z.-M. Geng, J.-B. Zheng, X.-X. Zhang, J. Tao, and L. Wang. Role of transforming growth factor-beta signaling pathway in pathogenesis of benign biliary stricture. *World Journal of Gastroenterology*, 2008.
- S. Germain, M. Howell, G. M. Esslemont, and C. S. Hill. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. *Genes & Development*, 2000.
- A. K. Ghosh, W. Yuan, Y. Mori, S.-j. Chen, and J. Varga. Antagonistic Regulation of Type I Collagen Gene Expression by Interferon- γ and Transforming Growth Factor- β integration at the level of p300/CBP transcriptional coactivators. *Journal of Biological Chemistry*, 2001.
- I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 2005.
- R. R. Gomis, C. Alarcón, W. He, Q. Wang, J. Seoane, A. Lash, and J. Massagué. A FoxO–Smad synexpression group in human keratinocytes. *Proceedings of the National Academy of Sciences*, 2006.
- M. Gossen and H. Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, 1992.
- M.-J. Goumans, G. Valdimarsdottir, S. Itoh, F. Lebrin, J. Larsson, C. Mummery, S. Karlsson, and P. ten Dijke. Activin Receptor-like Kinase (ALK)1 Is an Antagonistic Mediator of Lateral TGF β /ALK5 Signaling. *Molecular Cell*, 2003.
- E. Grönroos, I. J. Kingston, A. Ramachandran, R. A. Randall, P. Vizán, and C. S. Hill. Transforming Growth Factor β Inhibits Bone Morphogenetic Protein-Induced Transcription through Novel Phosphorylated Smad1/5-Smad3 Complexes. *Molecular and Cellular Biology*, 2012.
- G. R. Grotendorst. Connective tissue growth factor: A mediator of TGF- β action on fibroblasts. *Cytokine & Growth Factor Reviews*, 1997.
- G. R. Grotendorst, H. Okochi, and N. Hayashi. A novel transforming growth factor beta response element controls the expression of the connective tissue growth factor gene. *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research*, 1996.

Bibliography

- S. Grünberg, L. Warfield, and S. Hahn. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Structural & Molecular Biology*, 2012.
- A. Hahn, J. Heusinger-Ribeiro, T. Lanz, S. Zenkel, and M. Goppelt-Struebe. Induction of connective tissue growth factor by activation of heptahelical receptors. *Journal of Biological Chemistry*, 2000.
- M. P. Hall, J. Unch, B. F. Binkowski, M. P. Valley, B. L. Butler, M. G. Wood, P. Otto, K. Zimmerman, G. Vidugiris, T. Machleidt, M. B. Robers, H. A. Benink, C. T. Eggers, M. R. Slater, P. L. Meisenheimer, D. H. Klaubert, F. Fan, L. P. Encell, and K. V. Wood. Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *ACS Chemical Biology*, 2012.
- C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, D. G. Spiller, J. J. Mullins, D. A. Rand, J. R. E. Davis, and M. R. H. White. Dynamic Analysis of Stochastic Transcription Cycles. *PLoS Biol*, 2011.
- S. Hayes, A. Chawla, and S. Corvera. TGF β receptor internalization into EEA1-enriched early endosomes: Role in signaling to Smad2. *The Journal of Cell Biology*, 2002.
- C.-H. Heldin and A. Moustakas. Signaling Receptors for TGF- β Family Members. *Cold Spring Harbor Perspectives in Biology*, 2016.
- A. G. Hill. Initiators and Propagators of the Metabolic Response to Injury. *World Journal of Surgery*, 2000.
- C. S. Hill. Nucleocytoplasmic shuttling of Smad proteins. *Cell Research*, 2009.
- K. Hiragami-Hamada, S. Soeroes, M. Nikolov, B. Wilkins, S. Kreuz, C. Chen, I. A. De La Rosa-Velázquez, H. M. Zenn, N. Kost, W. Pohl, A. Chernev, D. Schwarzer, T. Jenuwein, M. Lorincz, B. Zimmermann, P. J. Walla, H. Neumann, and T. a. Baubec. Dynamic and flexible H3K9me3 bridging via HP1 dimerization establishes a plastic state of condensed chromatin. *Nature Communications*, 2016.
- J. W. K. Ho, Y. L. Jung, T. Liu, B. H. Alver, S. Lee, K. Ikegami, K.-A. Sohn, A. Minoda, M. Y. Tolstorukov, A. Appert, S. C. J. Parker, T. Gu, A. Kundaje, N. C. Riddle, E. Bishop, T. A. Egelhofer, S. S. Hu, A. A. Alekseyenko, A. Rechtsteiner, D. Asker, J. A. Belsky, S. K. Bowman, Q. B. Chen, R. A.-J. Chen, D. S. Day, Y. Dong, A. C. Dose, X. Duan, C. B. Epstein, S. Ercan, E. A. Feingold, F. Ferrari, J. M. Garrigues, N. Gehlenborg, P. J. Good, P. Haseley, D. He, M. Herrmann, M. M. Hoffman, T. E. Jeffers, P. V. Kharchenko, P. Kolasinska-Zwierz, C. V. Kotwaliwale, N. Kumar, S. A. Langley, E. N. Larschan, I. Latorre, M. W. Libbrecht, X. Lin, R. Park, M. J. Pazin, H. N. Pham, A. Plachetka, B. Qin, Y. B. Schwartz, N. Shores, P. Stempor, A. Vielle, C. Wang, C. M. Whittle, H. Xue, R. E. Kingston, J. H. Kim, B. E. Bernstein, A. F. Dernburg, V. Pirrotta, M. I. Kuroda, W. S. Noble, T. D. Tullius, M. Kellis, D. M. MacAlpine, S. Strome, S. C. R. Elgin, X. S. Liu, J. D. Lieb, J. Ahringer, G. H. Karpen, and P. J. Park. Comparative analysis of metazoan chromatin organization. *Nature*, 2014.

- S. Hocine, P. Raymond, D. Zenklusen, J. A. Chao, and R. H. Singer. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nature Methods*, 2013.
- A. Holmes, D. J. Abraham, S. Sa, X. Shiwen, C. M. Black, and A. Leask. CTGF and SMADs, Maintenance of Scleroderma Phenotype Is Independent of SMAD Signaling. *Journal of Biological Chemistry*, 2001.
- G. Hornung, R. Bar-Ziv, D. Rosin, N. Tokuriki, D. S. Tawfik, M. Oren, and N. Barkai. Noise–mean relationship in mutated promoters. *Genome Research*, 2012.
- S. Itoh and P. ten Dijke. Negative regulation of TGF- β receptor/Smad signal transduction. *Current opinion in cell biology*, 2007.
- S. Itzkovitz, A. Lyubimova, I. C. Blat, M. Maynard, J. van Es, J. Lees, T. Jacks, H. Clevers, and A. van Oudenaarden. Single-molecule transcript counting of stem-cell markers in the mouse intestine. *Nature Cell Biology*, 2011.
- S. Iyer-Biswas, F. Hayot, and C. Jayaprakash. Stochasticity of gene products from transcriptional pulsing. *Physical Review E*, 2009.
- L. M. Jakt, S. Moriwaki, and S. Nishikawa. A continuum of transcriptional identities visualized by combinatorial fluorescent in situ hybridization. *Development*, 2013.
- S. K. Jang and E. Wimmer. Cap-independent translation of encephalomyocarditis virus RNA: Structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes & Development*, 1990.
- R. Janknecht, N. J. Wells, and T. Hunter. TGF- β -stimulated cooperation of Smad proteins with the coactivators CBP/p300. *Genes & Development*, 1998.
- K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser. Robust single-particle tracking in live-cell time-lapse sequences. *Nature methods*, 2008.
- T. Juven-Gershon, J.-Y. Hsu, J. W. Theisen, and J. T. Kadonaga. The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology*, 2008.
- P. Kafri, S. E. Hasenson, I. Kanter, J. Sheinberger, N. Kinor, S. Yunger, and Y. Shav-Tal. Quantifying β -catenin subcellular dynamics and cyclin D1 mRNA transcription during Wnt signaling in single living cells. *eLife*, 2013.
- A. Kalo, I. Kanter, A. Shraga, J. Sheinberger, H. Tzemach, N. Kinor, R. H. Singer, T. Lionnet, and Y. Shav-Tal. Cellular Levels of Signaling Factors Are Sensed by β -actin Alleles to Modulate Transcriptional Pulse Intensity. *Cell reports*, 2015.
- C. J. M. Kane, P. A. Hebda, J. N. Mansbridge, and P. C. Hanawalt. Direct evidence for spatial and temporal regulation of transforming growth factor *b1* expression during cutaneous wound healing. *Journal of Cellular Physiology*, 1991.

Bibliography

- J. S. Kang, T. Alliston, R. Delston, and R. Derynck. Repression of Runx2 function by TGF- β through recruitment of class II histone deacetylases by Smad3. *The EMBO Journal*, 2005a.
- Y. Kang, P. M. Siegel, W. Shu, M. Drobnjak, S. M. Kakonen, C. Cordón-Cardo, T. A. Guise, and J. Massagué. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell*, 2003.
- Y. Kang, W. He, S. Tulley, G. P. Gupta, I. Serganova, C.-R. Chen, K. Manova-Todorova, R. Blasberg, W. L. Gerald, and J. Massagué. Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway. *Proceedings of the National Academy of Sciences*, 2005b.
- P. Kavsak, R. K. Rasmussen, C. G. Causing, S. Bonni, H. Zhu, G. H. Thomsen, and J. L. Wrana. Smad7 Binds to Smurf2 to Form an E3 Ubiquitin Ligase that Targets the TGF β Receptor for Degradation. *Molecular Cell*, 2000.
- R. A. Kellogg, C. Tian, M. Etzrodt, and S. Tay. Cellular Decision Making by Non-Integrative Processing of TLR Inputs. *Cell Reports*, 2017.
- A. Khmelinskii, P. J. Keller, A. Bartosik, M. Meurer, J. D. Barry, B. R. Mardin, A. Kaufmann, S. Trautmann, M. Wachsmuth, G. Pereira, W. Huber, E. Schiebel, and M. Knop. Tandem fluorescent protein timers for *in vivo* analysis of protein dynamics. *Nature Biotechnology*, 2012.
- D. Koinuma, M. Shinozaki, A. Komuro, K. Goto, M. Saitoh, A. Hanyu, M. Ebina, T. Nukiwa, K. Miyazawa, T. Imamura, and K. Miyazono. Arkadia amplifies TGF- superfamily signalling through degradation of Smad7. *The EMBO Journal*, 2003.
- S. Kondo, S. Kubota, Y. Mukudai, N. Moritani, T. Nishida, H. Matsushita, S. Matsumoto, T. Sugahara, and M. Takigawa. Hypoxic regulation of stability of connective tissue growth factor/CCN2 mRNA by 3'-untranslated region interacting with a cellular protein in human chondrosarcoma cells. *Oncogene*, 2006.
- Z. Konkoli. Exact equilibrium-state solution of an intracellular complex formation model: $kA \rightleftharpoons P$ reaction in a small volume. *Physical Review E*, 2010.
- T. Kouzarides. Chromatin Modifications and Their Function. *Cell*, 2007.
- J. N. Kuehner, E. L. Pearson, and C. Moore. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*, 2011.
- G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz, and U. Alon. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nature Genetics*, 2004.
- D. R. Larson. What do expression dynamics tell us about the mechanism of transcription? *Current Opinion in Genetics & Development*, 2011.
- D. R. Larson, R. H. Singer, and D. Zenklusen. A single molecule view of gene expression. *Trends in cell biology*, 2009.

- D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science (New York, N.Y.)*, 2011.
- D. R. Larson, C. Fritsch, L. Sun, X. Meng, D. S. Lawrence, and R. H. Singer. Direct observation of frequency modulated transcription in single cells using light activation. *eLife*, 2013.
- A. Laugesen and K. Helin. Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell*, 2014.
- C. Le Roy and J. L. Wrana. Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. *Nature Reviews Molecular Cell Biology*, 2005.
- A. Leask and D. J. Abraham. TGF- β signaling and the fibrotic response. *The FASEB Journal*, 2004.
- A. Leask, S. Sa, A. Holmes, X. Shiwen, C. M. Black, and D. J. Abraham. The control of *ccn2* (ctgf) gene expression in normal and scleroderma fibroblasts. *Molecular Pathology*, 2001.
- A. Leask, C. P. Denton, and D. J. Abraham. Insights Into the Molecular Mechanism of Chronic Fibrosis: The Role of Connective Tissue Growth Factor in Scleroderma. *Journal of Investigative Dermatology*, 2004.
- E. Lécuyer, N. Parthasarathy, and H. M. Krause. Fluorescent in situ hybridization protocols in *Drosophila* embryos and tissues. *Methods in Molecular Biology (Clifton, N.J.)*, 2008.
- T. L. Lenstra, J. Rodriguez, H. Chen, and D. R. Larson. Transcription Dynamics in Living Cells. *Annual Review of Biophysics*, 2016.
- I. Lestas, G. Vinnicombe, and J. Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 2010.
- M. J. Levesque and A. Raj. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nature Methods*, 2013.
- J. M. Levsky and R. H. Singer. Fluorescence in situ hybridization: Past, present and future. *Journal of cell science*, 2003.
- J. M. Levsky, S. M. Shenoy, R. C. Pezo, and R. H. Singer. Single-Cell Gene Expression Profiling. *Science*, 2002.
- L. Levy, M. Howell, D. Das, S. Harkin, V. Episkopou, and C. S. Hill. Arkadia Activates Smad3/Smad4-Dependent Transcription by Triggering Signal-Induced SnoN Degradation. *Molecular and Cellular Biology*, 2007.
- X. Lin, M. Liang, and X.-H. Feng. Smurf2 Is a Ubiquitin E3 Ligase Mediating Proteasome-dependent Degradation of Smad2 in Transforming Growth Factor- β Signaling. *Journal of Biological Chemistry*, 2000.

Bibliography

- X. Lin, X. Duan, Y.-Y. Liang, Y. Su, K. H. Wrighton, J. Long, M. Hu, C. M. Davis, J. Wang, F. C. Brunicardi, Y. Shi, Y.-G. Chen, A. Meng, and X.-H. Feng. PPM1A functions as a Smad phosphatase to terminate TGF β signaling. *Cell*, 2006.
- T. Lionnet and R. H. Singer. Transcription goes digital. *EMBO reports*, 2012.
- T. Lionnet, K. Czaplinski, X. Darzacq, Y. Shav-Tal, A. L. Wells, J. A. Chao, H. Y. Park, V. de Turris, M. Lopez-Jones, and R. H. Singer. A transgenic mouse for *in vivo* detection of endogenous labeled mRNA. *Nature Methods*, 2011.
- X. Liu, Y. Sun, R. A. Weinberg, and H. F. Lodish. Ski/Sno and TGF- β signaling. *Cytokine & Growth Factor Reviews*, 2001.
- S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 1982.
- R. S. Lo and J. Massagué. Ubiquitin-dependent degradation of TGF- β -activated Smad2. *Nature Cell Biology*, 1999.
- A. Loewer and G. Lahav. We are all individuals: Causes and consequences of non-genetic heterogeneity in mammalian cells. *Current Opinion in Genetics & Development*, 2011.
- P. Lönn, A. Morén, E. Raja, M. Dahl, and A. Moustakas. Regulating the stability of TGF β receptors and Smads. *Cell Research*, 2009.
- E. Lubeck and L. Cai. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods*, 2012.
- J. Ma and M. D. Wang. DNA supercoiling during transcription. *Biophysical Reviews*, pages 1–13, 2016.
- M. J. Macias, P. Martin-Malpartida, and J. Massagué. Structural determinants of Smad function in TGF- β signaling. *Trends in Biochemical Sciences*, 2015.
- A. Mandic, D. Strebinger, C. Regali, N. E. Phillips, and D. M. Suter. A novel method for quantitative measurements of gene expression in single living cells. *Methods*, 2017.
- C. Mao, C. R. Brown, E. Falkovskaia, S. Dong, E. Hrabeta-Robinson, L. Wenger, and H. Boeger. Quantitative analysis of the transcription control mechanism. *Molecular Systems Biology*, 2010.
- R. M. Martin, J. Rino, C. Carvalho, T. Kirchhausen, and M. Carmo-Fonseca. Live-Cell Visualization of Pre-mRNA Splicing with Single-Molecule Sensitivity. *Cell Reports*, 2013.
- J. Massagué. TGF- β Signal Transduction. *Annual Review of Biochemistry*, 1998.
- J. Massagué. TGF β signalling in context. *Nature Reviews Molecular Cell Biology*, 2012.
- J. Massagué, J. Seoane, and D. Wotton. Smad transcription factors. *Genes & Development*, 2005.

- J. Massagué. TGF in Cancer. *Cell*, 2008.
- Massagué Joan and Xi Qiaoran. TGF- control of stem cell differentiation genes. *FEBS Letters*, 2012.
- K. J. Mavrikakis, R. L. Andrew, K. L. Lee, C. Petropoulou, J. E. Dixon, N. Navaratnam, D. P. Norris, and V. Episkopou. Arkadia Enhances Nodal/TGF- β Signaling by Coupling Phospho-Smad2/3 Activity and Turnover. *PLOS Biology*, 2007.
- A. Mazo-Vargas, H. Park, M. Aydin, N. E. Buchler, K. Weis, and J. Lippincott-Schwartz. Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy. *Molecular Biology of the Cell*, 2014.
- A. McDavid, L. Dennis, P. Danaher, G. Finak, M. Krouse, A. Wang, P. Webster, J. Beechem, and R. Gottardo. Modeling Bi-modality Improves Characterization of Cell Cycle on Gene Expression in Single Cells. *PLOS Computational Biology*, (7):e1003696, 2014.
- C. Meyer, P. Godoy, A. Bachmann, Y. Liu, D. Barzan, I. Ilkavets, P. Maier, C. Herskind, J. G. Hengstler, and S. Dooley. Distinct role of endocytosis for Smad and non-Smad TGF- β signaling regulation in hepatocytes. *Journal of Hepatology*, 2011.
- A. Miyawaki, T. Nagai, and H. Mizuno. Mechanisms of protein fluorophore formation and engineering. *Current Opinion in Chemical Biology*, 2003.
- J. R. Moffitt and X. Zhuang. Chapter One - RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). In G. S. Filonov and S. R. Jaffrey, editors, *Methods in Enzymology*, Visualizing RNA Dynamics in the Cell. Academic Press, 2016.
- N. Molina, D. M. Suter, R. Cannavo, B. Zoller, I. Gotic, and F. Naef. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proceedings of the National Academy of Sciences of the United States of America*, 2013.
- N. Moullan, L. Mouchiroud, X. Wang, D. Ryu, E. G. Williams, A. Mottis, V. Jovaisaite, M. V. Frochaux, P. M. Quiros, B. Deplancke, R. H. Houtkooper, and J. Auwerx. Tetracyclines Disturb Mitochondrial Function across Eukaryotic Models: A Call for Caution in Biomedical Research. *Cell Reports*, 2015.
- E. E.-D. A. Moussad and D. R. Brigstock. Connective Tissue Growth Factor: What's in a Name? *Molecular Genetics and Metabolism*, 2000.
- A. Moustakas and C.-H. Heldin. The regulation of TGF β signal transduction. *Development*, 2009.
- S. Muehlich, I. Cicha, C. D. Garlich, B. Krueger, G. Posern, and M. Goppelt-Struebe. Actin-dependent regulation of connective tissue growth factor. *American Journal of Physiology-Cell Physiology*, 2007.

Bibliography

- T. Muramoto, I. Müller, G. Thomas, A. Melvin, and J. R. Chubb. Methylation of H3K4 Is Required for Inheritance of Active Transcriptional States. *Current Biology*, 2010.
- G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden. Systematic identification of signal-activated stochastic gene regulation. *Science*, 2013.
- J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 2006.
- R. H. Newman, M. D. Fosbrink, and J. Zhang. Genetically Encodable Fluorescent Biosensors for Tracking Signaling Dynamics in Living Cells. *Chemical Reviews*, 2011.
- D. Nicolas, N. E. Phillips, and F. Naef. What shapes eukaryotic transcriptional bursting? *Molecular BioSystems*, 2017.
- D. Nicolas, B. Zoller, D. M. Suter, and F. Naef. Modulation of transcriptional burst frequency by histone acetylation. *Proceedings of the National Academy of Sciences*, 2018.
- F. J. Nicolás, K. D. Bosscher, B. Schmierer, and C. S. Hill. Analysis of Smad nucleocytoplasmic shuttling in living cells. *Journal of Cell Science*, 2004.
- H. Nishida, T. Suzuki, S. Kondo, H. Miura, Y.-i. Fujimura, and Y. Hayashizaki. Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. *Chromosome Research*, 2006.
- H. Ochiai, T. Sugawara, and T. Yamamoto. Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Research*, 2015.
- G. Orphanides and D. Reinberg. A Unified Theory of Gene Expression. *Cell*, 2002.
- H. W. Park, Y. C. Kim, B. Yu, T. Moroishi, J.-S. Mo, S. W. Plouffe, Z. Meng, K. C. Lin, F.-X. Yu, C. M. Alexander, C.-Y. Wang, and K.-L. Guan. Alternative Wnt Signaling Activates YAP/TAZ. *Cell*, 2015.
- P. Paszek, S. Ryan, L. Ashall, K. Sillitoe, C. V. Harper, D. G. Spiller, D. A. Rand, and M. R. H. White. Population robustness arising from cellular heterogeneity. *Proceedings of the National Academy of Sciences*, 2010.
- J. M. Pedraza and A. van Oudenaarden. Noise Propagation in Gene Networks. *Science*, 2005.
- S. G. Penheiter, H. Mitchell, N. Garamszegi, M. Edens, J. Jules J. E. Doré, and E. B. Leof. Internalization-Dependent and -Independent Requirements for Transforming Growth Factor β Receptor Signaling via the Smad Pathway. *Molecular and Cellular Biology*, 2002.
- M. K. Phanish, N. A. Wahab, P. Colville-Nash, B. M. Hendry, and M. E. C. Dockrell. The differential role of Smad2 and Smad3 in the regulation of pro-fibrotic TGF1 responses in human proximal-tubule epithelial cells. *Biochemical Journal*, 2006.

- C. E. Pierreux, F. J. Nicolás, and C. S. Hill. Transforming Growth Factor β -Independent Shuttling of Smad4 between the Cytoplasm and Nucleus. *Molecular and Cellular Biology*, 2000.
- J. L. Plank and A. Dean. Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. *Molecular Cell*, 2014.
- W. Pych. A Fast Algorithm for Cosmic-Ray Removal from Single Images. *Publications of the Astronomical Society of the Pacific*, 2004.
- C. Querfeld, B. Eckes, C. Huerkamp, T. Krieg, and S. Sollberg. Expression of TGF-*b1*, -*b2* and -*b3* in localized and systemic scleroderma. *Journal of Dermatological Science*, 1999.
- A. Raj and A. van Oudenaarden. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, (2):216–226, Oct. 2008.
- A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. Stochastic mRNA synthesis in mammalian cells. *PLoS biology*, 2006.
- A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 2008.
- R. A. Randall, M. Howell, C. S. Page, A. Daly, P. A. Bates, and C. S. Hill. Recognition of Phosphorylated-Smad2-Containing Complexes by a Novel Smad Interaction Motif. *Molecular and Cellular Biology*, 2004.
- F. Reiter, S. Wienerroither, and A. Stark. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development*, 2017.
- S. Ross and C. S. Hill. How the Smads regulate transcription. *The International Journal of Biochemistry & Cell Biology*, 2008a.
- S. Ross and C. S. Hill. How the Smads regulate transcription. *The International Journal of Biochemistry & Cell Biology*, 2008b.
- S. Ross, E. Cheung, T. G. Petrakis, M. Howell, W. L. Kraus, and C. S. Hill. Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. *The EMBO Journal*, 2006.
- P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, Nov. 1987. ISSN 0377-0427. doi: 10.1016/0377-0427(87)90125-7.
- C. E. Runyan, H. W. Schnaper, and A.-C. Poncelet. The Role of Internalization in Transforming Growth Factor *b1*-induced Smad2 Association with Smad Anchor for Receptor Activation (SARA) and Smad2-dependent Signaling in Human Mesangial Cells. *Journal of Biological Chemistry*, 2005.

Bibliography

- M. D. Ryan and J. Drew. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *The EMBO Journal*, 1994.
- H. Ryu, M. Chung, M. Dobrzyński, D. Fey, Y. Blum, S. S. Lee, M. Peter, B. N. Kholodenko, N. L. Jeon, and O. Pertz. Frequency modulation of ERK activation dynamics rewires cell fate. *Molecular Systems Biology*, 2015.
- S. Sainsbury, C. Bernecky, and P. Cramer. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 2015.
- R. Samarakoon, M. Goppelt-Struebe, and P. J. Higgins. Linking cell structure to gene regulation: Signaling events and expression controls on the model genes PAI-1 and CTGF. *Cellular signalling*, 2010.
- A. Sanyal, B. R. Lajoie, G. Jain, and J. Dekker. The long-range interaction landscape of gene promoters. *Nature*, 2012.
- A. Saunders, L. J. Core, and J. T. Lis. Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology*.
- P. Schmid, P. Itin, G. Cherry, C. Bi, and D. A. Cox. Enhanced expression of transforming growth factor-beta type I and type II receptors in wound granulation tissue and hypertrophic scar. *The American Journal of Pathology*, 1998.
- B. Schmierer and C. S. Hill. Kinetic Analysis of Smad Nucleocytoplasmic Shuttling Reveals a Mechanism for Transforming Growth Factor β -Dependent Nuclear Accumulation of Smads. *Molecular and Cellular Biology*, 2005.
- D. Schulz, V. R. T. Zanotelli, J. R. Fischer, D. Schapiro, S. Engler, X.-K. Lun, H. W. Jackson, and B. Bodenmiller. Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Systems*, pages 25–36.e5, 2018.
- B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. Global quantification of mammalian gene expression control. *Nature*, 2011.
- A. Senecal, B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller, and X. Darzacq. Transcription factors modulate c-Fos transcriptional bursts. *Cell reports*, 2014.
- J. Seoane, H.-V. Le, L. Shen, S. A. Anderson, and J. Massagué. Integration of Smad and Forkhead Pathways in the Control of Neuroepithelial and Glioblastoma Cell Proliferation. *Cell*, 2004.
- G. Serini, M.-L. Bochaton-Piallat, P. Ropraz, A. Geinoz, L. Borsi, L. Zardi, and G. Gabbiani. The Fibronectin Domain ED-A Is Crucial for Myofibroblastic Phenotype Induction by Transforming Growth Factor- β 1. *The Journal of Cell Biology*, 1998.
- K. Shah and S. Tyagi. Barriers to transmission of transcriptional noise in a c-fos c-jun pathway. *Molecular Systems Biology*, 2013.

- V. Shahrezaei and P. S. Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, 2008.
- N. C. Shaner, P. a Steinbach, and R. Y. Tsien. A guide to choosing fluorescent proteins. *Nature methods*, 2005.
- W. Shi, C. Sun, B. He, W. Xiong, X. Shi, D. Yao, and X. Cao. GADD34–PP1c recruited by Smad7 dephosphorylates TGF β type I receptor. *The Journal of Cell Biology*, 2004.
- Y. Shi and J. Massagué. Mechanisms of TGF- β Signaling from Cell Membrane to the Nucleus. *Cell*, 2003.
- Y. Shi, Y.-F. Wang, L. Jayaraman, H. Yang, J. Massagué, and N. P. Pavletich. Crystal Structure of a Smad MH1 Domain Bound to DNA: Insights on DNA Binding in TGF- β Signaling. *Cell*, 1998.
- M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science*.
- Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Molecular Cell*, 2014.
- A. Singh and M. Soltani. Quantifying Intrinsic and Extrinsic Variability in Stochastic Gene Expression Models. *PLOS ONE*, 2013.
- A. Singh, B. Razooky, C. D. Cox, M. L. Simpson, and L. S. Weinberger. Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. *Biophysical Journal*, 2010.
- S. Skylaki, O. Hilsenbeck, and T. Schroeder. Challenges in long-term imaging and quantification of single-cell dynamics. *Nature Biotechnology*, 2016.
- L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding. General properties of transcriptional time series in *Escherichia coli*. *Nature Genetics*, 2011.
- B. Sorre, A. Warmflash, A. H. Brivanlou, and E. D. Siggia. Encoding of temporal signals by the TGF- β pathway and implications for embryonic patterning. *Developmental cell*, 2014.
- F. Spitz and E. E. M. Furlong. Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, 13, 2012.
- J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad. Cellular Noise Regulons Underlie Fluctuations in *Saccharomyces cerevisiae*. *Molecular Cell*, 2012.
- J. Strasen, U. Sarma, M. Jentsch, S. Bohn, C. Sheng, D. Horbelt, P. Knaus, S. Legewie, and A. Loewer. Cell-specific responses to the cytokine TGF β are determined by variability in protein levels. *Molecular Systems Biology*, 2018.

Bibliography

- S. L. Stroschein, W. Wang, S. Zhou, Q. Zhou, and K. Luo. Negative Feedback Regulation of TGF- β Signaling by the SnoN Oncoprotein. *Science*, 1999.
- M.-H. Sung, N. Li, Q. Lao, R. A. Gottschalk, G. L. Hager, and I. D. C. Fraser. Switching of the Relative Dominance Between Feedback Mechanisms in Lipopolysaccharide-Induced NF- κ B Signaling. *Sci. Signal.*, 2014.
- D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. 2011.
- Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, 2010.
- K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozzi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nature Communications*, 2016.
- S. Tay, J. J. Hughey, T. K. Lee, T. Lipniacki, S. R. Quake, and M. W. Covert. Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature*, 2010.
- J. P. Thiery, H. Acloque, R. Y. J. Huang, and M. A. Nieto. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, 2009.
- G. Thillainadesan, J. M. Chitilian, M. Iovic, J. N. G. Ablack, J. S. Mymryk, M. Tini, and J. Torchia. TGF- β -Dependent Active Demethylation and Expression of the p15^{ink4b} Tumor Suppressor Are Impaired by the ZNF217/CoREST Complex. *Molecular Cell*, 2012.
- I. Tirosh and N. Barkai. Two strategies for gene regulation by promoter nucleosomes. *Genome Research*, 2008.
- R. Tsutsumi, M. Masoudi, A. Takahashi, Y. Fujii, T. Hayashi, I. Kikuchi, Y. Satou, M. Taira, and M. Hatakeyama. YAP and TAZ, Hippo Signaling Targets, Act as a Rheostat for Nuclear SHP2 Function. *Developmental Cell*, 2013.
- S. Tyagi. Imaging intracellular RNA distribution and dynamics in living cells. *Nature Methods*, 2009.
- X. Varelas, R. Sakuma, P. Samavarchi-Tehrani, R. Peerani, B. M. Rao, J. Dembowy, M. B. Yaffe, P. W. Zandstra, and J. L. Wrana. TAZ controls Smad nucleocytoplasmic shuttling and regulates human embryonic stem-cell self-renewal. *Nature Cell Biology*, 2008.
- X. Varelas, P. Samavarchi-Tehrani, M. Narimatsu, A. Weiss, K. Cockburn, B. G. Larsen, J. Rossant, and J. L. Wrana. The Crumbs Complex Couples Cell Density Sensing to Hippo-Dependent Control of the TGF- β -SMAD Pathway. *Developmental Cell*, 2010.
- M. K. Vartiainen. Nuclear actin dynamics – From form to function. *FEBS Letters*, 2008.

- J. M. G. Vilar, R. Jansen, and C. Sander. Signal Processing in the TGF- β Superfamily Ligand-Receptor Network. *PLOS Computational Biology*, 2006.
- P. Vizán, D. S. J. Miller, I. Gori, D. Das, B. Schmierer, and C. S. Hill. Controlling Long-Term Signaling: Receptor Dynamics Determine Attenuation and Refractory Behavior of the TGF- β Pathway. *Sci. Signal.*, 2013.
- P. Vizán, D. S. J. Miller, I. Gori, D. Das, B. Schmierer, and C. S. Hill. Controlling Long-Term Signaling: Receptor Dynamics Determine Attenuation and Refractory Behavior of the TGF- β Pathway. *Sci. Signal.*, 2013.
- T. C. Voss and G. L. Hager. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics*, 2014.
- L. M. Wakefield and C. S. Hill. Beyond TGF β : Roles of other TGF β superfamily members in cancer. *Nature Reviews Cancer*, 2013.
- A. Warmflash, Q. Zhang, B. Sorre, A. Vonica, E. D. Siggia, and A. H. Brivanlou. 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*, 2012.
- K. Wegner, A. Bachmann, J.-U. Schad, P. Lucarelli, S. Sahle, P. Nickel, C. Meyer, U. Klingmüller, S. Dooley, and U. Kummer. Dynamics and feedback loops in the transforming growth factor β signaling pathway. *Biophysical Chemistry*, 2012.
- S. J. Wicks, S. Lui, N. Abdel-Wahab, R. M. Mason, and A. Chantry. Inactivation of Smad-Transforming Growth Factor β Signaling by Ca²⁺-Calmodulin-Dependent Protein Kinase II. *Molecular and Cellular Biology*, 2000.
- M. Z. Wilson, P. T. Ravindran, W. A. Lim, and J. E. Toettcher. Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control. *Molecular Cell*, 2017.
- R. Winzen, M. Kracht, B. Ritter, A. Wilhelm, C.-Y. A. Chen, A.-B. Shyu, M. Müller, M. Gaestel, K. Resch, and H. Holtmann. The p38 MAP kinase pathway signals for cytokine-induced mRNA stabilization via MAP kinase-activated protein kinase 2 and an AU-rich region-targeted mechanism. *The EMBO Journal*, 1999.
- V. Wosika and S. Pelet. Relocation sensors to quantify signaling dynamics in live single cells. *Current Opinion in Biotechnology*, 2017.
- D. Wotton, R. S. Lo, L.-A. C. Swaby, and J. Massagué. Multiple Modes of Repression by the Smad Transcriptional Corepressor TGIF. *Journal of Biological Chemistry*, 1999.
- M. Y. Wu and C. S. Hill. TGF- β Superfamily Signaling in Embryonic Development and Homeostasis. *Developmental Cell*, 2009.

Bibliography

- Q. Xi, W. He, X. H.-F. Zhang, H.-V. Le, and J. Massagué. Genome-wide Impact of the BRG1 SWI/SNF Chromatin Remodeler on the Transforming Growth Factor β Transcriptional Program. *Journal of Biological Chemistry*, 2008.
- H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac, and I. Golding. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nature Methods*, 2015.
- L. Xu, Y. Kang, S. Çöl, and J. Massagué. Smad2 Nucleocytoplasmic Shuttling by Nucleoporins CAN/Nup214 and Nup153 Feeds TGF β Signaling Complexes in the Cytoplasm and Nucleus. *Molecular Cell*, 2002.
- P. Xu, J. Liu, and R. Derynck. Post-translational regulation of TGF- β receptor and Smad signaling. *FEBS Letters*, 2012.
- K. Yagi, D. Goto, T. Hamamoto, S. Takenoshita, M. Kato, and K. Miyazono. Alternatively Spliced Variant of Smad2 Lacking Exon 3 COMPARISON WITH WILD-TYPE Smad2 AND Smad3. *Journal of Biological Chemistry*, 1999.
- X. Yan, Z. Liu, and Y. Chen. Regulation of TGF- β signaling by Smad7. *Acta Biochimica et Biophysica Sinica*, 2009.
- L. Yang, Y. Pang, and H. L. Moses. TGF- β and immune cells: An important regulatory axis in the tumor microenvironment and progression. *Trends in Immunology*, 2010.
- Q.-L. Ying, J. Nichols, I. Chambers, and A. Smith. BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. *Cell*.
- R. A. Young. Control of the Embryonic Stem Cell State. *Cell*, 2011.
- S. Yunger, L. Rosenfeld, Y. Garini, and Y. Shav-Tal. Single-allele analysis of transcription kinetics in living mammalian cells. *Nature Methods*, 2010.
- S. Yunger, L. Rosenfeld, Y. Garini, and Y. Shav-Tal. Quantifying the transcriptional output of single alleles in single living mammalian cells. *Nature Protocols*, 2013.
- J. Zhang, X. Zhang, F. Xie, Z. Zhang, H. van Dam, L. Zhang, and F. Zhou. The regulation of TGF- β /SMAD signaling by protein deubiquitination. *Protein & Cell*, 2014.
- Q. Zhang, Y. Yoon, Y. Yu, E. J. Parnell, J. A. R. Garay, M. M. Mwangi, F. R. Cross, D. J. Stillman, and L. Bai. Stochastic expression and epigenetic memory at the yeast HO promoter. *Proceedings of the National Academy of Sciences*, 2013.
- X. L. Zhang, N. Topley, T. Ito, and A. Phillips. Interleukin-6 Regulation of Transforming Growth Factor (TGF)- β Receptor Compartmentalization and Turnover Enhances TGF- β 1 Signaling. *Journal of Biological Chemistry*, 2005.

- Y. Zhang, C. Chang, D. J. Gehling, A. Hemmati-Brivanlou, and R. Derynck. Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase. *Proceedings of the National Academy of Sciences*, 2001.
- Y. E. Zhang. Non-Smad pathways in TGF- β signaling. *Cell Research*, 2009.
- B. Zhao, X. Ye, J. Yu, L. Li, W. Li, S. Li, J. Yu, J. D. Lin, C.-Y. Wang, A. M. Chinnaiyan, Z.-C. Lai, and K.-L. Guan. TEAD mediates YAP-dependent gene induction and growth control. *Genes & Development*, 2008.
- Z. Zi, Z. Feng, D. A. Chapnick, M. Dahl, D. Deng, E. Klipp, A. Moustakas, and X. Liu. Quantitative analysis of transient and sustained transforming growth factor- β signaling dynamics. *Molecular Systems Biology*, 2011.
- Z. Zi, D. A. Chapnick, and X. Liu. Dynamics of TGF- β /Smad signaling. *FEBS Letters*, 2012.
- A. Zieba, K. Pardali, O. Söderberg, L. Lindbom, E. Nyström, A. Moustakas, C.-H. Heldin, and U. Landegren. Intercellular Variation in Signaling through the TGF- β Pathway and Its Relation to Cell Density and Cell Cycle Phase. *Molecular & Cellular Proteomics*, 2012.
- B. Zoller, D. Nicolas, N. Molina, and F. Naef. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology*, 2015.
- C. J. Zopf, K. Quinn, J. Zeidman, and N. Maheshri. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. *PLoS Computational Biology*, 2013.

Onur Tidin

<https://people.epfl.ch/onur.tidin>

Date of birth : June 21, 1988

Email : onur.tidin@epfl.ch

onurtidin@gmail.com

EDUCATION

- **École Polytechnique Fédérale de Lausanne (EPFL)** Lausanne, Switzerland
PhD in Biotechnology and Bioengineering 2014 – 2018
- **École Polytechnique Fédérale de Lausanne (EPFL)** Lausanne, Switzerland
Master of Science in Electrical Engineering, Microelectronics 2011 – 2013
- **Bilkent University** Ankara, Turkey
B.Sc., Electrical and Electronics Engineering 2007 – 2011

PUBLICATIONS

Journals

- Blanchoud, S., Nicolas, D., Zoller, B., Tidin, O., and Naef, F. (2015). “CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings”, [Methods](#).
- S. B. Eryilmaz, O. Tidin, A. K. Okyay, (2011). “Plasmonic Nanoslit Array Enhanced Metal - Semiconductor - Metal Optical Detectors”, [IEEE Photon. Technol. Lett.](#)

Conferences

- Onur Tidin, Felix Naef, David Suter, “Input-output relationships underlying transcriptional dynamics”, 3rd International SystemsX.ch Conference on Systems Biology, September 4-7, 2017, ETH Zurich, Switzerland (Best Poster Award)
- Onur Tidin, Felix Naef, David Suter, “Input-output relationships underlying transcriptional dynamics”, All SystemsX.ch Day, September 1, 2016, Bern, Switzerland
- Onur Tidin, Felix Naef, David Suter “Single mammalian cell transcription bursting dynamics upon stimulation”, EPFL Bioengineering Day Poster Session, September, 19, 2014, Lausanne, Switzerland

SKILLS

Experimental: Luminescence and Fluorescence Microscopy, Cell Culture, Bacterial Culture, Cloning, Immunofluorescence, Western Blot, Microfabrication

Programming: MATLAB, Image Analysis Tools (ImageJ, Cell Profiler, iLastik), R, L^AT_EX, Shell, Python

Languages: English (fluent), French (intermediate), German (basic), Turkish (native)