Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M.Sc. Enoch Boasiako Antwi

born in: Kuntanase, Ghana

oral examination date: June 29, 2021

An optogenetics approach to decipher mammalian promoter decoding of transcription factor dynamics

Referees: **Prof. Dr. Ursula Kummer**

Prof. Dr. Barbara Di Ventura

Für meine Eltern, Esther und Samuel Antwi,

meine Verlobte Regina Gyasi

Acknowledgements

I would like to express my gratitude to Prof. Dr. Barbara Di Ventura for giving me the opportunity to work on this interesting interdisciplinary project. I would like to thank her for supervising this project with charisma and pushing for this project to be presented at many conferences. Thank you Barbara; for the thorough scientific discussion and being patient with me through our disagreements. Likewise, I appreciate you for providing the funds I needed through this long PhD period. I look forward to the time we have left and for future collaborations. I will also want to extend my appreciation to Prof. Dr. Ursula Kummer for gladly accepting to be my TAC supervisor. Thank you for your genuine interest in my well-being in all aspects of my PhD time and even before. I hope there will be more time and opportunity for me to express my gratitude for all the things you have done for me. Many thanks to Prof. Dr. Thomas Höfer for the interest in my project and the detailed discussions and comments during my TAC meetings. I will also like to extend my gratitude to Prof. Dr. Ursula Klingmüller for accepting to be part of my TAC and taking time off her busy schedule to attend them.

A special thanks goes to Prof. Dr. Thomas Brox for accepting to help with our challenging segmentation problem and dedicating PhD students to the problem. I would like to thank Özgün Çiçek and Yassine Marrakchi for taking on the challenge and generating the neuronal network that was used for the automated quantification of the protein expression microscopy data generated in this study. Thanks to you Yassine, for the discussions and all your hard work, it is really appreciated. My sincere gratitude goes to Dr. Mehmet Ali Öztürk for the interesting scientific and non-scientific discussions. Thank you for being warm and for listening to all my jokes. We will definitely have the beer time.

In addition, I would like to express my gratitude to Dr. Pierre Wehler for being my transcription factor dynamics partner for the most part of my PhD time. I would like to thank Dr. Navaneethan Palanisamy for sharing an office with me through my PhD period and listening to my scientific thoughts while trying to make sense of my data. Have a good life Eethan. A special thank you to Edoardo Romano (the Romano de roma of our lab) for willingly accepting to compete with me from the first day of our PhD life. I hope you have a good after-PhD life buddy. I also thank Daniel Weiss for all the jokes he made to create a lively lab environment and for correcting my Deutsch. Thank you Giada for taking the time to help proofread this thesis. My appreciation to Dagmar Krischke for creating an orderly lab and Simone Ugi for the administrative support. I would like to express my gratitude to all past and present members of the Di Ventura lab for making PhD life bearable.

I would like to thank my parents, Mr. and Mrs. Antwi, for their support, love and encouragement through my academic life. To my siblings, I say thank you for all the love and for cheering me on through the tough times. Special thanks to Regina Gyasi, my fiancée and soon to be wife for the love and support. Thank you for the fun times that helped me escape the stress momentarily. I look forward to more fun days ahead.

To all the friends I made along my academic journey, I say thank you very much for the good times we had. God bless you all!!!.

Contributions

All data were generated by myself under the supervision of Prof. Dr. Barbara Di Ventura. ‘All data’ include but are not limited to the design and construction of expression vectors, cell culture and microscopy experiments, quantification of mRNA data, mathematical modelling as well as data analysis and presentation.

The neural network used for the automated quantification of protein expression data from microscopy experiments was generated by Özgün Çiçek and Yassine Marrakchi from the computer vision group headed by Prof. Dr. Thomas Brox (paragraph 5.1.8).

Through this thesis, I use the term ‘we’ when concepts or experiments were obtained with help from others. Concepts developed and experiments performed by myself are described with the pronoun ‘I’.

Publications

Peer-reviewed publications:

I contributed to the following publications during my PhD:

Çiçek*, Ö., Marrakchi*, Y., Antwi, E.B., Di Ventura, B. & Brox, T (2020): Recovering the Imperfect: Cell Segmentation in the Presence of Dynamically Localized Proteins. *Lecture Notes in Computer Sci.* 12446: 85–93.

Geller, S.H., Antwi, E.B., Di Ventura, B. & McClean, M.N. (2019): Optogenetic Repressors of Gene Expression in Yeasts Using Light-Controlled Nuclear Localization. *Cellular and Molecular Bioengineering*. 12(5): 511–28.

The following publications are related to work done prior to my PhD:

Antwi, E.B., Olins, A., Teif, V.B., Bieg, M., Bauer, T., Gu, Z., Brors, B., Eils, R., Olins, D. & Ishaque, N. (2020): Whole-genome fingerprint of the DNA methylome during chemically induced differentiation of the human AML cell line HL-60/S4. *Biology Open.* 9(2).

Antwi, E.B., Haanstra, J.R., Ramasamy, G., Jensen, B., Droll, D., Rojas, F., Minia, I., Terrao, M., Mercé, C., Matthews, K., Myler, P.J., Parsons, M. & Clayton, C. (2016): Integrative analysis of the Trypanosoma brucei gene expression cascade predicts differential regulation of mRNA processing and unusual control of ribosomal protein expression. *BMC Genomics.* 17(1): 306.

Teif, V.B., Gould, T.J., Clarkson, C.T., Boyd, L., Boyd, L., Antwi, E.B., Ishaque, N., Ishaque, N., Olins, A.L. & Olins, D.E. (2020): Linker histone epitopes are hidden by in situ higher-order chromatin structure. *Epigenetics and Chromatin.* 13(1): 26.

Poster and oral presentations as presenting author:

Results on decoding transcription factor dynamics have been presented at the following conferences and meetings: (Enoch B. Antwi, Barbara Di Ventura)

HBIGS PhD summer retreat, Überlingen, Germany. July 2017 (Oral presentation).

Advanced lecture course on systems biology (SYSBIO), Innsbruck, Austria. February 28 – March 6, 2018 (Poster).

International conference on systems biology of human diseases (SBHD), Los Angeles, USA. June 4-6, 2018 (Poster).

German Associaation of synthetic biology (GASB) II conference, Berlin, Germany. September 27-28, 2018 (Poster).

Results on transcription factor architecture have been presented at the following conferences and meetings: (Enoch B. Antwi, Barbara Di Ventura)

3rd Course on Multiscale Integration in Biological Systems, Paris, France. November 7-13, 2018 (short talk and Poster).

Abstract

Several transcription factors (TFs), such as the tumour suppressor p53, the immune response regulator NF-κB, the yeast stress response regulator Msn2 and others, exhibit different nuclear accumulation patterns (dynamics) depending on the upstream activating stimulus. TF dynamics thus encode information about the type and intensity of the stimulus perceived by a eukaryotic cell. TF dynamics are believed to govern cell fate because they lead to the activation of distinct sets of target genes. Studies on how information about either internal or external stimuli is transmitted through signalling pathways into specific cell fates have shown that promoters of target genes play a critical role in decoding the information encoded in TF dynamics. Earlier studies suggested that the binding affinities of the TF for the promoters of the different target genes may orchestrate the observed differential gene expression under different TF dynamics. It was later shown that nucleosome positioning and, as a consequence, promoter accessibility determines how rapidly a promoter gets activated, thus making it more or less sensitive to different TF dynamics. The distance between the core promoter and the TF binding sites as well as the core promoter itself have also been demonstrated to affect the expression of different target genes. Other studies on p53 and NF-κB measuring transcript levels of various target genes have revealed differences in the stability of transcripts belonging to early and late response genes. Some p53 target genes show oscillatory transcript levels in response to p53 pulses. In such studies, an external stimulus such as a cytokine, radiation or a chemical agent whose effects may not be fully understood were used to impose different TF dynamics. The lack of full clarity on the effects of the agents used to induce the TF dynamics may therefore undermine observations and the explanations given in such studies. Despite the progress made in understanding the role played by TF dynamics in gene expression regulation, it is still not clear which mammalian promoter elements contribute to decoding TF dynamics, and how they do so.

In this study, I constructed a library of synthetic optogenetic circuits consisting of a library of synthetic light-responsive TFs and a library of promoters designed with well-studied elements to investigate the relationship between TF dynamics and promoter activation in mammalian cells. Such a synthetic biology approach allows us to minimize the complexity, which is inevitable when studying endogenous pathways.

I observed that there is a threshold for the time the TF must remain bound to the cognate responsive elements (REs) at the promoter (TF dwelling time) for transcription to be successfully initiated. The TF dwelling time is set by the affinity of the TATA binding protein (TBP) for the TATA-box (TB). A high-affinity TATA-box consents efficient assembly of the transcription pre-initiation complex (PIC), which reduces the TF dwelling time required for transcription initiation. The affinity of the TF for the REs defines the TF concentration (amplitude) threshold necessary to achieve the required TF dwelling time. Consequently, promoters with low-affinity REs and TATA-box filter out low-frequency pulsatile signals, but are activated by sustained TF signals. Additionally, reducing DNA looping efficiency by increasing the distance between the REs and the TATA-box, turns an otherwise TF dynamics-insensitive promoter into a promoter that can distinguish TF dynamics. I also show that the efficiency of translation initiation is critical for differential expression of target genes in response to different TF dynamics observed at the protein level.

Finally, I investigated a different type of synthetic TF bearing only the DNA binding domain (DBD) which interacts with a light-responsive co-regulator that bears the transactivation domain (TAD). This scenario resembles several natural TFs, such as the TEAD/YAP pair. I found that this system is very sensitive to the interaction strength between the DBD- and TAD-bearing proteins. Furthermore, a high concentration of nuclear DBD-bearing TF impedes gene expression due to the competition for the REs between its free and TAD-bearing protein-bound fractions. These observations will help to further understand gene expression regulation by dynamics and how TEAD concentration in mammalian cells can be targeted in cancers where TEAD/YAP is dysregulated.

Zussammenfassung

Mehrere Transkriptionsfaktoren (TFs), wie der Tumorsuppressor p53, der Immunregulator NF-κB, der Hefestressregulator Msn2 und andere, zeigen je nach Aktivierungsstimulus unterschiedliche Zellkernakkumulationsmuster (Dynamik). Die TF-Dynamik codiert somit Informationen über die Art und Intensität des von einer eukaryotischen Zelle wahrgenommenen Stimulus. Es wird angenommen, dass die TF-Dynamik das Zellschicksal bestimmt, da sie zur Aktivierung unterschiedlicher Zielgenen führt. Studien darüber, wie Informationen über interne oder externe Stimuli über Signalwege in bestimmte Zellschicksale übertragen werden, haben gezeigt, dass Promotoren von Zielgenen eine entscheidende Rolle bei der Dekodierung der in der TF-Dynamik enthaltenen Informationen spielen. Frühere Studien deuteten darauf hin, dass die Bindungsaffinitäten des TF für die Promotoren der verschiedenen Zielgene die beobachtete differentielle Genexpression unter verschiedenen TF-Dynamiken steuern können. Später wurde gezeigt, dass die Nukleosomenpositionierung und damit die Zugänglichkeit des Promotors bestimmen, wie schnell ein Promotor aktiviert wird, wodurch er mehr oder weniger empfindlich für unterschiedliche TF-Dynamiken wird. Es wurde auch gezeigt, dass der Abstand zwischen dem Zellkernernpromotor und den TF-Bindungsstellen sowie dem Zellkernernpromotor selbst die Expression verschiedener Zielgene beeinflusst. Andere Studien zu p53- und NF-κB-Messungen der Transkriptionsniveaus verschiedener Zielgene haben Unterschiede in der Stabilität von Transkripten gezeigt, die zu frühen und späten Antwortgenen gehören. Einige p53-Zielgene zeigen oszillatorische Transkriptniveaus als Reaktion auf eine oszillierende Zellkernernakkumulation von p53. In solchen Studien wurde ein externer Stimulus wie Zytokine, Strahlung oder chemische Mittel verwendet, deren Auswirkungen möglicherweise nicht vollständig verstanden werden, um die TF-Dynamik zu bewirken. Der Mangel an Verständnis über die Wirkungen der Mittel, die zur Induktion der TF-Dynamik verwendet werden, kann daher die Beobachtungen und die in solchen Studien gegebenen Erklärungen untergraben. Trotz der Fortschritte beim Verständnis der Rolle der TF-Dynamik bei der Regulation der Genexpression ist immer noch nicht klar, welche Säugetier-Promotorelemente zur Dekodierung der TF-Dynamik beitragen und wie sie dies tun.

In dieser Studie konstruierte ich eine Bibliothek synthetischer optogenetischer Schaltkreise, die aus einer Bibliothek synthetischer lichtempfindlicher TFs und einer Bibliothek von Promotoren besteht, die mit gut verstandenen Elementen entworfen wurden, um die Beziehung zwischen TF-Dynamik und Promotoraktivierung in Säugetierzellen zu untersuchen. Ein solcher Ansatz der synthetischen Biologie ermöglicht es uns, die Komplexität zu minimieren, die bei der Untersuchung endogener Pfade unvermeidlich ist.

Ich beobachtete, dass es einen Schwellenwert für die Zeit gibt, in der der TF an die responsiven Elemente (REs) im Promotor (TF-Verweilzeit) gebunden bleiben muss, damit die Transkription erfolgreich initiiert werden kann. Die TF-Verweilzeit wird durch die Affinität des TATA-Bindungsproteins (TBP) zu einer TATA-Box (TB) definiert. Eine hochaffine TATA-Box führt zu einer starken Bindung und einer effizienten Assemblierung des Transkriptions-Pre-Initiation-Komplexes (PIC). Dies reduziert die für die Transkriptionsinitiierung erforderliche TF-Verweilzeit. Die Affinität des TF zu den REs definiert auch den TF-Konzentrationsschwellenwert (Amplitudenschwelle), der erforderlich ist, um die erforderliche effektive TF-Verweilzeit zu erreichen.Folglich filtern Promotoren mit niedrigaffinen REs und

TATA-Box niederfrequente pulsierende Signale heraus, werden jedoch durch anhaltende TF-Signale aktiviert . Die Verringerung der Effizienz der DNA-Schleifenbildung durch Erhöhen des Abstands zwischen den REs und der TATA-Box wird ein ansonsten TF-dynamikunempfindlicher Promotor zu einem Promotor, der die TF-Dynamik unterscheiden kann. Ich zeige auch, dass die Effizienz der Translationsinitiierung für die differentielle Expression von Zielgenen als Reaktion auf unterschiedliche TF-Dynamiken, die auf Proteinebene beobachtet werden, entscheidend ist.

Schließlich untersuchte ich einen anderen Typ von TF, der nur die DNA-Bindungsdomäne (DBD) trägt, welche mit einem auf Licht ansprechenden Co-Regulator interagiert, der die Transaktivierungsdomäne (TAD) trägt. Dieses Szenario ähnelt mehreren natürlichen Szenarien, z. B. dem TEAD / YAP-Paar. Ich fand heraus, dass dieses System sehr empfindlich auf die Wechselwirkungsstärke zwischen den DBD- und TAD-tragenden Proteinen reagiert. Darüber hinaus behindert eine hohe Konzentration von nuklearem DBD-tragendem TF die Genexpression aufgrund der Konkurrenz zwischen seinem freien und der durch TAD-tragendes Protein gebundenen Fraktion um die REs. Diese Erkenntnisse weden dazu beitragen, die Regulation der Genexpression durch die TF-Dynamik besser zu verstehen und dabei helfen wie die TEAD-Konzentration in Säugetierzellen bei Krebserkrankungen, bei denen TEAD / YAP fehlreguliert ist, gezielt eingesetzt werden.

Contents

Acknowledgements i
Contributions iii
Publications v
Abstract vii
Zussammenfassung ix
Contents xi
INTRODUCTION 1
1.1 Transcription factors 1
1.1.1 Transcriptional repressors 1
1.1.2 Transcriptional activators 1
1.2 Promoters 4
1.2.1 Core promoter elements 5
1.3 Transcription initiation 6
1.3.2 Pre-mRNA processing 10
1.3.3 Translation 10
1.4 Transcription factor dynamics 12
1.4.1 Instances of transcription factor dynamics in eukaryotic cells 13
1.4.2 Promoter decoding of transcription factor dynamics 15
1.4.3 Light-control of transcription factor dynamics 17
1.5 Transcription factor architecture 18
1.5.1 Yes-associated protein (YAP) as a co-activator 19
1.5.2 TEA domain (TEAD) transcription factors 20
1.6 Mathematical modelling 21
1.6.1 Mathematical modelling of transcription factor dynamics 22
1.7 Aims of the study 23
2 RESULTS 25
2.1 Decoding TF dynamics by promoters of target genes 25
2.1.1 Design and construction of synTF 25
2.1.2 Construction of the synthetic promoter library 28
2.1.3 One-plasmid system 30
2.1.4 Reporter expression induced by different TADs 32
2.1.5 Experimental setup and data quantification 33
2.1.6 A promoter with high affinity REs and TATA-box cannot distinguish between sustained and pulsatile TF dynamics 36
2.1.7 Lower TBP affinity for the TATA-box alone does not enable decoding of TF dynamics 39

2.1.8 A promoter with high affinity TB but low affinity REs is sensitive to synTF amplitude but not frequency 40
2.1.9 A promoter with low-affinity REs and TATA-box is sensitive to TF dynamics . 41
2.1.10 Mathematical model of gene expression 43
2.1.11 Low DNA looping efficiency leads to low protein expression but allows for differential expression in response to different TF dynamics 46
2.1.12 Reduced mRNA translatability improves the ability of a promoter to distinguish different TF dynamics 49
2.2 Transcription factor architecture and how it impacts gene expression 51
2.2.1 Design and construction of synTEAD/synYAP transcription factor and characterization of synYAP nuclear accumulation dynamics 51
2.2.2 SynTF has lower nuclear accumulation background and higher dynamic range than synYAP 53
2.2.3 Higher synYAP/synTEAD affinity and lower synTEAD concentration improve the functionality of the TF 55
2.2.4 synTF outperforms any synTEAD/synYAP TF construct 56
2.2.5 SynYAP can be translocated in and out of the nucleus by interacting with a partner protein 58
3 DISCUSSION AND OUTLOOK 63
3.1 Transcription factor dynamics 63
3.1.1 Choice of experimental setup for studying TF dynamics impacts interpretation of data 63
3.1.2 Promoters requiring high TF amplitudes or long TF dwelling time distinguish well between different TF dynamics 64
3.1.3 Translation initiation and mRNA structure affect target gene expression in response to different TF dynamics 67
3.1.4 Higher reporter expression by pulsatile TF dynamics: is it possible? 68
3.1.5 A theoretical exercise: modifying the mathematical model to obtain higher reporter expression under pulsatile dynamics 69
3.1.6 Outlook on TF dynamics 70
3.2 TF architecture 71
3.2.1 Strong TEAD/YAP interaction is necessary for effective reconstitution of a functional TF 71
3.2.2 Diffusion and protein-protein interaction drive YAP nuclear translocation 71
3.2.3 Intermediate TEAD levels drive better expression of target genes 72
3.2.4 Relevance and outlook of synthetic TEAD/YAP studies 72
4 REFERENCES 73
5 MATERIALS AND METHODS 91
5.1 Decoding transcription factor dynamics 91
5.1.1 Plasmid Construction 91

5.1.2 LOVtrap experimental setup 91
5.1.3 Nascent mRNA visualization construct: 91
5.1.4 Mammalian cell culture and transfection: 91
5.1.5 Imaging of synTF dynamics and reporter expression: 92
5.1.6 Transcript visualization and quantification: 92
5.1.7 Flexibility and bendability of 601 and Random sequence 93
5.1.8 Automated quantification protein signals: 93
5.1.9 Mathematical model of synthetic promoter response to synTF 94
5.1.10 Parameter fitting 96
5.2 TF architecture 96
5.2.1 Construction of synTEAD/synYAP 96
5.2.2 LOVtrap 97
5.2.3 Blue light illumination of synYAP co-activator 97
5.2.4 Construction of synTEAD/synYAP without LINuS 97
5.2.5 QUANTIFICATION AND STATISTICAL ANALYSIS 97
6 APPENDIX 99
6.1 Supplementary Figures 99
6.2 Supplementary Tables 104
6.3 Supplementary Sequences 115
6.4 List of Figures 119
6.5 List of Tables 120
6.6 Abbreviations 121

INTRODUCTION

1.1 Transcription factors

Gene expression in eukaryotic cells is a complex multi-step biological process and includes transcription – the process of copying information on the genomic DNA into messenger RNA (mRNA) and translation – making of functional proteins from the information transcribed on the mRNA. Transcription is constitutive for some genes, mostly house-keeping genes that need to be expressed continually to ensure the homeostasis and survival of the cell [1,2]. Expression of other genes is cell type- and developmental stage-specific. Expression of regulated genes may also be in response to an external stimulus or stress [3–5]. Transcription factors (TFs) serve as the primary regulatory proteins for modulating the expression of stimulus-induced genes [6–8]. The level of gene expression triggered by stress-induced TFs is dependent on the type or level of stress [9]. TFs regulate expression by binding to their specific cis-regulatory sequences also termed responsive elements (REs) in the promoter of their target genes. Functionally, TFs are classified into two groups: (I) repressors that inhibit the expression of their target genes [10–12] and (II) activators that induce or enhance the expression of their target genes [8,13].

1.1.1 Transcriptional repressors

Repressor proteins impair the transcriptional activation of genes. The mammalian transcriptional repressor can act passively by blocking the DNA binding site of transcriptional activators or directly sequestering the activator in a sub-cellular compartment other than the nucleus, thus causing effective inhibition of transcription induction. Repressors can also actively remodel the chromatin structure making the activator binding site inaccessible [14,15]. Nucleosomes for example can locally or globally inhibit transcription by sequestering the DNA around the promoters of genes [16]. Other global transcriptional repressors act by targeting the general transcription machinery either via sequestration or post-translational modification of components of the transcription pre-initiation complex (PIC) or RNA polymerase II [12]. The PIC is the state in which RNAP II and the general transcription factors are all bound to the core promoter but remain in an inactive conformation for initiation of transcription [17]. Gene-specific transcriptional repression occurs locally when repressors reduce the concentration of functional activators via post-translational modifications which prevent activators from binding to their target sites or interacting with the co-activators or general transcription machinery [12,14].

1.1.2 Transcriptional activators

Transcriptional activators are transcription regulatory proteins that induce or improve transcription of a gene or set of genes. Gene-specific DNA-binding TFs generally contain at least two domains: the DNA binding domain (DBD) and the transactivation domain (TAD) [18,19]. The DBD recognizes specific DNA sequences hence directs the TF to binding sites of target

genes. The DNA recognition by DBD is based on its specific structure and amino acid sequence. In contrast, the TAD domain has no defined structural motif. Transcriptional activators may consist of both DBD and TAD in the same protein molecule or may consist of only the DBD, and thus, require interaction with a co-activator that bears the TAD to induce gene expression [20–22]. Once the TF is recruited to the right target gene via the DBD, the TAD serves as the effector for recruiting and assembling the general transcription factors and/or chromatin remodelling proteins to help initiate transcription. The DBD and TAD domains of a TF can be separated from each other and fused to heterologous domains to target new genes [19].

1.1.2.1 DNA binding domains

TF are classified into families defined by the structures of the DBDs. DNA binding structures are categorized by a few structurally conserved motifs [19,23]. Despite the existence of a wide variety of TFs, the largest classes which account for over 80% of all known human TFs are the C2H2 zinc fingers, homeodomains and basic helix-loop-helix (bHLH) [7,23]. The C2H2 zinc finger TFs contain the Cys2His2 zinc finger motifs in their DBD. While being the largest class of all putative human TFs, there are no known function for about 80% of them [24]. Homeodomain TFs are the largest family of TFs that bind to DNA using the helix-turn-helix (HTH) structure. Homeodomain is a 60 amino acid HTH structure [23,25–27]. The HTH superfamily is prevalent in eukaryotes; its members however do not share sequence similarity with their counterparts in prokaryotes [28]. The bHLH superfamily of TFs consist of two highly conserved domains that have distinct functions. The two domains are consist of a total of ~60 amino acids. The amino-terminal (N-terminal) domain is a basic domain that binds to the six nucleotide DNA consensus sequence called E-box. At the carboxyl-terminal (C-terminal) is the HLH domain which is used for the protein-protein interaction with other protein units to form homo or heterodimers [29–31].

1.1.2.1.1 Binding affinity of the DNA binding domain

DBDs of TFs recognize specific DNA sequence called the responsive element or operator sequence in bacteria. Most TFs in eukaryotes bind to 6-20bp sequence degenerately [32–34]. RE sequence and the flanking sequences of the core RE affect the TF binding by dictating the DNA structure. The right DNA structure allows for TF binding to the major or minor grooves depending on TF type [35]. It is however generally assumed in bioinformatics that nucleotide positions contribute to binding affinity independently [32].

1.1.2.1.2 Specificity of the DNA binding domain

The DBD of a TF binds preferentially to its binding sequence although it also binds non-specifically while searching for its specific sequence. The preference of TF for a given nucleotide at a specific position is in the most part determined by the physical interaction between the accessible edges of the base pairs (bp) and the TF amino acid side chains that make the contact – an interaction termed as base readout [36,37]. TF can as well recognize the sequence dependent DNA bending and unwinding at their binding sites [36,38,39]. This DNA structure recognition mechanism is known as shape readout.

In addition, cooperative interactions at the TF binding site is known to increase the specificity of the regulatory process thus, making DNA recognition by TFs more robust [32,40]. There are two forms of cooperative interactions that can increase the specificity of DNA-TF interaction. The first is the direct cooperativity in which TFs form oligomeric complexes at the site of strong interaction with the DNA [41]; and the indirect cooperativity where TFs facilitate each other's

binding via local chromatin remodelling [32]. Similarly, multiple binding sites and the distance between them contribute to the specificity and affinity of TF [35] since binding of one site can influence the binding at another site in close proximity [42]. Specificity of DNA binding to its binding sites is also influenced by the abundance of TF [43].

1.1.2.1.3 LexA DNA binding domain

The 202 amino acid *E. coli* protein, LexA is a member of the winged HTH DBD family of proteins and folds into two structurally distinct domains which are joined together by a flexible linker region. The winged HTH motif is a variant of the HTH DBD [44]. LexA forms dimers using its carboxyl terminal domain and binds to its operator sequence using the HTH structure in its amino terminal domain [45]. LexA is a DNA damage regulatory repressor protein in *E. coli*. LexA and its binding sequences are well-studied which makes it appropriate for systematic study of DNA binding affinity and its effect on gene expression. LexA binds to different operator sequences with varying affinities. The operator sequence of LexA has the consensus sequence CTGTN$_8$ACAG. The N$_8$ spacer length is invariant but LexA affinity for the operator changes when the N$_8$ sequence is varied [46].

Table 1.1: Sequences of LexA responsive elements.

Sequence	Name	K_D	Source
CTGTATATATATACAG	Consensus	0.80nM	[46,47]
CTGTATATAAAACCAG	recN1	<2nM	[47]
CTGTATGATCATACAG	recA	1.67nM	[46]
CTGTAAAAAAAAACAG	A-tract spacer	5.64nM	[46]

1.1.2.2 Transactivation domains

The transactivation domains (TAD) also known as the transcription activation domains are mostly unstructured in solution but form specific structures upon binding to their target. This implies that it is the binding partner of the TAD that serves as a template for shaping the unstructured TAD thus, allowing for TADs such as VP16 to interact with different components of the general transcription machinery [48,49]. This mechanism creates a system in which TADs are not restricted in their ability to recruit the transcription initiation complex since they have evolved flexible ways of interacting with multiple components of the transcription PIC [49].

The eukaryotic TADs are mainly classified with respect to their amino acids composition. They can be rich in acidic amino acid residues (example p53), in glutamine residues (example sp1) or in proline residues (example AP-2). Members of each class of TAD can interact with various components of the general transcriptional machinery such as the TATA binding protein (TBP) associated factors (TAFs) and the TFIIB [50,51]. The most abundant and thus well studied TADs are the ones characterized by an over representation of acidic and hydrophobic amino acids, however, the exact amino acid sequence are variable. Mutational analysis of p53 and RelA showed that the ability of TADs to induce transcription is more sensitive to mutations in the hydrophobic amino acid patch than mutation of acidic amino acids which defines this class of TADs [52,53]. This implies that it is rather the overall negative charge and not the exact sequence that is the important determinant of a TAD's potency [19]. The acidic TADs for example, have

no specific conserved three-dimensional structure when not bound to their partners but adopt α-helical conformation when bound to interacting proteins. Acid TADs are strong activators whereas the proline-rich and glutamine-rich TADs are considerably weaker gene activators [19].

Eukaryotic TADs can also be functionally classified into those that induce transcription initiation and those that induce transcription elongation on the basis of the different contacts they make with general transcription machinery [54]. Generally, TAD induces transcription initiation by stabilizing the pre-initiation complex or by facilitating promoter clearance and enhancing the elongation rate [54–58]. PIC stabilization can be achieved via interactions with subunits of the pre-initiation complex. TAD of VP16 for example, interacts directly with multiple subunits of the PIC such as the TBP. Promoter clearance by TADs arise when they recruit chromatin remodelling proteins such as the Spt-Ada-Gcn5 acetyltransferase (SAGA) multi-protein complex and other chromatin remodelers to clear nucleosomes that may be blocking elongation.

1.1.2.2.1 VP16 transactivation domain

A typical example of a strong acidic TAD that induces transcription initiation is found on the *Human Herpes Simplex Virus 1* Protein 16 (VP16). The TAD of VP16 is subdivided into two regions; H1 and H2 which can independently induce transcription in vivo and in vitro [59]. The VP16 TAD on its own just as other acidic TADs, has no specific three-dimensional structure when it is not bound. It however, adopts α-helical conformation when bound to components of general transcription factors such as TFIID [60,61]. It can directly interact with TBP, TFIIB and the SAGA histone acetylase complex in vivo [62]. Many subunits of the general transcription machinery interact with the VP16 TAD mostly via the H1 region. The H2 region can also bind directly to the histone acetyltransferase (HAT) CREBBP both in vivo and in vitro. The human Mediator co-activator complex has been shown to bind to both the H1 and H2 regions [63].

1.2 Promoters

The promoter of a gene is a DNA sequence that defines the site of transcription initiation by the RNA polymerase II (RNAP II) and is generally located upstream (5’ end) of the corresponding gene. Transcription by RNAP II initiates at a defined position called the transcription start site (TSS), which is embedded within the core promoter. The core promoter is a short sequence spanning ~50bp upstream and ~50bp downstream of the TSS. The core promoter sequence serves as the binding platform for RNAP II and its associated general transcription factors [64]. The core promoter is therefore the minimal DNA sequence capable of inducing non-regulated or basal transcription [17,65]. The transcription activity induced by the core promoter is susceptible to chromatin modelling and activation by cis-regulatory elements such as enhancers [66–68]. The core promoter can be grouped into three [69]; (I) the tissue/gene-specific core promoters (with sharply defined initiation site), (II) house-keeping core promoters (with dispersed initiation sites) and (III) the developmental gene core promoters (marked by long single CGI or multiple CGIs).

1.2.1 Core promoter elements

The core promoter contains functionally distinct DNA motifs called core promoter elements. The core promoter of eukaryotic genes contains different promoter elements – most of which have fixed positions relative to the TSS. Some regularly occurring core promoter elements are the TATA-box [70–72], the Initiator sequence (INR) [73], upstream and downstream transcription factor IIB (TFIIB) recognition element (BREu and BREd respectively) [74], motif ten element (MTE) [75], downstream promoter element (DPE) [76] and the downstream core promoter elements (DCE) [77]. Other core promoter elements occur on special classes of gene promoters such as ribosomal protein gene promoters [78]. Some of these elements are synergistic when they occur together on the same promoter, but they do not all necessarily occur together on a single promoter. For example, the most well studied core promoter element – the TATA-box occurs in the core promoter of less than 30% of all mammalian protein coding genes [79] and it is present in combination with INR and/or BREu-like elements about 30% of the time [80]. The TATA-box is located ~30bp upstream of the TSS [81] in the sharp/focused core promoter [69]. The TATA-box is highly conserved in eukaryotes from yeast to human but it is present in a minority of core promoters but enriched in sharp/focused core promoters [75,82]. The TATA-box is recognized and bound by the TBP [83]; a subunit of TFIID complex. The interaction between TATA-box and the general transcription factors plays role in the determination of the TSS thus explaining why it is found at a fixed position relative to the TSS [84,85].

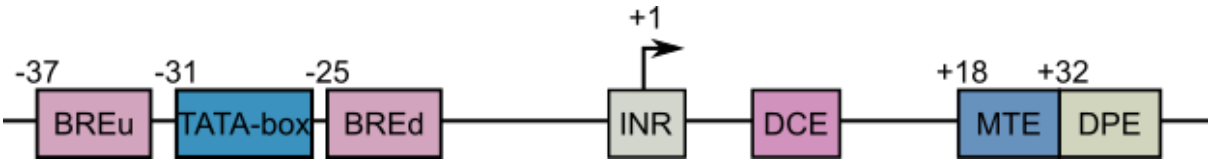

Figure 1.1 | Eukaryotic core promoter elements. Upstream and downstream transcription factor IIB (TFIIB) recognition element (BREu and BREd respectively), Initiator sequence (INR), motif ten element (MTE), downstream promoter element (DPE) together with the TATA-box have fixed position relative to the transcription start site which is denoted by the arrow. The downstream core promoter elements (DCE) does not have a fixed position on the promoter.

The initiator element is another such motif with a fixed position relative to the TSS and directly overlaps the TSS [86]. INR sequence is more abundant compared to the TATA-box and its consensus sequence differ among eukaryotes [82]. In humans the INR motif was formerly defined as a pyrimidine followed by a purine such that the purine is the first nucleotide to be transcribed [87]. For focused human core promoters, higher information content INR motif was has been identified and several nucleotides outside the dinucleotide motif have been suggested to be crucial for transcription initiation at least in vitro studies [88]. In TATA-less promoters the INR motif often occur in combination with the downstream promoter element (DPE) which is located downstream of the TSS [76]. Several TFIID subunits are suggested to bind the DPE and the strict INR-DPE spacing is thought to be important for coordinating TFIID binding [85,89]. Other human core promoter elements are the TFIIB recognition elements (BREs) [74,90] and the downstream core elements (DCEs) [77] which are less abundant. These elements are bound by different general transcription factors in vitro, thus suggesting a potential role in PIC recruitment and assembly [91].

1.2.1.1 Nucleators of the transcriptional pre-initiation complex

The TATA-box and INR are core promoter elements that act as nucleators of the PIC [78]. The recruitment of the PIC to the promoter is considered a rate-limiting step in eukaryotic gene transcription [92]. The TATA-box is generally considered to be a stronger PIC nucleator than INR. That notwithstanding, the majority of human core promoters (~76%) lacks TATA-like elements but rather have high GC content. Only about 24% of human genes contain any form of TATA-like element or high AT-rich sequences in their promoters. Furthermore, only ~10% of all the TATA-box-containing promoters have the canonical TATA-box sequence (TATAWAWR). On the other hand, ~46% of human core promoters contain the consensus INR sequence (YYANWYY) which may occur in combination with a TATA-box. Only ~30% of human core promoters are TATA-less genes which contain the INR sequence. The INR is the simplest functional core promoter element capable of direct transcription initiation in the absence of the TATA-box. As a nucleator, the INR motif enhances the binding of the general transcription factor TFIID [73]. It is noteworthy that ~46% of all human promoters lack both TATA-like and consensus INR elements [93].

Gene ontology (GO) analysis indicates that TATA-less human genes are frequently associated with housekeeping processes while those with a TATA-box are mostly highly regulated by either biotic or stress signals [93]. This is consistent with the notion that promoters of housekeeping genes in vertebrate organisms are often TATA-less and/or associated with CpG islands, while cell type-specific or highly regulated genes often have TATA boxes [94–96]. These findings indicate that genes with different core promoter elements generally control different biological process. In addition to the observations about the human genome, *S. cerevisiae* genomic data indicate that TATA-bearing genes are generally regulated and associated with inducible stress-related response, while TATA-less genes are generally related to housekeeping functions [97].

1.3 Transcription initiation

1.3.1.1 Recruitment of the transcriptional pre-initiation complex

To assemble RNAP II PIC capable of initiating transcription in both mammals [98] and yeast [99] a minimal set of general transcription factors are required: TBP, TFIIE, TFIIF and TFIIH. The TATA-box binding protein normally exist as a subunit of the TFIID transcription factor protein complex but it can independently assemble the PIC. At other promoters DNA recognition is assisted by the TFIIB factor which binds to DNA elements (BREs) upstream or downstream to the TATA-box [74,90].

The TFIIE is known to serve as a loading factor for the TFIIH [100–102]. TFIIE also helps to induce transcription in systems with supercoiled templates for which TFIIH is not required, suggesting that TFIIE helps in promoter melting [103,104]. TFIIF was first identified as RNAP II binding factor suggesting that it facilitates the loading of the RNAP II into the PIC [105]. Further structural and functional studies point to a role of TFIIF in stabilizing TFIIB within the PIC [106,107]. TFIIF modified by casein kinase 2 (CK2) has been observed to still support the formation of transcriptionally competent PIC despite not being stably retained within the PIC [106]. This effect of phosphorylation on TFIIF complicates the interpretation of earlier work. TFIIA is not required to assemble a transcriptionally active PIC, but has been shown to stabilize TBP-DNA

interaction, thus increasing the ratio of TBP-bound templates and generally impacting the interaction between the TFIID complex and the TATA-box [108–111].

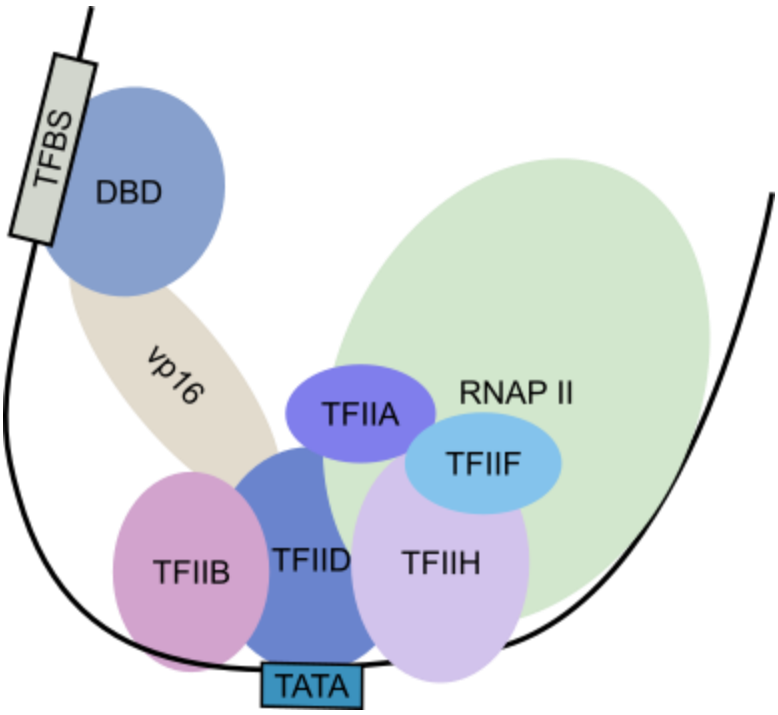

Figure 1.2 | Depiction of transcription pre-initiation complex stabilized by vp16. The residues 436-47 of the H1 region of the VP16 transactivation domain (vp16) fused to a DBD interact with TATA binding protein – a subunit of TFIID to stabilize the PIC for successful initiation of transcription in eukaryotic cells.

1.3.1.2 DNA loop formation

Upon TF binding to the RE, the DNA forms a loop to bring the TF-bound RE and the core promoter into close proximity to allow the TF to interact with and stabilize the PIC [112,113]. The free energy of looping or J-factor (looping efficiency) depends on the length, flexibility and phasing of the DNA [114–117]. It is, therefore, a simplification to say that the efficiency of looping needed for successful PIC stabilization is dependent on the distance between the RE(s) and the core promoter. The J-factor can be considered as the effective concentration of one end of the loop in the vicinity of the other [118], thus providing a quantitative measure of the energetics of DNA loop formation. This parameter has been also used to compare the role of sequence in DNA cyclization and nucleosome positioning [119].

The impact of DNA loop formation on gene expression was nicely demonstrated using the natural IFN-β binding site, which is located immediately upstream of the core promoter of the INF-β gene [120]. It was shown that transcription initiation was abrogated by placing the binding site 560 bp away from the core promoter, a distance at which inherent looping of the DNA was inefficient. However, the DNA looping and transcription could be restored by placing the heterologous Sp1 RERE just upstream of the core promoter, which allowed the interaction between the two TFs, thus re-establishing DNA loop formation. DNA-Sp1 complexes have been shown to self-associate thus supporting the observed looping that brought together the two distant DNA segments [112]. Similarly, the λ repressor could establish interaction between

the core promoter and IFN-β binding site when they were placed several kilobases away from each other – and λ repressor binding sites were placed near both elements. However, when other REs were scattered between the enhancer and the promoter, they served as decoys which led to formation of non-productive loops thus, acting as insulator elements.

Characterization of the 601TA sequence derived from the 601 nucleosome positioning sequence showed that it has a J-factor for cyclization 5-30 folds greater than a random sequence and its looping efficiency was dependent on the DNA phasing [121,122]. Mechanically, the 601TA and the random sequence generally differ in bendability and as such differ in their ability to form loops that facilitate transcription. Despite the difference in the J-factor of 601TA and the random sequences, it was observed that the concentration at which looping is maximal was the same for both sequence but the looping efficiency was higher for 601TA at all concentrations [119]. For example, the J-factors of loops of 94bp long random and 601TA sequences differed by an order of magnitude with the 601TA nucleosome positioning sequence being more flexible.

The sequence dependence of looping efficiency was observed only at 94bp, except when RE was added to the loop sequence in which case there was restoration of a consistent length-independent sequence dependence [119]. The TF impose distinct boundary conditions which has effects on the associated looping efficiency hence yielding multiple looped configurations for the given DNA length. Genomic DNA unlike naked DNA is organized into chromatin by first organizing DNA strands into nucleosomes [123–125], thus making the chromatin state of the DNA important in terms of looping efficiency since nucleosomes can impose structural constraint on genomic DNA.

1.3.1.3 Transcription by RNA polymerase II

Gene transcription by RNAP II is initiated when TF bind near the TSS. These factors can then act indirectly by recruiting other factors that modify the chromatin structure, or directly by interacting with components of the general transcription machinery. Both modes of action by the TFs results in the recruitment of the general transcription factors to the core promoter [126]. RNAP II with unmodified carboxyl-terminal hepta-peptide repeat domain (CTD) is recruited to the core promoter and forms the PIC through interactions with the general transcription factors [127,128]. The unmodified CTD has a high affinity for the Mediator complex [127,129,130]; a transcription co-activator that bridges the interaction between the TFs bound to the REs and the general transcription factors at the core promoter [131]. The strong affinity of unmodified CTD for the Mediator complex is as a result of multiple hydrophobic interactions and hydrogen bonds its subunits form with the Mediator complex [127,129,130].

Since the RNAP II is not capable of driving separation of the template and non-template strands, it relies on the XPB helicase subunit of TFIIH to melt promoters in which there is no super helical tension in the template DNA [132,133]. The TFIIH melts 11-15 base pairs of DNA surrounding the transcription start site [132–134] and the template strand of the promoter is then positioned within the active cleft of RNAP II to form the open complex. Transcription initiation starts with the synthesis of the first phosphodiester bond of RNA. RNAP II is released from the rest of the general transcription factors after synthesizing about 30 bases of RNA to transition from transcription initiation to transcription elongation [17,135]. Promoter escape is facilitated by the phosphorylation of Ser5 and Ser7 residues of the CTD by the general transcription factor TFIIH subunit Cyclin-dependent kinase 7 (CDK7). The phosphorylation lowers the affinity of RNAP II for the Mediator complex by disrupting the hydrogen bonds [127,136–

[138]. Several other residues of the CTD are dynamically phosphorylated and dephosphorylated after the promoter escape throughout the transcription process [139]. The unmodified CTD is mainly implicated in transcription initiation while the post-translationally modified CTD is associated with transcription elongation and pre-mRNA processing [135,140].

1.3.1.3.1 RNA polymerase II pausing

RNAP II pausing occurs after it has escaped the promoter as a result of association with pausing factors DRB-sensitivity-inducing factor (DSIF) [141] and negative elongation factor (NELF) [142].

Transitioning of a paused RNAP II to a productively elongating state requires the activity of positive transcription factor b (P-TEFb) [143–145] which phosphorylates the DSIF-NELF complex causing NELF to dissociate from RNAP II. This transforms DSIF to elongation promoting state [146,147]. P-TEFb is also responsible for phosphorylating other serine residues within the RNAP II CTD which creates the platform for RNA processing factors and chromatin modelling factors to bind in order to aid productive RNA synthesis. Stably paused RNAP II downstream of the promoter markedly reduces transcription re-initiation after promoter escape [148,149].

1.3.1.4 Visualization and quantification of transcription

There are many strategies for mRNA visualization and quantification, but most of them give only a snapshot about localization or abundance of the mRNA of interest. To allow for time course imaging of mRNA in live cells, specific binding of RNA binding proteins has been used to directly follow mRNA expression dynamics. One such method is based on direct mRNA tagging with fluorescent proteins using the MS2 bacteriophage system. The bacteriophage MS2 coat protein (MCP) binds specifically to an RNA hairpin sequence also called MS2 binding site (MBS), which can be cloned into the mRNA of interest [150,151]. Extending the mRNA with multiple repeats of the MBS stem loop and expressing a fluorescent protein fused to MCP enables time course imaging of mRNA expression dynamics and localization in live cells [152].

The phage PP7 RNA hairpin or binding site (PBS) can also be cloned into the mRNA and be detected with the PP7 coat protein (PCP) fused to a fluorescent protein [153]. In combination with the MCP and its MBS, live cell imaging of two mRNA species can be performed simultaneously. Moreover, split fluorescent proteins can be fused to different coat proteins, which results in “background-free” imaging. There is only low background fluorescence in cells except when the split proteins are brought into close proximity when the two coat proteins bind to their cognate hairpin structures [154]. The drawback for using such mRNA visualization systems is that, the size of both repeated hairpin structures and the coat proteins fused to fluorophores can interfere with the normal mRNA function and localization [155,156]. Effort is ongoing to improve the mRNA visualization technique. For example, the MBSV6 reporter system has a hairpin structure that has lower affinity for MCP, which allows for mRNA degradation while preserving detection of single molecules in live cell imaging [157].

That notwithstanding, cloning the MBS and PBS into mRNA should not interfere with normal mRNA transcription kinetics. Subsequently, quantitative visualization with fluorescently tagged MCP or PCP should be indicative of normal transcription kinetics. The level of mRNA transcription or amplitude can be determined from the brightness of nuclear mRNA foci, which

is proportional to the number of nascent mRNAs whose stem loops have been transcribed and bound by the coat protein. It has been demonstrated that, for two genes with equal transcription rates, the foci brightness is proportional to the length of the 3' UTR following MBS or PBS loops transcription [158]. Therefore, for two mRNAs with a 3' UTR of the same length, the difference in the brightness of the mRNA foci should be directly proportional to the transcription rate of the mRNA.

1.3.2 Pre-mRNA processing

For stability and efficient translation of the transcript, mRNAs go through three essential processing steps. (I) Five prime capping of the mRNA – a process requiring three enzymatic activities: RNA triphosphatase, guanylyltransferase and 7-methyltransferase [159]. The RNA triphosphatase acts on the terminal nucleotide to remove the γ-phosphate after RNAP II has transcribed the first 25-30 nucleotides [160]. The second enzyme; the guanylyltransferase then transfers GMP from GTP to form GpppN which is subsequently methylated by the 7-methyltransferase enzyme. Humans have a bifunctional capping enzyme that consists of both N-terminal RNA triphosphate and C-terminal guanylyltransferase activities. The mammalian capping enzyme binds directly to the phosphorylated CTD of the elongating RNAP II through its guanylyltransferase domain [161]. This process couples transcript capping directly to early stage transcription thus stabilizing it. Five prime capping of mRNA protects it from 5'-3' exonucleases in the nucleus and the cytosol [162,163]. (II) The precise removal of the non-coding intron sequences from the pre-mRNA so that full length proteins can be expressed. This reaction is catalysed by the spliceosomes in eukaryotes [160]. (III) The Endonucleolytic cleavage of the transcript at 10-30 nucleotides downstream of a signal (AAUAAA in mammals) which is followed by polyadenylation (poly(A)) at the 3' end [164]. Similar to the 5'-capping, poly(A) tail is essential for the stability and translatability of the mRNA [165]. Transcript cleavage requires the activity of multiple proteins including cleavage/polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF) and two additional cleavage factors (CFIm and CFIIm) in humans. It is the poly(A) polymerase (PAP) that adds the poly(A) tails to the 3'-OH which gets exposed after cleavage [160,164]. The pre-mRNA processes; mRNA capping, splicing and polyadenylation together with mRNA stability govern mRNA abundance post-transcription [166–168], thus influencing gene expression.

1.4 Translation

Translation is one step in the gene expression process in which ribosomes synthesize functional proteins using the information transcribed on mRNA. The initiation of translation is considered a rate limiting step in mRNA translation [169]. Eukaryotic translation begins with the identification of the first codon by the translational PIC. The translation initiation codon is predominately identified by scanning mechanism, although initiation on smaller proportion of mRNAs is mediated by the internal ribosome entry site (IRES). In the scanning mechanism (reviewed in [170]), every triplet in the 5'-end of the mRNA is inspected for complementarity to the anticodon of methionyl initiator transfer RNA (Met-tRNA$_i$). Eukaryotic initiator factors (eIFs); eIF1, eIF1A, eIF3 and eIF5 stimulate the binding of Met-tRNA$_i$ to the small ribosomal

subunit (40S). The Met-tRNAi/40S complex then forms a ternary complex with eIF2-GTP. The resulting translation PIC then binds to the 5'-end of mRNA activated by eIF4F and poly(A)-binding protein. Secondary structures that impede ribosomal attachment and the scanning mechanism are removed by the RNA helicases (reviewed in [171]). The eIF5 stimulates the hydrolysis of eIF2-GTP in the scanning pre-initiation factor – a reaction which is impaired at a non-AUG triplet. The eIF1 and the C-terminal end of the eIF1A must then be displaced from the P site of 40S ribosome to allow for base-pairing between Met-tRNA$_i$ and the AUG codon and the subsequent release of phosphate from eIF2-GDP. The joining of the 60S subunit is mediated by a second GTPase, eIF5B to produce elongation competent 80S initiation complex [170,172,173].

1.4.1.1 Ribosome scanning

The translation PIC scans the five prime untranslated region (5' UTR) using the Met-tRNA$_i$ anticodon to identify the AUG codon. Scanning depends on 40S conformation conducive for processive movement along the mRNA and unwinding of mRNA structures to enable the mRNA to thread through the mRNA-binding cleft of 40S to expose successive codon triplets in the P site [170]. Although ribosomes have bidirectional movement, they exhibit a bias towards 5' to 3' movement during scanning – a process which occurs at about 8 bases/s [174,175]. The scanning process involves a series of forward movements (5' to 3') which are interspersed by short backward movements or 3' to 5' excursions [176]. Increasing the 5' UTR was not seen to affect the translation efficiency in yeast [174] but one can reason that there is a higher chance of secondary structures forming in longer 5' UTRs.

1.4.1.2 5' untranslated region

The length of the 5' UTR and secondary structures therein can govern the probability that a scanning ribosome will find the first translation codon before falling off the mRNA [177,178]. The median length of human 5'UTR is 218 nucleotides [179,180]. Some mRNAs are however leaderless; meaning some mRNAs completely lack the 5' UTR. All mammalian mitochondria mRNA species are leaderless [181] but aside that leaderless mRNA are very rare in higher eukaryotes [182]. Some other mRNAs in humans have very short 5' UTR about 12 nucleotides on the average and undergo scanning-free translation initiation [178]. In contrast, other mRNAs have 5' UTRs that are highly structured [177] and can regulate ribosome entry. Example of a structural element in the 5' UTR is the iron responsive element which affects the translation of a set of mRNAs that are important for the homeostasis of iron [183,184]. Positioning of stable G-quadruplex structures (RG4) close to the 5' cap also contributes to repression of translation in vitro and in vivo [185,186].

1.4.1.2.1 Upstream open reading frame

Structured 5' UTR may also contain upstream open reading frame (uORF). Approximately 50% of mammalian mRNAs harbour at least one uORF (which are mostly less than 30 codons) upstream of the main protein coding open reading frame [187,188] and there are observations that suggest a good proportion of that are translated [189–191]. In less than 50% of the time, ribosome that have translated the uORF resumes scanning and reinitiate at a downstream site [171]. Events at the uORF stop codon post termination may be conventional with the release of 60S

subunits prior to deacylated tRNA, however, some 40S subunits remain on the mRNA to resume scanning. The 40S subunits at this stage are incapable of re-initiation since they lack the eIF2-ternary complex (TC), however this can be acquired during scanning [171]. Short length uORF is essential for rescanning and re-initiation efficiency as they decrease sharply with increasing length of the uORF [192–194]. In addition, it has been shown that efficient initiation at the uORF significantly reduces the expression of the main ORF thus, making the whole mRNA unstable [195]. Since both stable structures in the uORF and the translation efficiency of the uORF affect re-initiation, it can be said that it is not the length per se, but the leakiness of the start codon that leads initiation at the protein coding ORF [195].

1.4.1.3 Ribosome pausing

There are three events that may cause ribosomes to pause after translation initiation. First, ribosomes pause at rare codons as they are forced to wait longer for the corresponding rare tRNA to incorporate the right amino acid. Limitation of some amino acids is sensed by the mTORC1 and GCN2 signalling pathways which regulate global translation accordingly. On the other hand, amino acids whose limitations are not efficiently sensed by the mTORC1 and GCN2 pathways, can also regulate translation through ribosome pausing. For example the loss of arginine charging of tRNA during its limitation has been observed to cause ribosome pausing at two out of six codon [196]. Second, cis-elements within the mRNA affect ribosome translation rates since the presence of secondary structure temporarily stalls ribosomes while they attempt to unwind them to be able to continue translation. Lastly, ribosomes may be stalled due to nascent peptide interaction with the ribosome within the peptide channel [197]. Generally, ribosome pausing causes a decrease in protein production and premature termination of translation but does not cause reduction in mRNA abundance [196].

1.5 Transcription factor dynamics

Most TFs bind to the REs in the promoters of multiple genes and thus regulate the expression of more than one target gene [198,199]. The genes under the regulation of the same TF can exhibit differential gene expression. Some target genes may be fully expressed while others will show different levels of expression or may not be expressed at all, depending on the external stimulus or developmental stage of the cell. [198,200,201]. This difference in the expression of target genes, regulated by the same TF, can lead to different cellular responses or cell states [198]. Such behaviour has been shown to hold true for many master TFs [198,200,201]. But how does the cell achieve this? It has been observed that the differential expression of target genes is dependent on the amplitude and the frequency of nuclear import and export/degradation of TFs (TF dynamics) [198]. TF dynamics is thought to encode the type and intensity of the different external stimuli [202,203].

1.5.1 Instances of transcription factor dynamics in eukaryotic cells

TF dynamics is defined in this dissertation as the pattern of TF concentration in the nucleus over time. Persistent presence of TF in the nucleus would be referred to as sustained dynamics while intermittent nuclear accumulation, separated by nuclear export, will be termed pulsatile dynamics. Earlier observation of this phenomenon was made in the extracellular signal-regulated kinase (ERK) pathway. It was observed that two growth factor, the nerve growth factor (NGF) and the epidermal growth factor (EGF), triggered a different cell fate in neuronal precursors in rat. EGF causes cell proliferation while NGF leads to cell differentiation. At first glance one might have assumed that separate signalling pathways were induced by each of the stimuli thus leading to different cellular outcome. Further studies revealed that the ERK signalling is activated by both stimuli but with different dynamics [204–206]. This suggested that differentiation or proliferation were not necessarily stimulus specific but rather regulated by ERK dynamics. The sustained activation of ERK induced by NGF leads to differentiation, while transient ERK activity triggered by EGF permits proliferation [203]. These differences in the response to EGF and NGF achieved via the dynamics of ERK are a consequence of differences in the identity and connectivity of various pathway components.

Studies of signalling molecules have identified additional molecules that encode upstream signals or stimuli into their dynamics. Different inflammatory stimuli have been shown to induce different dynamics of nuclear factor-κB (NF-κB) TF. NF-κB activation by tumour necrosis factor-a (TNFα) results in prolonged nuclear accumulation and a subsequent transcription of the negative regulator or NF-κB inhibitor alpha (IKBα). The negative feedback loop generates oscillations of transcriptionally active NF-κB [207–210]. Bacterial lipopolysaccharide (LPS) on the other hand triggers slow nuclear accumulation that leads to a single sustained NF-κB activity pattern [211–213]. It has been shown that NF-κB oscillations (pulses) triggered by TNFα induce expression of inflammatory response genes while sustained NF-κB nuclear accumulation triggered by LPS causes, in addition to the expression of inflammatory response genes, the expression of genes associated with adaptive immune response and cytokines secretion [211].

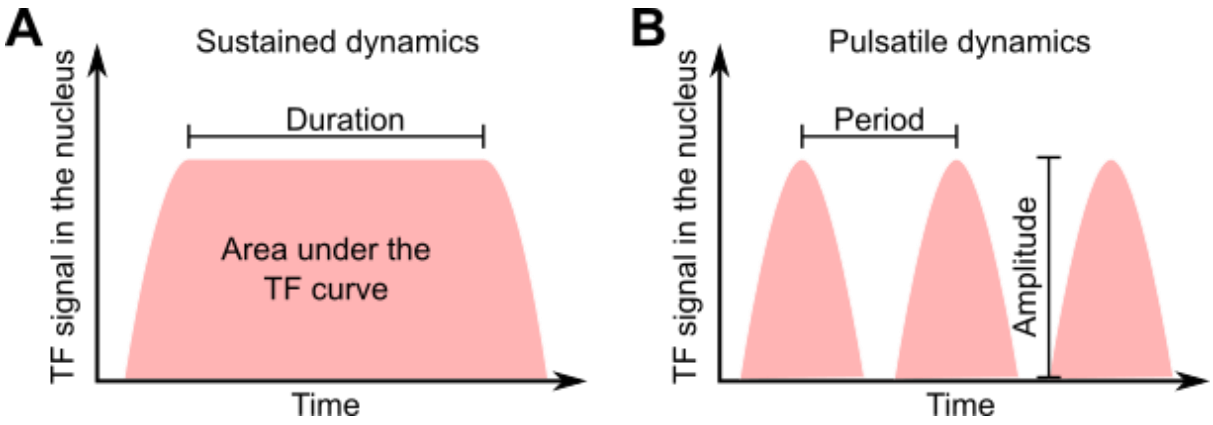

Figure 1. 3 | Illustration of transcription factor dynamics. Measurable features of TF dynamics known to encode information about the stimulus. Period and duration of a pulsatile dynamics define the frequency of pulses. Amplitude describes the peak of nuclear accumulation whereas area under the curve describes the cumulative nuclear TF accumulation.

Mechanistically, transient activation of NF-κB by TNFα is mediated by IkBα in a negative feedback loop. TNF receptor activation leads to the activation of the IkB kinase complex which by phosphorylating IkB triggers its degradation via ubiquitination. Free NF-κB induces the expression of its target genes including IkB leading to a subsequent inhibition of NF-κB. In addition to IkBα, A20 is a NF-κB target gene that controls the long term NF-κB dynamics in response to persistent TNFα. A20 protein acts upstream of the pathway and compared to IkBα has a longer half-life, which explains why the long term phase of NF-κB oscillations are dampened [198,214,215]. In contrast, sustained NF-κB activation triggered by LPS is a result of positive feedback achieved through an autocrine pathway involving the de novo production of TNFα. LPS activation of the Toll-like receptor 4 (TLR4) induces the expression of TNFα and thus the activation of the TNF receptor. The delay between the TLR4 and the TNF dependent NF-κB activation has been suggested to give rise to the sustained NF-κB activation induced by LPS [213]. It must also be noted that NF-κB activity is dose dependent; increasing TNFα concentration for example leads to shorter delays between NF-κB pulses while increasing frequency of TNFα stimulation causes pulses with smaller amplitude [207,216,217].

The dynamics of the tumour suppressor p53 are also associated with different cellular outcomes. DNA double strand breaks caused by γ-radiation triggers pulsatile p53 dynamics which are associated with transient cell-cycle arrest [218]. The pulses have a fixed amplitude and duration but the number of pulses is however, γ-radiation dose dependent [219,220]. In contrast, UV radiation elicits a single sustained nuclear p53 accumulation which leads to the apoptosis of cells [218]. Sustained p53 activation, triggered by UV, has a dose dependent amplitude and duration [221]. The different dynamics of p53 in response to γ- and UV radiations result from specific feedbacks in the DNA damage response network [221]. P13 kinase-related kinases, which are upstream of the network, relay the DNA damage signal to p53. The activity of p53 in turn activates two core negative-feedback loops: the first involving the E3 ubiquitin ligase Mdm2 and p53 and the second being between the phosphatase Wip1 and p53. The difference lies in the latter: the response to γ-radiation has an additional negative feedback between p53 and ATM which is facilitated by Wip1 [198].

Other studies in *Saccharomyces cerevisiae* has revealed that the stress response master TF Msn2 shows transient increase in nuclear accumulation in response to glucose limitation or high osmolarity. The duration of nuclear Msn2 increase exhibit dose-dependence response with a fixed amplitude. It has also been observed, in single cell studies, that the initial pulse induced by glucose limitation and osmotic stress is followed by a series of Msn2 pulses. The frequency of the pulses is dependent on the intensity of glucose limitation and not affected by intensity of osmotic stress. On the other hand, oxidative stress induces sustained nuclear Msn2 accumulation which has amplitude that is H_2O_2 dose dependent [202].

The observations above suggest that dynamics of the same signalling molecule can encode the identity and quantity of different upstream stimuli, which, when decoded leads to different cellular responses. The obvious scientific question is: what are the molecular bases for detecting features of TF dynamics and the subsequent decoding of the information carried in TF dynamics?

1.5.2 Promoter decoding of transcription factor dynamics

Deciphering how TF dynamics are decoded to yield differential cellular response has been challenging. However, some clues can be drawn from the differential activation of JNK, NF-κB and NFAT in response to calcium concentration dynamics. JNK and NF-κB have low affinity for calcium and, since they require high calcium concentration for their activity, they respond to strong transient calcium bursts. This, in combination with their slow degradation rate, makes the downstream factors stay elevated after just a brief stimulation with calcium. NFAT, on the other hand, has a high affinity for calcium and high degradation rate. As such, low but sustained calcium levels will activate NFAT over JNK and NF-κB [222].

Molecular schemes with similar mechanism have been proposed in the decoding of Msn2 TF dynamics. Differences in TF binding affinity together with the kinetics of the promoter have been suggested to regulate the cellular response to different Msn2 dynamics [198,202]. Through modelling and experimentations, promoters of Msn2 target genes have been classified according to how they process Msn2 dynamics. This revealed that multiple and distinct gene expression programs could be encoded into the dynamics of Msn2. It was observed that activation of the promoters of target genes was nucleosome occupancy dependent and that combination of pulsatile TF dynamics and slow promoter activation leads to high expression noise. Nucleosome positioning between the RE(s) and the core promoter modulated how fast the promoter responds to TF binding [201]. These findings implied a trade-off between information transfer and gene expression noise. The noise limits how much information can be encoded in the dynamics of a single TF and reliable decoding of that information by promoters of target genes [201].
Cis-regulatory motifs in variants of SIP18, a target gene of Msn2, have been shown to affect the activation time and amplitude threshold of the promoter. The number of Msn2 binding sites, their accessibility and their distance from the TATA-box were used to explain 90% of variance in amplitude threshold and activation time of target gene promoters [223]. Furthermore, the promoter activation time and amplitude threshold could be decoupled by modulating the number of Msn2 binding sites. Moreover, promoter types could as well be switched. The amplitude threshold was dependent on the number and accessibility of Msn2 binding sites whereas the activation time of promoters was dependent on the distance of the binding sites from the TATA-box. It has also been shown in studies using optogenetic-like *Neurospora* GATA-type TF White Collar Complex (WCC) that a TF that regulates transcription burst frequency induces strong transcriptional response in promoters with weak TF binding. It was further observed that differential regulation of WCC target genes was as a result of different transcription initiation rates [200]. The transcription initiation rates were in turn tuned by both core promoter and distance (in bp) between WCC binding site and the core promoter. These observations were in accordance with the observations made in Msn2 from *S. cerevisiae*.

In mammalian cells, TF dynamics and DNA binding activity control both the specificity and level of target gene expression. Mammalian TFs are however, very highly regulated through post translation modifications which tune their specificity and binding affinities. For instance, UV activation causes sustained p53 nuclear accumulation in mammalian cells but also induces the phosphorylation of Ser46 which directs p53 to promoters of pro-apoptotic genes [224]. In contrast, lys320 acetylation induced by γ-radiation activation favours p53 binding to cell cycle arrest genes [225]. Co-factors of p53 such as ASPP proteins and BRN3B also target p53 to pro-

apoptotic genes thus, favouring their binding by p53. In contrast, iASPP, Hzf and BRN3A inhibit this form of p53 targeting [226–231]. These findings complicate the explanation of differential expression of p53 target genes. Alteration of chromatin remodelling using HAT inhibitors has shown that p53 binding is affected by chromatin structures [232]. It has also been demonstrated that, by dictating the differential RNA polymerase II recruitment and transcription initiation in a p53 fashion, core promoter elements play a critical role in how target genes respond to p53 dynamics [233].

The core promoter of p21 for example, directs rapid TATA box dependent assembly of RNAP II preinitiation complex but permits few rounds of RNAP II re-initiation, whereas PIC formation is very inefficient at the core promoter of Fas/APO1 but permits multiple rounds of transcription initiation [234]. This suggests two intrinsic DNA encoded properties or parameters of p53 induced transcriptional response. First, the kinetics of the gene induction and second, the duration of expression resulting from frequent re-initiation. These are fundamental properties of diverse core promoters in the programming of p53 dynamics and target gene response [234]. It can therefore be said that p53 dynamics does not directly control cell fate. However, coupling the dynamics with the binding affinities of target genes and the core promoters confers an appropriate time window for target gene expression. It becomes apparent that thresholding in response to p53 dynamic amplitude, refractory period in response to duration modulation and the filtering of frequency modulation are essential for differential gene expression in response to p53 dynamics.

Target activation is heterogeneous at the single cell level with fewer cells responding at lower TF doses. Cells can however encode analogue parameters such as TF peak intensity, response time, and number of pulses to modulate the expression outcome [235]. Computational simulation of ERK activation of target genes revealed that genes with promoters that are highly sensitive to ERK could be induced even at low ERK activity. Highly sensitive promoters are thus expected to respond to constant stimuli. In contrast, promoters with intermediate or low ERK sensitivity would be induced only at high ERK activity [158]. Dynamics of NFκB nuclear accumulation and DNA binding activity control both the specificity and the expression level of target genes [210]. Combination of quantitative experimentation and mathematical modelling showed that some genes with a long mRNA half-lives allow for effective decoding of NFκB dynamics but were not sufficient to account for all target genes. Chromatin processes such as slow rate of transition between inactive and NFκB bound enhancer state could also help decode NFκB dynamics [236].

1.5.2.1 Differential translation and stability of transcripts impact protein abundance

Efficiency of mRNA translation has been shown to affect the abundance of both mRNA and protein post transcription. The 5'-UTR element at position -3 to -1 of the start codon; the kozak sequence (KZ) together with mRNA secondary structures, and out of frame upstream AUGs have been shown to have strong effects on translation initiation [237]. Transcripts with weak KZ are less sensitive to reduction in global elongation rate but more sensitive to global initiation rates compared to transcripts with strong KZ [238]. Sequences immediately upstream of the kozak sequence are also crucial in protein synthesis: single or multiple point mutations in the upstream of the KZ highly affect protein levels [239]. Translation efficiency study in single living cells showed that translation efficiency also decreased dramatically as 5' UTR hairpin stability was increased from ΔG = 25 to 35 kcal/mol. Moreover, shifting this hairpin to the 5' cap was

able to negatively modulate translation more than 50 fold. At constant distance between the 5' cap and the hairpin, increasing GC content also decreased translation efficiency [240]. Similarly, unstructured A-rich element in the 5' UTR destabilized mRNAs in the absence of translation despite enabling cap-independent translation [195].

In studies of p53 dynamics, short mRNA half-lives have been shown to produce pulses of gene expression. Transcriptional profiling in single cells demonstrated that expression of a subset of p53 target genes was coordinated across time within single cells but this coordinated expression was weakened by p53 pulses [241]. A group of genes found in NFκB studies were observed to encode highly unstable mRNA transcripts. Reducing the transcription of these genes caused a rapid decrease in the amount of the corresponding mRNA transcripts, since their transcription was suppressed by IkBα and A20. Such instability of the mRNA transcripts is intrinsic to the target genes and therefore, is NFκB dynamics dependent [242]. The observed instability was associated with AU-rich sequences which are known to destabilize mRNA. In agreement with this observation, in yeast, most genes expressed under transient stress are highly unstable whereas genes expressed under slow persistent stress were mostly stable [243]. Efficiency of mRNA translation and mRNA stability have not been directly implicated in differential expression of target genes in response to TF dynamics so far or used to explain the phenomenon. But could mRNA translation efficiency impact differential expression of target genes?

1.5.3 Light-control of transcription factor dynamics

Different strategies have been used to generate different dynamics in various studies of transcription factors dynamics. Earlier studies of ERK, NFκB and p53 dynamics in mammalian cells have used different stimuli such as radiations and hormones to induced different TF dynamics as described in section 0. Chemical agents have also been used to induce TF dynamics. For instance, Nutlin-3 together with γ-radiation has been used to generate sustained p53 dynamics. Nutlin-3 is a small molecule that binds to Mdm2 – the p53 inhibitor and in effect inhibits p53 protein degradation. Inhibition of Mdm2 after γ-radiation switched the organic p53 pulses into sustained p53 signals [218]. Similarly, a chemical agent was used to control nuclear localization of Msn2 in *S. cerevisiae* [202]. This strategy involved introduction of analog-sensitive mutation into the catalytic isoforms of protein kinase A (PKA), hence enabling the selective and reversible inhibition of PKA activity and the subsequent control of Msn2 localization with the small molecule 1-NM-PP1 [201,202,244].

The caveat for using such external or chemical agents for the generation of TF dynamics is the modulation of molecules higher upstream of the signalling pathway of the transcription factor of interest. In addition to generating different TF dynamics, the different stimuli or modulated molecules might affect other components in the pathway or even trigger completely different pathways, which may then contribute to the observed downstream cellular response. This leads to a plethora of effects which may be hard to decipher thus not readily explainable. PKA for example, controls many transcriptional regulators and Msn2 is just one of them. As a result, modulating its activity with 1-NM-PP1 may lead to Msn2 target gene expression changes that may not be caused only by Msn2 dynamics but rather a combination of multiple expression regulation by other TFs/pathways that are sensitive to PKA activity. Sustained

NFκB activity triggered by LPS also suggest that more regulation mechanisms may be at play during LPS activation of cells compared to activation with TNFα.

To quantitatively and causally elucidate the relationship between TF dynamics and transcription of target genes, a fast method that specifically and reversibly localizes a TF of interest to the nucleus is required. Inducible nuclear accumulation is considered a good strategy since many natural TFs behave similarly. Speed and reversibility are essential for the generation of high resolution nuclear accumulation and export in response to an external stimulus. Modular optogenetic strategies are well suited for this purpose. Although light can be toxic under high intensities, under low intensities it does not regulate nor interfere with expression programs in mammalian cells and so can serve as a good alternative for imposing TF dynamics. Opto-SOS system for example has enabled precise regulation of ERK dynamics and helped to study how different target genes specifically respond to different dynamics [158,245]. It must however, be pointed out that Opto-SOS acts upstream of ERK pathway. Other strategies have been implemented to achieve shuttling of POI in and out of the nucleus [246,247]. CLASP is one of such optogenetic strategies that has been used to directly control the nuclear localization of Crz1 to study how that impacts the expression of its cognate promoters [248].

Our lab developed a light-inducible nuclear localization system called LINuS, which allows reversibly translocating a protein of interest from the cytosol into nucleus upon blue light exposure [247]. This system consists of the second light oxygen voltage (LOV2) domain from *Avena sativa* phototrophin I, which has a C-terminal helix (the Jα helix) that docks to the core domain in the dark and undergoes a reversible conformational change when exposed to blue light. A nuclear localization signal (NLS) fused to a truncated Jα helix is inaccessible in the dark due to the interaction between the helix and the core domain (Figure 1.4), but accessible to importins when the Jα helix unfolds and undocks from the core domain after blue light exposure. The dynamic nature of the caging and passive diffusion causes a significant nuclear accumulation of the protein of interest (POI) fused to LINuS in the dark. A nuclear export signal (NES) was therefore added to the fusion protein to retain POI in the cytosol during the dark phase. Hence, allowing for sufficient nuclear accumulation only when the strong NLS is exposed upon blue light activation [247].

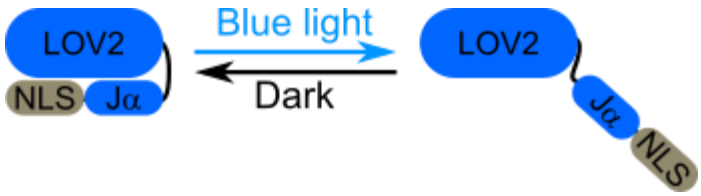

Figure 1.4 | Light-inducible nuclear localization signal (LINuS). Schematics showing the fusion of the nuclear localization signal (NLS) to the Jα helix of the AsLOV2 domain. The Jα helix docks to the core of LOV2 in the dark causing the NLS to be caged. Blue light induces a conformational change of the Jα helix which leads to the uncaging of the NLS.

1.6 Transcription factor architecture

1.6.1 Yes-associated protein (YAP) as a co-activator

The highly conserved mammalian Hippo pathway controls organ size and suppresses tumorigenesis by promoting apoptosis and blocking cell proliferation. The Yes-associated protein (YAP) and the WW domain-containing TF 1 (WWTR1 or TAZ) are co-activators of the TEA domain (TEAD) family of proteins. TEAD proteins interact with YAP/TAZ to form a transcriptionally active heterodimeric TF that then regulates the expression of cell proliferation and pro-survival genes. YAP lacks a DBD but contains a TAD. YAP and TAZ are master transcription regulators for a variety of physiological functions including cell volume control, glucose uptake, cell proliferation, migration and metabolism. Dysregulation of YAP/TAZ, therefore, has a significant effect on cell fate. YAP and TAZ are thought to be redundant in their function and are regulated by the Hippo pathway in a similar manner. Their development, physiological and structural differences however, suggest differences in their regulation and downstream functions. In fact, YAP inactivation has a greater effect on cellular processes, e.g. glucose uptake, cell spreading and migration, than TAZ inactivation [249].

The Hippo pathway inhibits the TEAD/YAP complex transcription factor by phosphorylating YAP and thereby promoting its cytosolic sequestration and degradation [250]. The two key mammalian hippo pathway core kinases STE20-like protein kinase 1/2 (MST1/2) and the Salvador family WW domain containing protein 1 (SAV1), interact with each other to activate the tumour suppressor kinase 1/2 (LATS1/2) and MOB kinase activator 1 (MOB1) large complex. LATS1/2-MOB1 complex then phosphorylates YAP as a downstream effector on ser127 and/or TAZ on ser89 which leads to the sequestering of YAP/TAZ in the cytosol. Cytosolic sequestering is facilitated by interaction between the phosphorylated YAP/TAZ and the cytosolic 14-3-3 proteins. Once in the cytosol, it is subjected to degradation via the ubiquitin-proteasome pathway [251–254]. In the absence of active LATS1/2 proteins, YAP/TAZ is activated by Rho GTPases. Activated YAP/TAZ translocate into the nucleus where it binds to members of the TEA domain family of DNA binding proteins.
YAP/TAZ co-activators have many domains and motifs including the WW domains, the TEA domain family member transcription factor binding domain, the SH3-bindiing domain, the coiled-coil domain, transactivation domain and the PDZ binding motifs [255–257]. NMR studies of the TEAD binding domain of YAP have shown that it is natively unfolded and that binding to TEAD induces localized conformational changes in the YAP protein. The short segment of YAP that adopts the extended conformation forms extensive contacts with the rigid surface of TEAD. Functional assays have thus, defined an extensive conserved surface of TEAD2 YAP binding domain for effective binding of YAP [250].

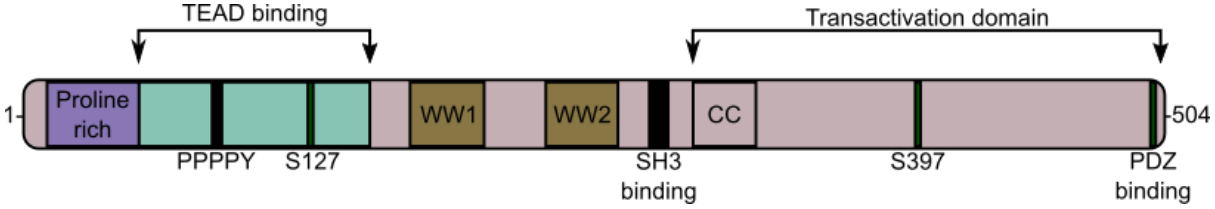

Figure 1.5 | Structure of Yes-associated protein. The scheme depicts key structural domains and phosphorylation sites that are crucial for regulating YAP activity and interactions with other binding partners.

The nuclear phosphoprotein p73 can also engage in a physical association with YAP. The association is facilitated by the WW domain of YAP and the PPPPY motif of p73. The terminal tyrosine of the PPPPY motif of p73 was shown to be essential for this association with YAP – an observation which is expected for ligands that interact with the group I WW domains [258]. Other transcription factors such as SMAD, RUNX2, TBX5 and ERBB4 have been shown to interact with YAP/TAZ in the nucleus [259,260]. However, it is the association with TEAD that is the most studied. The interaction between TEAD and YAP/TAZ stimulates cell proliferation, survival, migration and anti-apoptosis related genes such as CTCG, CYR61 and ITGB2 [254,261]. YAP/TAZ studies in cultured cells have revealed that culture conditions affect the cytoplasmic or nuclear YAP/TAZ localization. Moreover, the Hippo pathway and its activity are impacted by cell density and shape as well as mechanical tension and biochemical signals [259,260]. For example, stiff culture surfaces such as Collagen I promote nuclear accumulation of YAP/TAZ, whereas cytosolic sequestering of YAP/TAZ is maintained on soft surfaces [262–264]. YAP/TAZ also accumulate more in the cytosol under high cell density than under low cell density [265]. F-actin polymerization and G-actin depolymerisation serve as regulators of the Hippo pathway [266]. Inhibitors of Rho-GTPases and myosin light chain kinase prevent nuclear accumulation of YAP/TAZ [267,268]. Hippo pathway can be activated by diverse processes and factors which implies that YAP/TAZ localization can also be regulated by the cytoskeletal matrix or even cell culture conditions [269].

1.6.2 TEA domain (TEAD) transcription factors

Four TEAD genes (coding for so-called TEA/ATTS domain transcription factors) are expressed in mammals. TEAD proteins are widely expressed but expression of each member of the TEAD protein family is tissue-specific [270–272]. The mammalian TEAD1-4 proteins possess a DBD, but not a TAD. The DBD of the TEAD proteins is composed of a helix-turn-helix homeodomain fold [271] and dictates the DNA binding specificity of TEAD proteins. TEAD was first identified as a binder to the GT-IIc and SpH enhancers of the Simian Virus 40. The TEA/ATTS domain of the TEAD family is highly conserved in all members and directs binding to the MCAT (5'-GGAATG-3') sequence [271,273,274]. With the development of ChIP-seq, the definition of TEAD DNA binding sites has been expanded to include other sequences other than the MCAT sequence [275,276]. Despite the increase in our knowledge of TEAD-DNA binding, the MCAT sequence remains the predominant binding partner of TEAD [275]. This is not surprising in that, TEAD1-4 binds strongly to MCAT sequence with a K_D ranging from 16-38nM [277]. Cooperative binding of TFs has been shown to improve binding specificity [278]. There are multiple TEAD binding sites in promoters of TEAD/YAP target genes such as CTGF[279], suggesting that cooperativity may be relevant or perhaps necessary for the binding of TEAD to its responsive element. In fact, many groups have observed cooperativity in DNA binding by TEAD proteins in the absence of co-factors. Structure-function studies have identified that the formation of L1 loop upon TEAD1 binding is key for the cooperative binding to tandem MCAT DNA sequences [252,271,280].

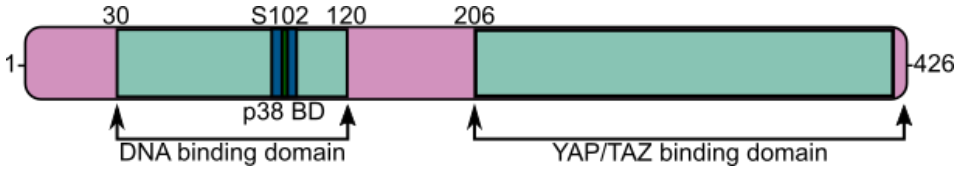

Figure 1.6 | Structural details of TEA domain protein 1. A diagram showing the essential binding domains and phosphorylation site that modulates DNA and p38 binding.

The C-terminal domain of TEAD is equally conserved, particularly the residues involved in the interaction with the co-activator proteins [281,282]. The crystal structure of the YAP binding domain of human TEAD2 showed that it adopts an immunoglobulin-like β-sandwich fold with two extra helix-turn-helix inserts [250].
TEAD1-4 proteins are the main transcription factors for the YAP/TAZ transcription co-activators of the Hippo pathway. YAP/TAZ remain the most well studied co-activators and regulators of the TEAD transcriptional activity. However, there are several other transcriptional co-activators that have been identified to interact with TEAD. The Vestigial-like (VGLL) protein family which has four members (VGLL1-4) has been shown to interact with TEAD to regulate expression of TEAD target genes [283,284]. Members of the VGLL family of proteins bind to regions of TEAD that overlap with the YAP/TAZ binding domain, and hence, can compete with the YAP/TAZ co-activators for TEAD binding [285,286]. VGLL4 binding to TEAD for instance inhibits TEAD binding by YAP/TAZ and as a consequence inhibits expression of TEAD target genes. In effect, VGLL4 suppresses cell growth thus, serving as a tumour suppressor [285–287].

TEAD4 has also been shown to directly interact with transcription factor 4 (TCF4) through its TEAD domain to facilitate the transactivation by TCF4 which then lead to the expression of Wnt target genes. Similar to TEAD/YAP, VGLL4 binding to TEAD4 inhibits the interaction between TEAD and TCF4, hence, inhibiting TEAD4/TCF4-induced target gene expression. However, VGLL4 does not compete with TCF4 for TEAD binding. Similarly, activator protein 1 (AP-1) has been shown to interact directly with TEAD [288,289]. AP-1 occupies the same chromatin sites as TEAD and the presence of AP-1 has been shown to be essential in inducing expression of target genes involved in tumour growth and progression [288,289]. Other proteins shown to interact with TEAD include the poly ADP-ribose polymerase (PARP) [275], serum response factor (SRF) [290], myocyte enhancer factor 2 (MEF2) [291], and myc-associated factor X (MAX) [292].

1.7 Mathematical modelling

A mathematical model of a biological system is a simplified representation of a biological system in the form of equations that serves to organize the information available about the biological process, to eventually understand and quantify it [293]. Mathematical models of dynamical biological systems consider changes in the molecules over time and can be categorized into deterministic or stochastic models. In a deterministic model the change in reaction speeds depend on the concentration of the molecules (variables) and the parameters of the model. In stochastic models, velocities of reactions are also dependent on randomness in the system. Randomness arises from the uncertainty that exists in systems that contain rate limiting elements. A mathematical model also allows us to study the effect of perturbations in components and/or environmental conditions on the behaviour of the biological system, thus

allowing for control and optimization. In other words, mechanistic mathematical modelling helps to quantitatively analyse and temporally resolve dynamic biological processes to be able to make predictions that can be experimentally verified [294].

Modelling approaches (reviewed in [295]) can be based on differential equations, Bayesian equations, stochastic modelling, agent-based modelling and other new approaches. Development of a mathematical model is an approximation process since simplifications are introduced. Simplifications must however be logical in chemical or physical terms and mathematically valid. Mathematical formulation of biological systems involves definition of key variables and functional expression of their relationship with other variables in the system. Dependent variables are elements that change over time according to the state of the biological system while independent variables are elements that can be controlled during the experiment. Parameters of a mathematical model set the internal and external constraints on the system. Numerical values of parameters are determined from experimental data from perturbation experiments [296,297] or from prior biological knowledge [298]. In combination with fluxes gathered from experimentations, mathematical models also integrate kinetic data and other information available on the elements of the system.

Parameter estimation is a recurring processing in the model building process and deals with finding the numerical values of parameters that best describes the given biological system from experimental data [299–303]. The experimental data must come from variables that represent a given time point and its evolution over time. The quality of the model should then be tested via some numerical quality assessment process and should include the evaluation of the steady states stability and robustness of the model [304,305]. Quality assessment must also evaluate features that describe the transient response to perturbations [305]. Quality assessment points out problems of consistency and reliability of the mathematical model and constitutes a cycle of model refinement which improves upon the initial model [306–309]. The improved conceptual model can then suggest further experiments which will further refine model; thus enriching our understanding of the biological system [310,311311].

1.7.1 Mathematical modelling of transcription factor dynamics

Mathematical modelling has been used to help better understand the role of TF dynamics in regulating the expression of target genes. Studies of ERK activity dynamics modelled the process with four equations that represented ERK nuclear import, mRNA transcription of dusp, DUSP protein expression and transcription of ERK-dependent immediate early gene. This model assessed the impact of ERK inhibition by DUSP on expression of IEGs [158]. Similarly, a mathematical model of IkB/ NF-κB signalling was developed based on biochemical data and validated using real-time single cell fluorescence imaging [209]. The model consisted of association and dissociation of protein complexes, synthesis of IkB proteins, IKK-dependent and independent IkB degradation and nuclear localization of IkB and NF-κB proteins [214].

Such models however did not explicitly define promoter states and how these states affected gene expression. Two promoters states have been used to model expression of reporter genes in *S. cerevisiae* [312–315]. However three promoter states have been used to account for slow response rate and delayed expression seen under some target genes of Msn2 TF [201]. The three promoter states were only defined as phenomenological variables rather than

biochemically defined promoter states. The three states were specified as unbound, bound and active [201]. Similarly, modelling of transcriptional burst cycles of gene promoters in mammalian and *Neurospora* cells, suggest that some promoters contain a promoter state which is refractory to TF activation [316–318]. The simplest mathematical model that accounts for these observations has been shown to be the three-state promoter model although other promoter states with a considerable lifetime may exist [200,319]. The three promoter states in this case were defined as unbound, bound/active and refractory promoter states. That notwithstanding there may be some mammalian gene promoters that can be described with a two-state promoter model while others will strictly need the three-state or more promoter models to be able to account for differences in transcriptional dynamics [248].

1.8 Aims of the study

Transcription factor dynamics have been shown to dictate differential expression of target genes by encoding the identity and intensity of a stimulus perceived by the cell. How mammalian promoters decode the information carried by TF dynamics is still poorly understood. I aim to assess the role played by different promoter elements in decoding TF dynamics using a minimal model system represented by well-defined synthetic TFs and promoters.

The synthetic TF should be easily translocated in and out of the nucleus in a short time frame preferably within minutes. The stimulus for inducing TF nuclear translocation should not interfere with endogenous processes so that the reporter output will not be confounded by other cellular processes. The reporter library shall be based on the same reporter gene to minimize differences across experiments due to parameters such as the DNA sequence of the coding region or translation efficiency, mRNA and protein stability etc. This will allow me to focus only on the role of defined promoter elements in dictating the gene expression output. For better understanding of the role they play in decoding TF dynamics, the selected promoter elements should be well characterized. After establishing the reporter library, I will impose different TF dynamics and analyse the resulting reporter expression, both at the mRNA and protein levels. The obtained data will then be incorporated into a mathematical model to gain a better mechanistic understanding of the system and to make predictions that shall be experimentally validated.

My second aim is to assess the difference between a TF that contains DB and TA domains (referred to as synTF) and one that has only the DBD, but needs to bind a co-regulator bearing the TAD. This last type of TF resembles the natural TEAD/YAP transcription factor system, thus I will refer to it as synTEAD/synYAP. I will use this system to induce expression of some selected reporters from my first reporter library to finally compare reporter gene expression induced by synTF and synTEAD/synYAP.

2 RESULTS

This chapter is divided into two sections. In section 2.1, I describe the experimental and computational results obtained with the synthetic light-responsive TF bearing both DBD and TAD (synTF). In section 2.2, I present the experimental results obtained with the synthetic light-responsive transcription system based on the DBD-bearing TF (synTEAD) which is localized to the nucleus and the light-responsive TAD-bearing co-activator (synYAP).

2.1 Decoding TF dynamics by promoters of target genes

2.1.1 Design and construction of synTF

A LexA based circuit was selected for this study since LexA is well-studied as a master DNA damage response TF which regulates expression of many target genes. LexA binds to the operator of its target genes with varying binding affinities and the K_D of many binding sequences have been determined in vitro. The availability of different binding sequence affinities and that fact that the circuit will be isolated in the mammalian system makes it very ideal for studying effect of TF dynamics and how it is decoded at the promoter without having to consider the role played by other pathways that may not be the focus of the study. I therefore started the construction of the synthetic TF by fusing the DBD of *E. coli* repressor protein LexA [320] to the TAD of the herpes simplex virus type 1 transcription factor VP16 [321]. Residues 436-47 of the H1 region of VP16 TAD were used in this study (henceforth referred to as vp16).

The TF was made light-responsive by fusing it to LINuS [247]. I tested several versions of LINuS consisting of different combinations of NES and NLS. TF1 was constructed by fusing synthetic TF LINuS with biNLS2 [247] and PKIt NES while TF2 was fused to LINuS with biNLS10 [247] and IkBα NES (Table 6.1). The mCherry fluorescent protein was added for visualization and tracking (Figure 2.1.1A). The uncaging of the NLS upon blue light (~ 480 nm) activation leads to increased accumulation of the synthetic TF in the nucleus. The nuclear import is reversible: when blue light is removed, the NLS docks back to the core of the AsLOV2 domain, leading to the effective export of the synthetic TF to the cytosol due to the added NES. Since there is always an equilibrium between the docked/folded state and the undocked/unfolded state for the Jα helix within AsLOV2 [322], even in the dark there is a small fraction of undocked helices (circa 1-2% of the entire population). This, together with potential passive diffusion of the ~80kDa fusion protein into the nucleus, accounts for the non-negligible amount of nuclear accumulation of TF1 constructed with PKIt NES and biNL2 in the dark, but showed high nuclear accumulation upon blue light activation (Figure 2.1.1B).

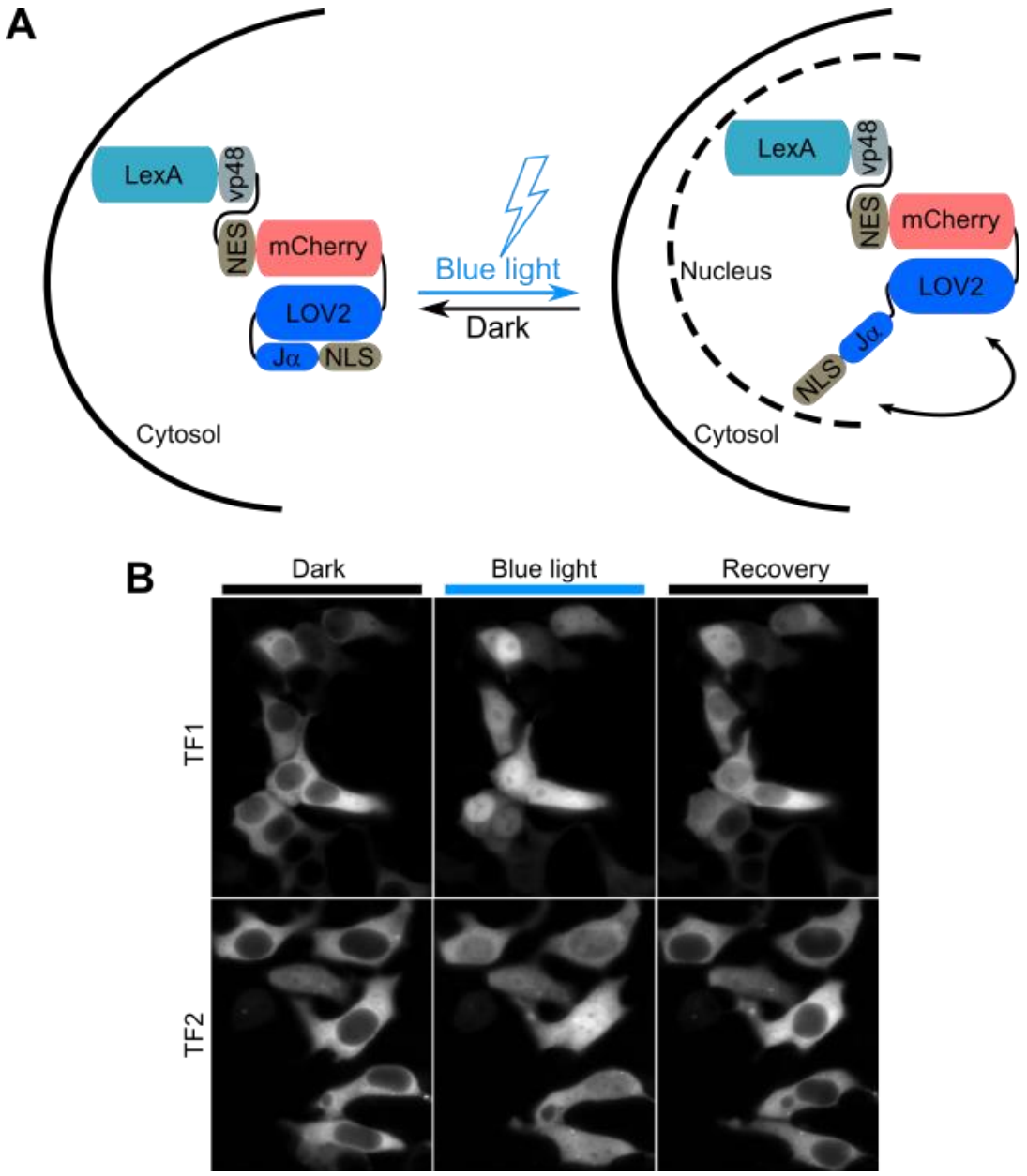

Figure 2.1.1 | Design and characterization of the synthetic TF. A) Scheme of synthetic TFs and mechanism of nuclear import. Synthetic TFs were composed of either the full-length bacterial repressor protein LexA or only its DBD fused to repeats of residues 436-47 of the VP16 TAD (vp16), a nuclear export signal (NES), mCherry for tracking and LINuS for light responsiveness. B) Nucleocytoplasmic translocation of synthetic TF TF1 and TF2 both of which consist of DBD and three repeats of vp16. TF1 has PKIt NES and biNLS2 based LINuS while TF2 has IkBα NES and biNLS10 based LINuS. Images were taken before (Dark) and right after 30 min (Blue light) of blue light activation. Images were taken again after 30 min incubation in the dark (recovery).

LOVTRAP

TF1 showed higher nuclear accumulation in dark compare to TF2 (Figure 2.1.1B). TF1 also showed higher nuclear accumulation upon blue light activation. I reasoned that TF1 would be a better TF candidate for this study if I could reduce the nuclear background while retaining the property of high nuclear accumulation upon blue light exposure. I therefore implemented the LOVTRAP technique [323] to sequester TF1 from the nucleus in the dark. The LOVTAP consists of a peptide that binds strongly to the LOV2 domain in the dark and unbinds upon blue light exposure (when the LOV2 domain is in an open conformation). The version of LOVTRAP used in this study consisted of the Zdk2 peptide which binds strongly to LOV2 with a modified Jα helix in the dark (LOV2 in closed conformation). Zdk2 was fused to a subcellular targeting sequence. Consequently, TF1 was localized to the compartment to which Zdk2 was targeted due to the interaction between Zdk2 and LOV2 domain in TF1. Without interaction between Zdk2 and the LOV2 domain under blue light, TF1 was expected to be released from the LOVTRAP so it could translocate to the nucleus. I however observed that nuclear accumulation of TF1 under blue light was much lower in the present of the LOVTRAP (Figure 2.1.2). The best TF1/LOVTRAP set up in terms of lower nuclear levels in the dark consisted of Zdk2 fused to NES (LOVtrap) (Figure 2.1.2B) but had nuclear accumulation similar to TF2 upon blue light activation. Since the LOVtrap/TF1 did not perform better than TF2 in terms of accumulation in the dark and upon blue light activation, I selected TF2 for further testing.

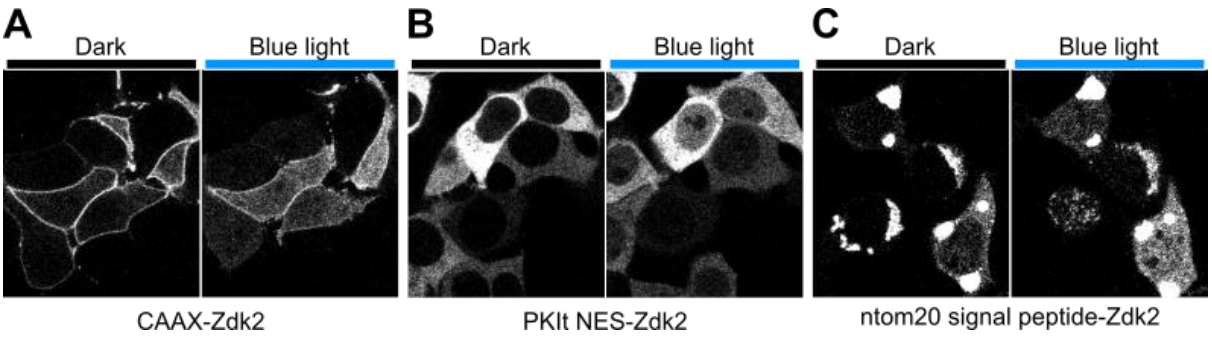

Figure 2.1.2 | Sequestering TF1 from the nucleus using LOVTRAP. A) The CAAX motif was used to target Zdk2 to the plasma membrane. The interaction between Zdk2 and the LOV2 domain of TF1 led to the sequestration of TF1 to the plasma membrane in the dark. B) PKIt NES was used to localize the Zdk2 to the cytosol. In this case TF1 was better retained in the cytosol with less nuclear accumulation in the dark. C) Mitochondrion targeting signal of ntom20 was used to target Zdk2 to the outer membrane of the mitochondria. TF1 was therefore localized to the outer membrane of the mitochondria in the dark. In all setups, only a fraction of TF1 was released after 30 min of blue light activation (blue light) which led to lower levels of TF1 being translocated to the nucleus of the cells when compared to TF1 without LOVtrap (Figure 2.1.1B).

TF2 and even TF1 with/without LOVTRAP induced very low expression of reporter under promoter p2. To improve reporter expression, I increased the promoter dwelling time of the synthetic TF by increasing its binding to the responsive elements. I therefore constructed synTF by replacing the LexA DBD with the full length LexA repressor protein amplified from TOP10 *E. coli* strain. The synTF was able to induce better expression of the reporter under promoter p2. I then tried to see if I could impose different nuclear TF dynamics synTF. Constant blue light activation led to a sustained nuclear accumulation of synTF that reached a peak after ~15-20 minutes of activation (Figure 2.1.3B). I obtained different nuclear

accumulation amplitudes by the intensity the blue light (Figure 2.1.3B-D). Due to the fast reversibility of LINuS, I was able to impose high frequency pulsatile activation of synTF by exposing cells to repeated cycles of 15 minutes blue light – 'ON', and 15 minutes dark – 'OFF' – which I call "15-15 pulses" (Figure 2.1.3C). By extending the dark phase I could generate pulses of lower frequency. Specifically, I used a protocol based on repeated cycles of 15 minutes blue light and 30 minutes dark, which gave rise to pulsatile dynamics I call "15-30 pulses" (Figure 2.1.3D). I observed that the time required to export synTF out of the nucleus depended on the level of the nuclear signal. Thus, more nuclear signal required longer time to completely revert to the initial levels and vice versa.

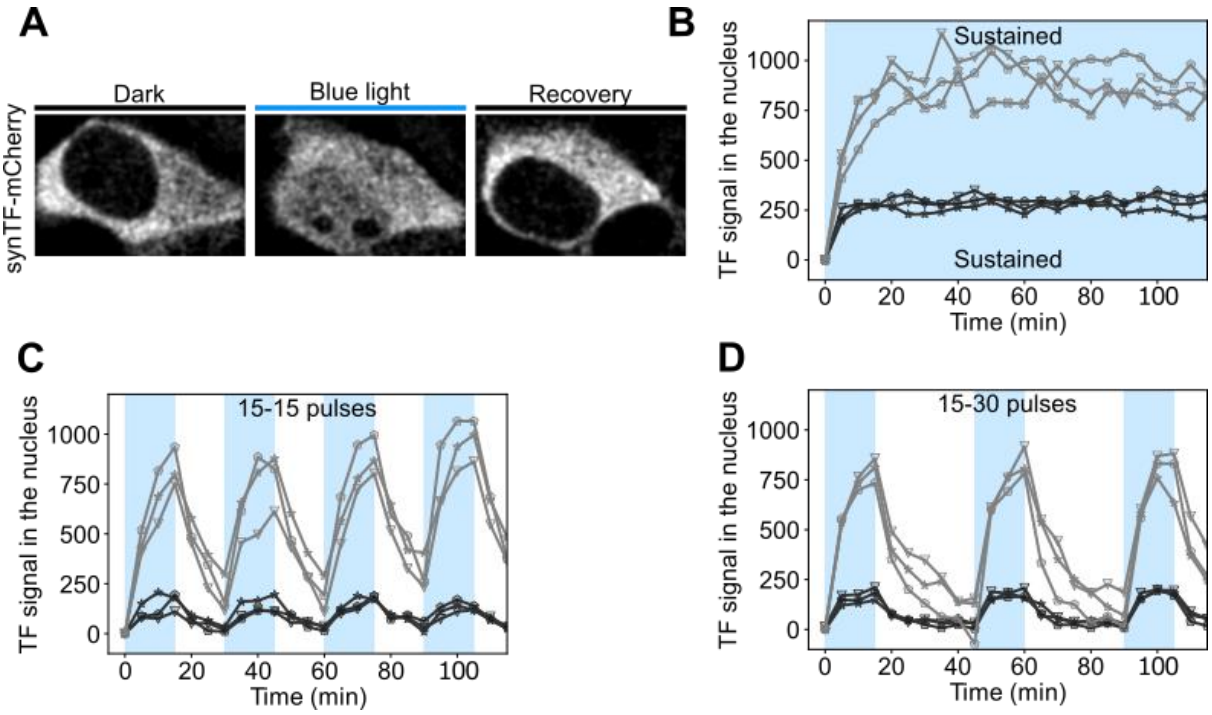

Figure 2.1.3 | Characterization of the synthetic transcription factor (synTF). A) Representative microscopy images showing nuclear accumulation of synTF upon blue light activation and recovery in the dark. B) Quantification of synTF in the nucleus of cells exposed to constant blue light. In this case, the nuclear signal of synTF is sustained. C) Quantification of synTF in the nucleus of cells exposed to 15 minutes blue light ('ON' phase) and 15 minutes darkness ('OFF' phase). In this case, synTF enters the nucleus in pulses. D) Quantification of synTF in the nucleus of cells exposed to 15 minutes blue light ('ON' phase) and 30 minutes darkness ('OFF' phase). Here, synTF enters the nucleus in pulses, but with lower frequency than those shown in C. Grey and black lines indicate cells illuminated with high and low intensities of blue light respectively. The shaded light blue regions indicate the 'ON' phases and the unshaded regions correspond to the 'OFF' phases.

2.1.2 Construction of the synthetic promoter library

I thought that a synthetic reporter would be ideal for this study since the synthetic circuit could be considered as an isolated system being free from endogenous regulation in the mammalian cells in which the circuit is studied. I wanted to eliminate any crosstalk with endogenous pathways, which would complicate the interpretation of the data from the differential expression of target genes under different synTF dynamics [236]. I selected to focus on the LexA

responsive element and two core promoter elements which can nucleate the transcription PIC: the TATA-box and the INR.

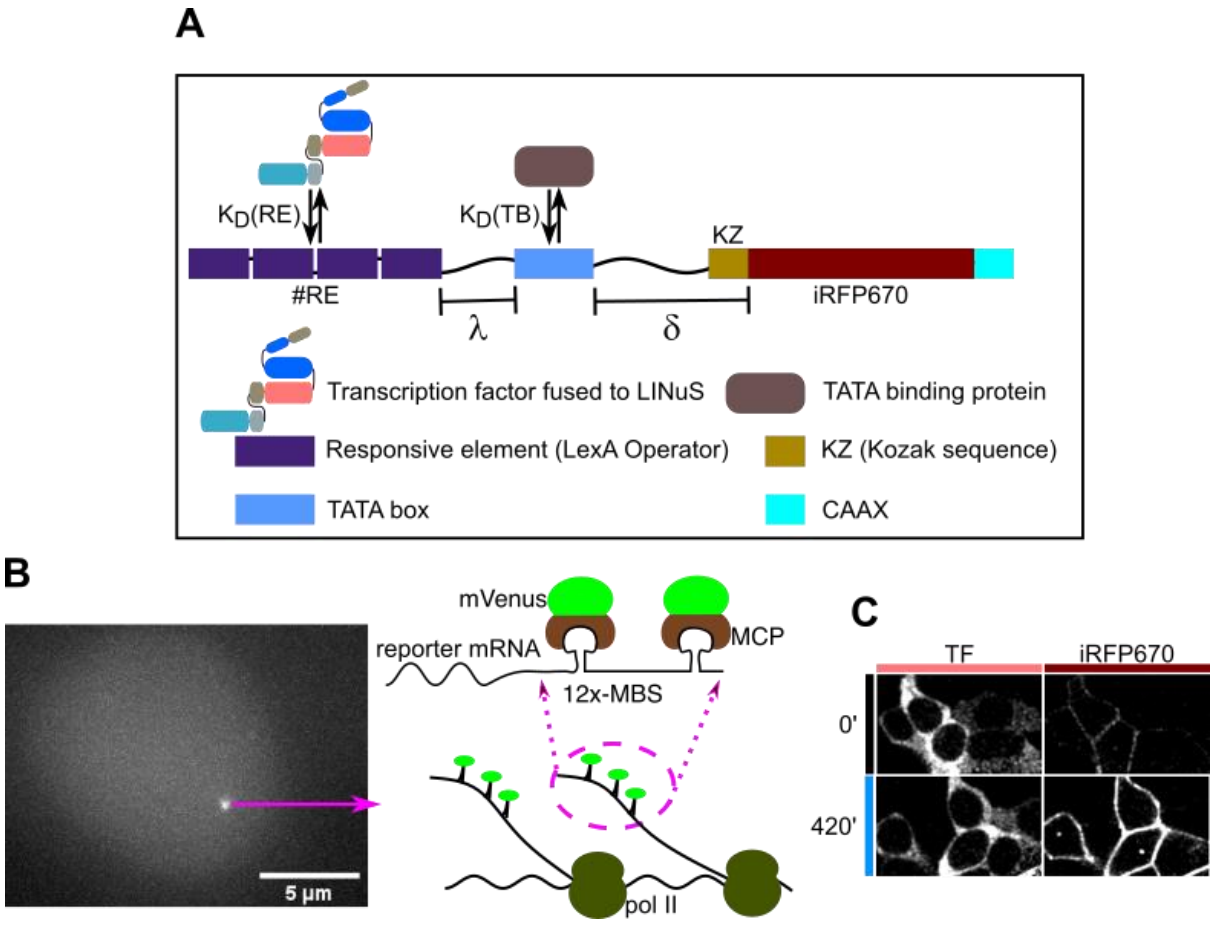

Figure 2.1.4 | Design and construction of reporter library and quantification of reporter expression. A) Scheme of a reporter showing the promoter elements and parameters used in generating the reporter library. Responsive element (RE) refers to LexA binding site that confers synTF responsiveness to the reporter whereas TATA-box serves as the core promoter. The nearest RE is placed λ base pairs upstream of the TATA-box which is itself placed δ base pairs upstream of the translation start site (ATG) of the iRFP670 reporter gene. SynTF binds to RE with affinity K_D(RE) and TBP binds to the TATA-box with affinity K_D(TB). The 5' UTR of the reporter transcript may or may not contain a kozak sequence (KZ). The CAAX motif was used to localize the reporter protein to plasma membrane. B) Visualization of reporter expression at the transcription level. Representative microscopy image of nascent mRNA transcription focus (left) and scheme showing the details of the focus (right). The 12x MS2 loops transcribed into the reporter mRNA are bound by mVenus-fused MS2 coat protein. A transcription focus consist of multiple nascent mRNAs being transcribed simultaneously. C) Representative confocal microscopy images of synTF and iRFP670 reporter protein expression before and after blue light activation. 0' and 420' represent images taken before blue light activation and 5 hours after 2 hrs of constant blue light activation respectively.

The reporter library was constructed by considering two main processes: (I) synTF binding to the cognate responsive elements, which is characterized by the affinity of LexA for its binding sites (the REs); and (II) the formation of the PIC, which is characterized by the affinity of the TATA-box binding protein for the TATA-box (TB) (Figure 2.1.4A). I varied the distance between the REs and the TB (defined as λ) to assess how it affects the coupling of synTF

binding and transcription initiation. To study the effect of translation initiation efficiency on TF dynamics decoding I used different sequences around the Kozak sequence. I also varied the distance between the TB and the translation start site (defined as δ) as well as the sequence of this DNA stretch to investigate the influence of the 5' UTR in the process. Considering that transcription starts approximately 30-31 bp from the start of the TB, varying δ basically affects the length of the 5' UTR of the corresponding transcript. I selected the far-red fluorescent protein iRFP670 as a reporter, since it can be imaged without activating the AsLOV2 domain. Since I expected low reporter expression under some promoters, I thought to localize the reporter to the plasma membrane for easier detection. To this aim, I fused to it the CAAX motif [324,325]. Since I kept the reporter sequence itself constant, I did not evaluate other factors such as transcription elongation rate and variability in mRNA stability that arise from different gene sequences. Expression of the reporter under the promoter bearing the initiator sequence without a TATA-box was not detected under the microscope and so was discontinued. Since the TATA-box is the core promoter element most often associated with genes that are highly regulated, I presumed that understanding how TATA-box containing promoters decode TF dynamics would be informative for many naturally occurring promoters.

The reporter library therefore consisted of promoters with different combinations of the above-mentioned promoter elements (Table 6.2). I combined responsive element sequences bound by LexA with different affinities (Table 1.1) with TB sequences bound by TBP with different affinities [326].

2.1.2.1 Quantifying the level of reporter transcription

To directly correlate TF activity with reporter gene expression on a shorter timescale than that achievable via quantification of reporter protein levels, I quantified also mRNA transcript levels. I turned to an established technique for in vivo mRNA quantification using live cell microscopy based on the interaction between RNA loops derived from the bacteriophage MS2 and its coat protein (MCP) [150,151]. I inserted twelve repeats of the MCP binding sequence (MBS) into the reporter gene after the stop codon (in the 3' UTR). Nascent mRNAs are visualized as foci in the nucleus using MCP fused to mVenus (Figure 2.1.4Figure 2.1.4B). The foci contain several transcribing nascent mRNAs. For better nascent mRNA foci visualization, I removed the terminator sequence from the reporter to achieve long transcripts that remain bound for a longer time to the DNA after transcription of the loops. This optimization step was important since I could hardly detect foci when I used shorter transcripts, as they got released shortly after transcription of the loops.

2.1.3 One-plasmid system

My first system consisted of two plasmids, one for the expression of synTF under the strong CMV promoter and one carrying the synthetic promoter responsive to synTF driving expression of the reporter gene. I realized that transient transfection of these two plasmids in mammalian cells resulted in population of cells that had different ratios of the two plasmids. This made it difficult to objectively compare experiments, since plasmids picked up by cells are not directly quantifiable. Consequently, I could not tell if cells with high synTF levels but low reporter expression accounted for an actual low activity of synTF or rather were a

consequence of a low number of reporter plasmids. To eliminate this dilemma, I decided to clone both, synTF and the synthetic promoter/reporter construct (synPromoter) on a single plasmid (referred to as synPlasmid). Transfection of HEK293T cells with such plasmid ensured the presence of the reporter in all cells showing a signal for the TF (Figure 2.1.6). Low expression of reporter in a cell with high expression of synTF would therefore mean that the promoter does not get activated in these cells.

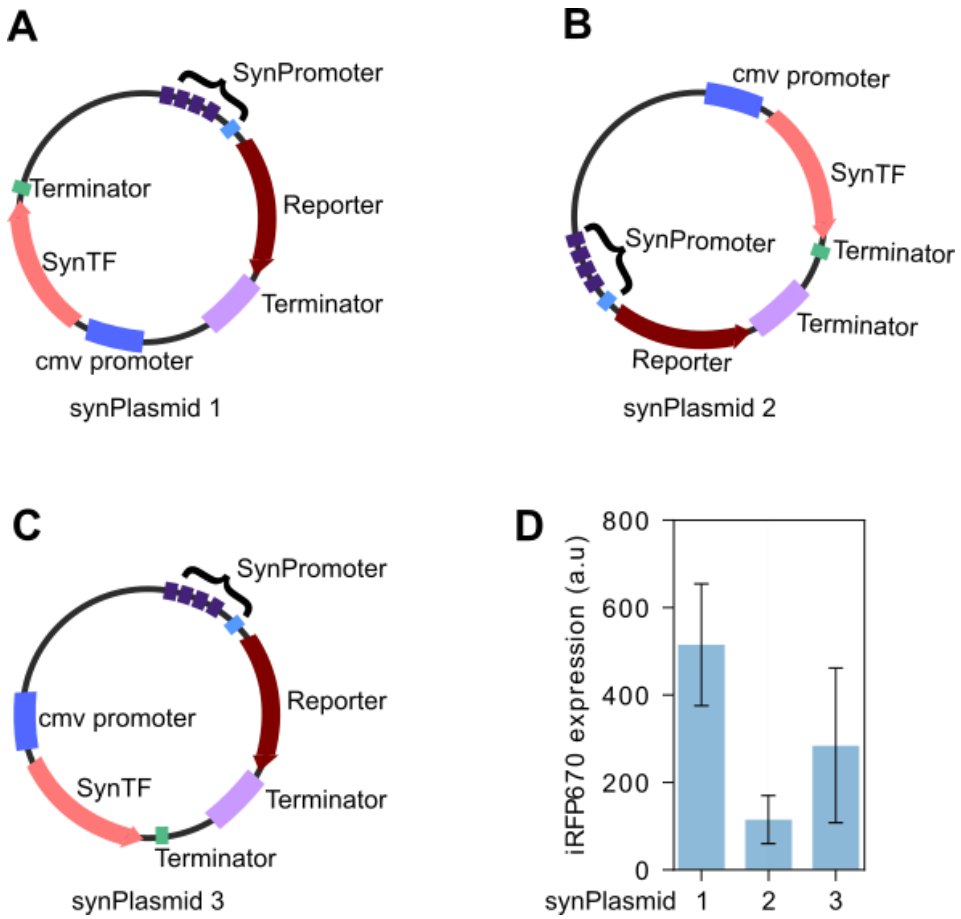

Figure 2.1.5 | Reporter expression under different plasmid arrangements. A) Tandem arrangement of reporter and synTF constructs on a single plasmid. The 3' end of the reporter was ~0.4kb upstream of the first base of the constitutive CMV promoter that drives expression of synTF while the 3' end of synTF was ~3.4kb from the 5' end of synPromoter. B) Convergent arrangement of synTF and reporter with the reporter being on the reverse strand. The 3' end of the reporter was ~0.75kb from the 3' end of synTF while the 5' ends of their promoters were ~3kb apart. C) Convergent arrangement of synTF and reporter with synTF being on the reverse strand. The 3' ends of synTF and the reporter were ~0.72kb away from each other while the 5' ends of their promoters were ~3kb apart. D) Expression of reporter under the p2 promoter 5 hours after 2 hours of constant blue light activation for the different plasmid arrangements. Error bars represent 95% confidence interval of the mean iRFP670 expression in at least 20 cells.

I arranged the synTF and synPromoter constructs in a "tandem configuration" (same orientation) on the plasmid with the 3' end of the reporter being ~0.4 kb upstream of the CMV promoter driving synTF. The 5' end of synPromoter was located ~3.5 kb from the 3' end of synTF (Figure 2.1.5A). This arrangement ensured that the constitutive expression of synTF will not directly cause the expression of the reporter gene. I also tested a "convergent"

arrangement by retaining synTF on the forward strand while placing the synPromoter on the reverse strand, so that they faced each other (Figure 2.1.5B). The CMV promoter and the synPromoter were ~3 kb apart to avoid any expression of reporter as a result of a bidirectional activity of CMV. Lastly, I tested another convergent arrangement in which the synPromoter was on the forward strand and synTF was placed on the reverse strand (Figure 2.1.5C). In this case, their 3' ends were ~0.72 kb apart, while the distance between their promoters was kept at ~ 3kb. These 3 arrangements yielded different reporter expression levels, with the tandem arrangement showing the highest expression and the convergent arrangement with the reporter on the reverse strand showing the lowest expression (Figure 2.1.5D). The convergent arrangement with synTF on the reverse strand was the noisiest among the three. All members of the reporter library tested in this study were therefore cloned together with synTF on single plasmid using the tandem arrangement.

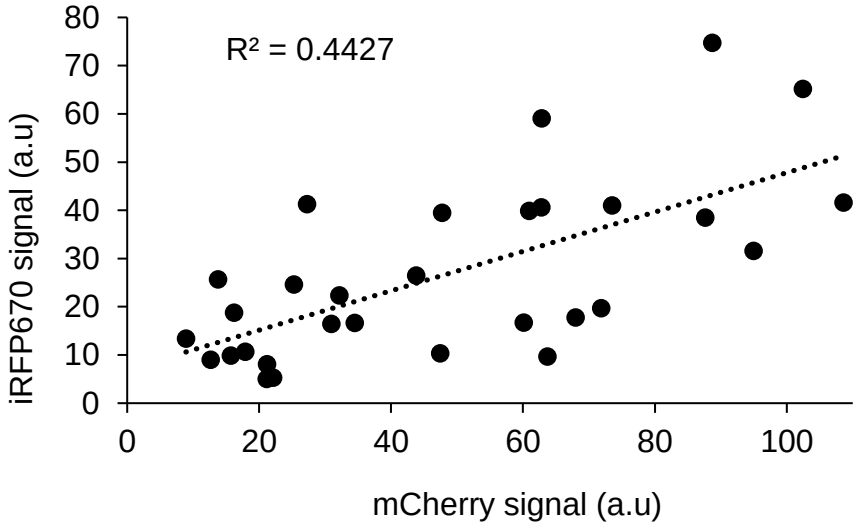

Figure 2.1.6 | Expression of reporter on synPlasmid1 shows a weak dependence on synTF levels. Reporter expression (iRFP670 signal) under promoter p2 on synPlasmid1 shows weak dependence on synTF levels (mCherry). The dotted trend line indicates the positive correlation between the reporter expression and synTF levels. Black dots represent data from single cell measurements.

2.1.4 Reporter expression induced by different TADs

TADs taken from different transcriptional activators induce transcription by recruiting different factors. Whereas some interact directly with subunits of the general transcription factors [50,51], other recruit chromatin remodelling proteins [327–330]. To test the impact of different TADs and their strength on transcription initiation, I tested the following TADs: (I) residues 436-47 of the H1 subdomain of the VP16 transactivation domain (vp16); (II) the first TAD of p53, and (III) the first TAD of Fox1Mc. A single copy of vp16 and Fox1Mc TAD1 induced very little transcription of the reporter under promoter p2 and in very few cells, while p53 TAD1 did not induce any observable expression of the same reporter. Interestingly, vp16 was able to induce a higher reporter expression when the reporter had a second TB (Figure 2.1.7, promoter p26), suggesting the distance between the TB and the TF REs may be essential for transcription initiation. Three repeats of both vp16 (vp48) and Fox1Mc (3xFox1Mc) were then tested for

their ability to induce expression of reporter under promoter p2 which has a single TATA-box (Figure 2.1.7). Reporter expression by vp48 was significantly improved, while that of 3xFox1Mc remained extremely low. I therefore selected vp48 as the TAD for studying the synTF dynamics since having higher reporter expression levels would help distinguish smaller differences between conditions. Our final working transcription factor therefore consisted of the full LexA protein fused to vp48 (Figure 2.1.1A).

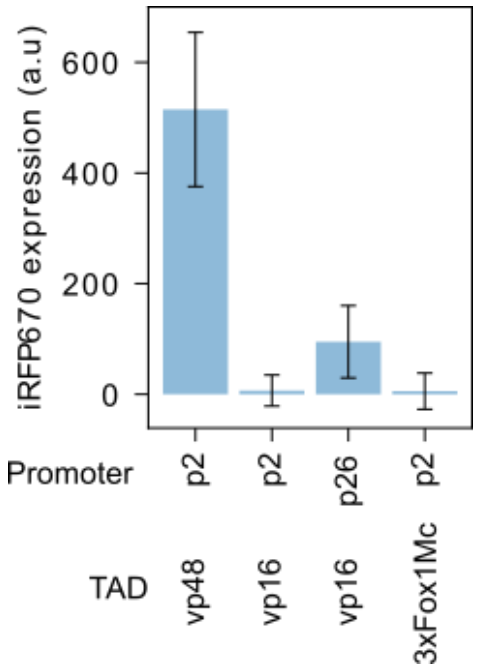

Figure 2.1.7 | Reporter expression is differentially induced by different transactivation domains. Expression of reporters under promoters p2 and p26 were induced by synthetic TF bearing vp48 (3x vp16), vp16 and 3x Fox1Mc TADs. Cells were activated with 2 hours of constant blue light which generated sustained accumulation of the TFS in the nucleus of activated cells. Data shown above were acquired 5 hours post the blue light activation. The error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

2.1.5 Experimental setup and data quantification

HEK293T cells transfected with synPlasmid1 (consisting of synTF and a member of the reporter library) were illuminated with different regimes of blue light to impose different nuclear accumulation dynamics of synTF. I used a hybrid epifluorescence/confocal microscope that allowed me to perform blue light illumination of whole cell populations in epifluorescence modality and acquire confocal images of the reporter and the TF. A very important aspect of this project was to quantify a high number of cells (in the hundreds) because of the higher variability that comes with transient transfection of the tens of reporter genes I was studying. High number of measurement helps to observe trends in such noisy data. The synTF and the reporter protein levels were therefore quantified using an automated software that was explicitly developed for this project (Figure 2.1.8A).

Fluorescent proteins that can used in combination with mCherry and iRFP670 without significant bleed-through of fluorescence and unwanted activation of the AsLOV2 domain are

limited. Consequently, the nuclei of cells were not labelled in the study. A neuronal network capable of instant segmentation of the cell nucleus [331] was built out of necessity for data quantification. Since transient transfection inevitably leads to a high variability of expression levels in the population, I clustered the data obtained from the automated quantification of the images into bins based on the nuclear synTF expression level (mCherry signal) at t=0. This allowed me to fairly compare cells focusing on the relation between the different synTF dynamics and promoter activation. I moreover stratified cells based on the nuclear accumulation of synTF upon blue light illumination. The transient transfection set up has the advantage that the impact of nuclear synTF amplitude on reporter expression can be assessed without the need for new experiments with different blue light intensities.

I decided to compare three types of dynamics: sustained signal, pulses with 30 min period and pulses with 45 min period (Figure 2.1.8B). Eight cycles of 15 min ON and 15/30 min OFF were used to achieve the same area under the curve obtained with 2 hours of sustained blue light activation. All activations were followed by a 5-hours resting period to allow for the maturation of the iRFP670 reporter protein [332,333]. I first tested enhanced YFP (eYFP) in place of iRFP670, but there was significant photo bleaching due to the blue light used for LINuS activation, thus I moved on to IRFP670.

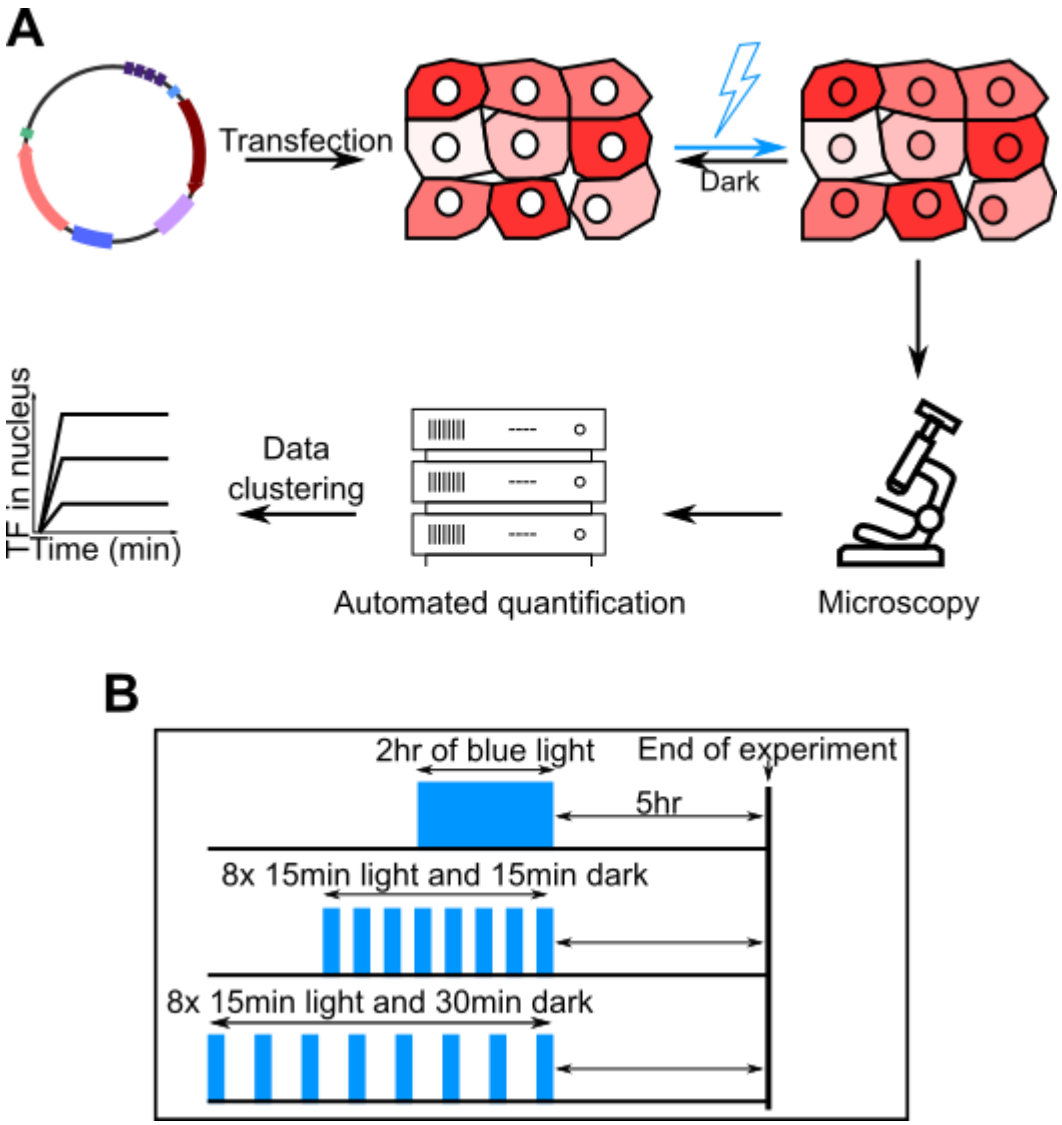

Figure 2.1.8 | Experimental setup for studying transcription factor dynamics. A) Experimental design and data analysis. HEK293T cells were transiently transfection with synPlasmid1 constructs bearing synTF and the reporter of interest. After 24hrs of transfection, cells were illuminated with different blue light activation regimes to impose different TF dynamics and time course reporter protein and synTF images acquired. Images were then automatically quantified and data analysed. B) Blue light activation regimes used to impose the different synTF dynamics. Eight repeats of 15 minutes pulsatile activation were used to ensure equal amount of blue light exposure and the subsequent equal area under the curve of synTF accumulated in the nucleus of cells. Experiments continued 5 hours post blue light activation to allow the iRFP670 protein to mature.

2.1.6 A promoter with high affinity REs and TATA-box cannot distinguish between sustained and pulsatile TF dynamics

Promoter p1 consists of four responsive elements with K_D of 1.61nM [47], a TATA-box with K_D of 2nM [326], 49 bp between the last RE and the TB (λ) and 69 bp between the TB and the translation start site (δ) (Figure 2.1.9A). This promoter induced a strong reporter protein expression when cells were subjected to the sustained synTF signal (Figure 2.1.9B). The average volume of the foci representing nascent mRNAs was observed to increase steadily from the beginning to the end of the two hours activation phase, remaining fairly stable one hour after blue light was switched off (Figure 2.1.9B, middle). The reporter protein was expressed at higher levels when cells were subjected to the 15-15 pulses. Nascent mRNA levels at each time point under 15-15 pulses were similar to those obtained with the sustained signal with no sign of transcription shutdown during the dark phases (Figure 2.1.9C, middle). An even higher protein expression was observed for cells activated with the lower frequency 15-30 pulses (Figure 2.1.9D, bottom). Similar to what was seen for the 15-15 pulses, nascent mRNA levels were similar to those obtained with sustained activation with no signs of expression shutdown during the 30 minutes dark phase (Figure 2.1.9D, middle). Reporter transcripts were only visualized during the blue light activation phase to avoid activating cells during the dark incubation times. I expected that the first visualization time point before every activation phase would indicate if transcription was shut down during the dark phase. However, the first measurements after the dark incubation phases showed no signs of a decrease in mean mRNA expression compared to the rest of the measurements taken during blue light activation. As mentioned above, pulsatile dynamics seemed to correlate with higher reporter protein levels than sustained dynamics; however, this is likely due to the fact that the experiment takes longer in this case (7, 9 and 11 hours for sustained, 15-15 and 15-30 min pulsatile activations, respectively). This reasoning comes from the fact that reporter expression under low amplitude sustained synTF dynamics led to similar protein expression as 15-30 pulsatile dynamics (Figure 6.3). The difference in mean protein expression at the later time points for the different dynamics was only significant when constant activation was compared to 15-30 pulses (Figure 2.1.9E). Furthermore, when I compared protein expression levels for the same dynamics but with two different synTF amplitudes, I found a significant difference only for constant activation (Figure 2.1.9F).

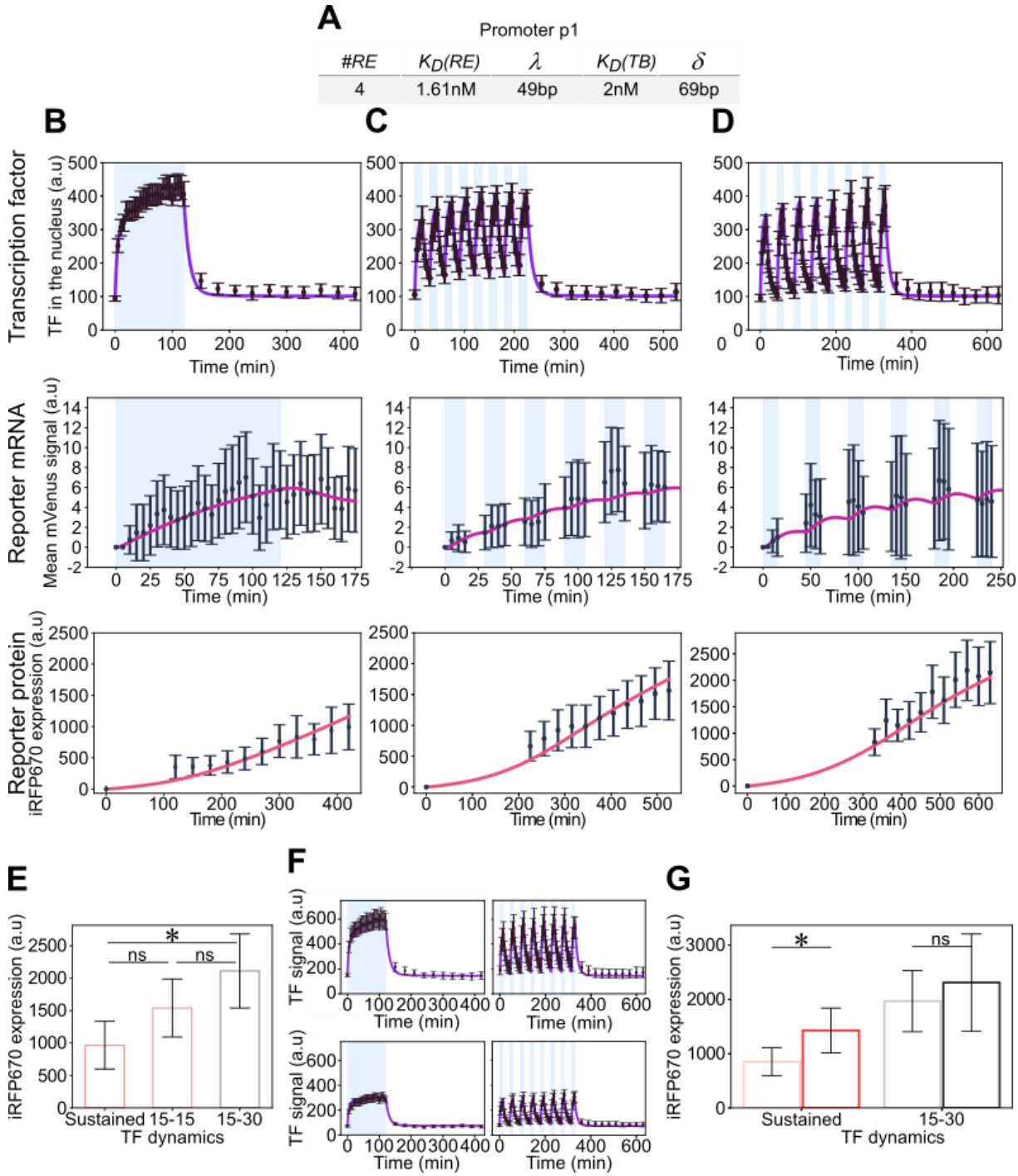

Figure 2.1.9 | A promoter with high affinity REs and TATA-box cannot distinguish between sustained and pulsatile TF dynamics. A) Table showing elements and paramters of promoter p1. The promoter consisted of four repeats of RE with affinity K_D of 1.61nM, TATA-box with affinity of 2nM K_D, and kozak sequence. B) Data for sustained synTF dynamics. Sustained accumulation of synTF in the nucleus of cells (top) when cells were exposured to 2h of constant blue light activation. C) High frequency pulsatile dynamics. 15-15 pulsatile accumulation of synTF in the nucleus of cells exposed to 15-15 pulses blue light activation (top). D) Data for low frequency pulsatile dynamics. 15-30 pulsatile accumulation of synTF in the nucleus of cells when exposed to 15-30 pulses of blue light activation (top). For B-D, mean nacent mRNA foci volume and iRFP670 reporter protein expression are shown in the middle and bottom panels respectively. Blue light activation in reporter protein quantification experiments were followed by 5 hrs in the dark to allow the iRFP670 reporter protein to mature.

E) Comparison of the expressed reporter protein induced by the different synTF dynamics at 5 hrs post activations. F) High and low amplitudes of sustained (left) and 15-30 pulsatile (right) synTF nuclear accumulation dynamics. G) Reporter protein expression at 5hrs post activation induced by the dynamics and amplitudes shown in (F). The lighter and darker colours mark low and high nuclear synTF amplitudes respectively. Six repeats of 15 min blue light activation instead of eight were used to generate the pulsatile dynamics when visualizing mRNA transcription. The trendlines in the reporter mRNA and protein plots in (B) were mathematically fitted while trendlines in (C&D) were model predictions. Error bars in all plots represents 95% confidence interval of the means of at least 20 individual cell measurements while ns and * represents p-value > 0.05 and p-value < 0.05 respectively.

2.1.7 Lower TBP affinity for the TATA-box alone does not enable decoding of TF dynamics

Promoter p2 has similar parameters as promoter p1 except I used a lower affinity TB (Figure 2.1.10A). With this promoter with high affinity REs (K_D of 1.61nM) and a lower affinity TB (K_D of 4nM)[326], for the sustained synTF signal the mean mRNA level overshoots before reaching steady state (Figure 6.1A), a behaviour called refractory [318]. The mean mRNA level was similar for all three dynamics (Figure 2.1.10B). In accordance to mRNA expression, the reporter protein expression data showed no significant difference for the different dynamics (Figure 2.1.10C, Figure 6.1). As for promoter p1, also promoter p2 was insensitive to synTF amplitude (Figure 2.1.10E, Figure 6.1). In conclusion, decreasing the affinity of TBP for the TB only decreases the overall achievable reporter expression level, but does enable the promoter to distinguish between different TF dynamics.

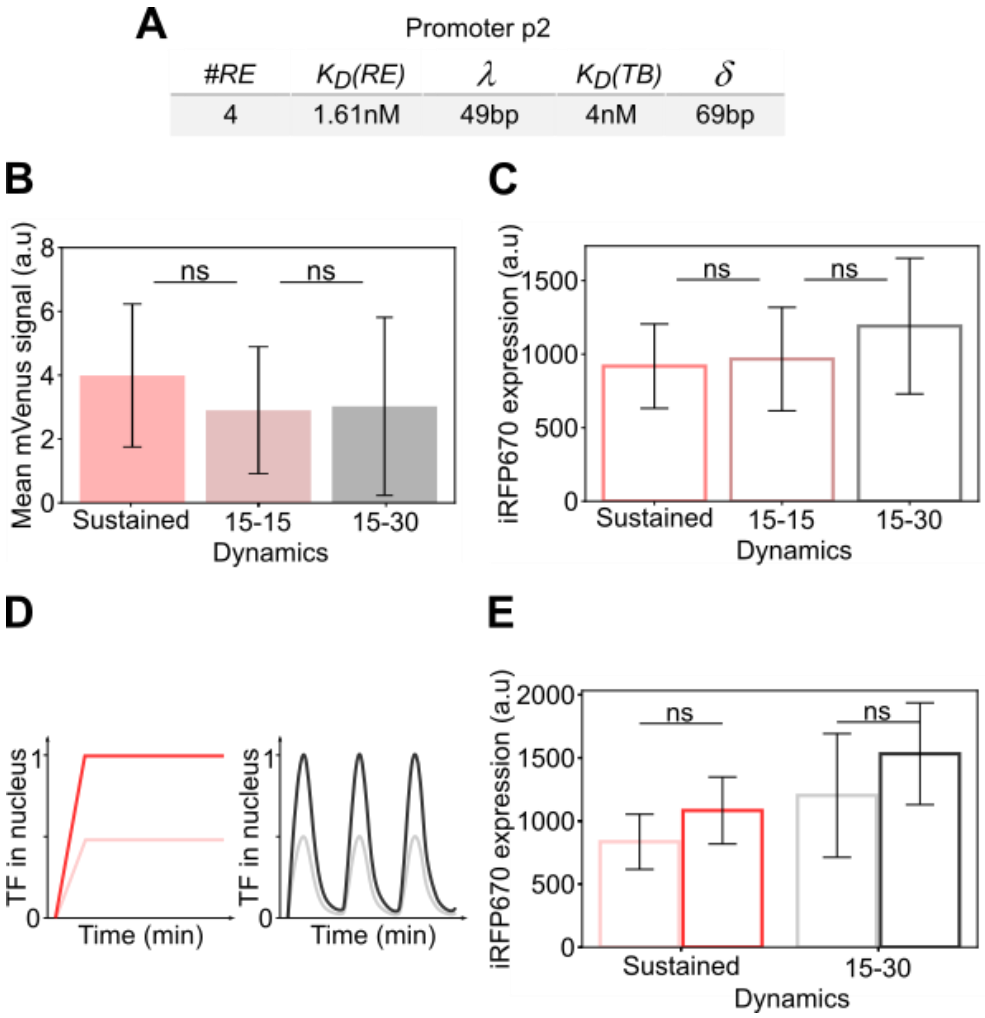

Figure 2.1.10 | Promoter with low affinity TATA-box and high affinity REs does not distinguish different synTF dynamics. A) Details of promoter p2 which has high affinity REs and lower TATA-box affinity. B) Mean mRNA expression induced by the different synTF dynamics. C) Mean reporter protein expression induced by the different synTF dynamics at 5 hours post blue light activation. D) Scheme showing the high and low amplitudes of sustained synTF in the nucleus (left) and high and low amplitudes of 15-30 pulsatile nuclear synTF (right). E) Comparison of reporter protein expression under high or low amplitudes of sustained (red) or 15-30 pulsatile (black) synTF dynamics. Error bars in all plots represents 95% confidence interval of the mean of at least 20 inidividual cells measurements. ns denotes p-value > 0.05.

2.1.8 A promoter with high affinity TB but low affinity REs is sensitive to synTF amplitude but not frequency

I constructed promoter p3 keeping the same TB same as in promoter p1, but using REs that are bound by LexA with more than 3.5-fold lower affinity (Figure 2.1.11A). The reporter protein expression induced by this promoter under sustained synTF dynamics was significantly lower than that obtained with promoters p1 and p2 (Figure 6.2). Reporter mRNA levels were significantly higher under sustained synTF dynamics compared to the pulsatile dynamics (Figure 2.1.11B, Figure 6.2). Promoter p3 exhibited no refractory behaviour in transcript expression since there was a gradual increase in mRNA expression which plateaued under sustained synTF dynamics. At the protein level, promoter p3 behaved similarly to promoter p1. Looking at reporter protein levels, I found that promoter p3 is sensitive to the amplitude of synTF regardless of whether the dynamics were sustained or pulsatile (Figure 2.1.11C, Figure 6.2). Considering that expression noise increases as the affinity of the REs decreases, I find that the significance in the difference between protein levels for the different synTF amplitudes is an interesting observation.

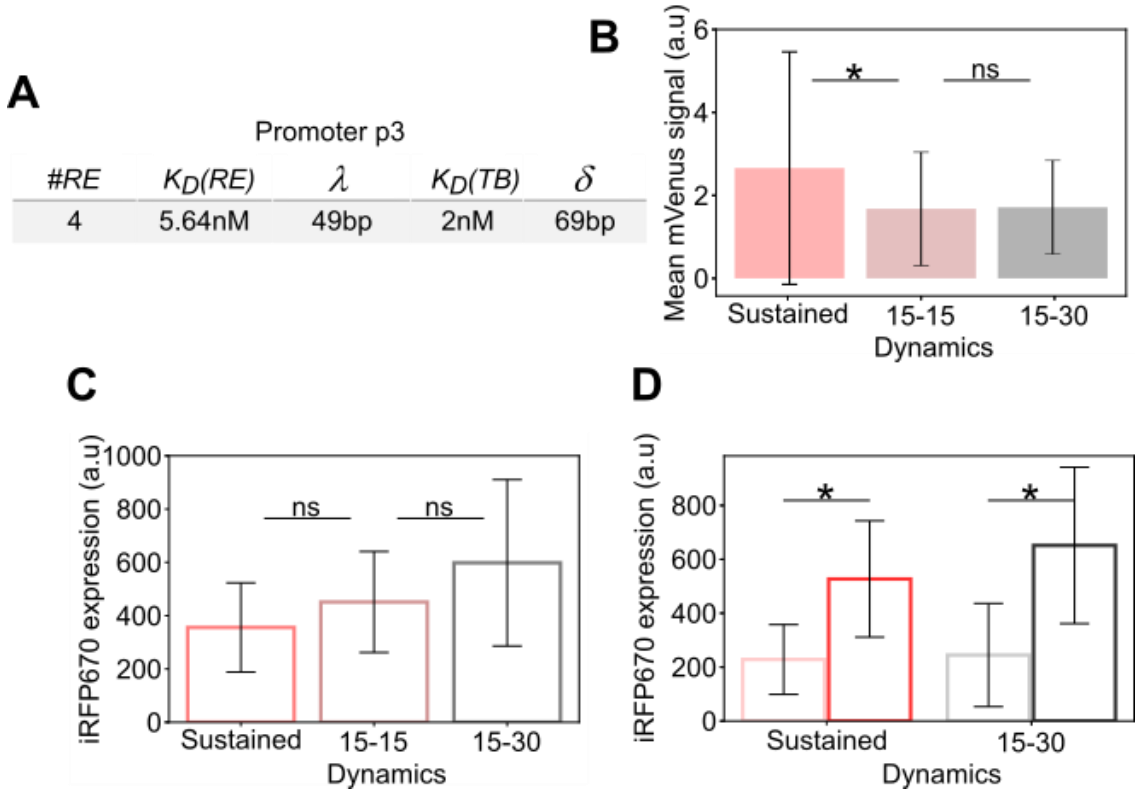

Figure 2.1.11 | Promoter with high affinity TATA-box and low affinity responsive element efficiently integrates nuclear synTF into reporter gene expression. A) Details of promoter with high affinity TATA-box and low affinity REs. B) Mean mRNA expression induced by promoter p3 under the different TF dynamics. C) Comparing reporter protein expression induced by the different dynamics at 5 hrs post blue light activations. D) Comparison of reporter protein expressions induced by high and low amplitudes of sustained (red) and 15-30 pulsatile (black) synTF dynamics. The light and dark colours denote low and high amplitudes respectively. Error bars in all plots represent 95% confidence interval of the mean of at least 20 individual cells measurements. **ns** and ***** represent p-value > 0.05 and p-value < 0.05 respectively.

2.1.9 A promoter with low-affinity REs and TATA-box is sensitive to TF dynamics

I constructed promoter p4 using low affinity RE and TATA-box (Figure 2.1.12A). Nascent mRNA expression induced by this promoter under sustained synTF dynamics showed a strong delayed refractory response (Figure 2.1.12B, middle). The mRNA expression lag time was ~12.5 min, hence 15 min blue light activation before each dark phase in the pulsatile activations barely started expression before synTF is translocated back into the cytosol. As a result, the 15-15 min pulsatile activation resulted in a lower mRNA expression (Figure 2.1.12C, middle), while the 15-30 min pulsatile activation barely resulted in any mRNA expression in the first 120 minutes (Figure 2.1.12D). After 30 minutes incubation in the dark, the amount of synTF in the nucleus decreases down to levels similar those found in this compartment prior to blue light activation (Figure 2.1.12). This means that, in every 'ON' phase of the 15-30 pulses, transcription initiation on promoter p4 has to be restarted by synTF. Mean mRNA levels at each time point were therefore significantly higher for the sustained dynamics compared to the pulsatile ones (Figure 2.1.13B). Zooming in into the first 60 minutes of activation, I observed that mRNA levels increased steadily for the sustained activation (Figure 2.1.13A). In contrast, for pulsatile synTF dynamics, there was no such steady increase in mRNA levels. Similarly, despite the high noise, sustained synTF dynamics corresponded to significantly higher protein levels compared to pulsatile dynamics (Figure 2.1.13C). The strong refractory behaviour exhibited by this promoter makes the gene expression obtained from it more stochastic since exiting the refractory state after the initial overshoot is a random process [200,316,319]. Although reporter protein expression under promoter p4 show a twofold difference when triggered by high and low amplitudes of sustained dynamics, the difference is not significant due to the high variability in expression (Figure 2.1.13D). Pulsatile dynamics on the other hand led to very low and highly variable protein levels. It is therefore not possible to conclude whether in this case synTF amplitude plays a role.

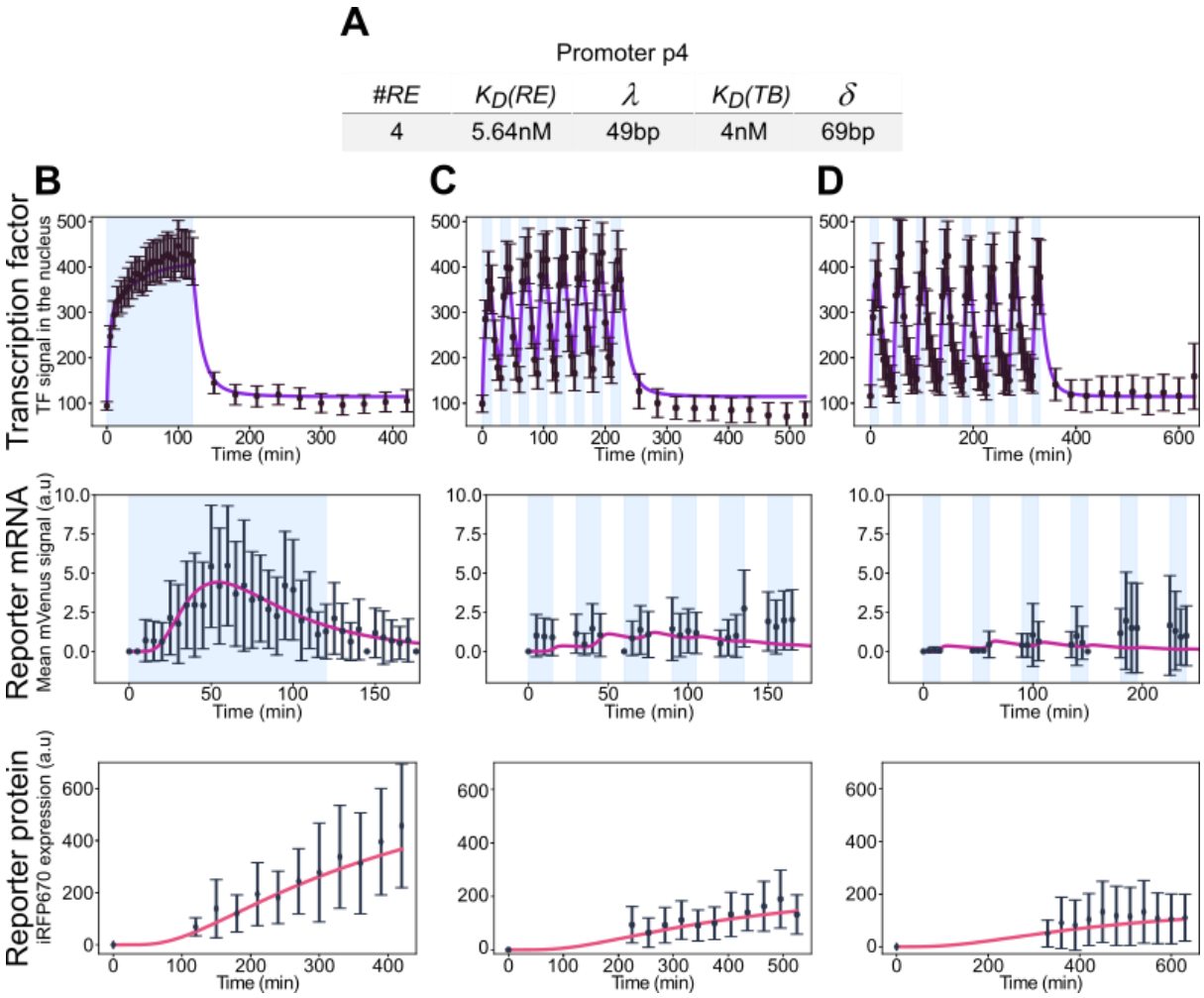

Figure 2.1.12 | Promoter with a low affinity REs and TATA-box displays a delayed refractory response under sustained synTF dynamics. A) Promoter p4 elements and parameter details. B) SynTF translocation and reporter expression data under sustained synTF dynamics. Sustained accumulation of synTF in the nucleus of cells (top) 2h constant blue light activation followed by 5hrs in the dark. C) High frequency pulsatile dynamics. 15-15 pulsatile accumulation of synTF in the nucleus of cells (top) when they were exposed to 15-15 pulsatile blue light activation. D) Low frequency pulsatile dynamics. 15-30 pulsatile nuclear accumulation dynamics of synTF upon exposure to 15-30 pulsatile blue light activation (top). For B-D, mean nacent mRNA foci volume and iRFP670 reporter protein expression are shown in the middle and bottom panels respectively. Six repeats of 15 min blue light activation instead of eight were used to generate the pulsatile dynamics during live cell imaging of mRNA. The trendlines in the reporter mRNA and protein plots in (B) were mathematically fitted while trendlines in (C&D) were model predictions. Error bars in all plots represents 95% confidence interval of the mean of at least 20 individual cell measurements.

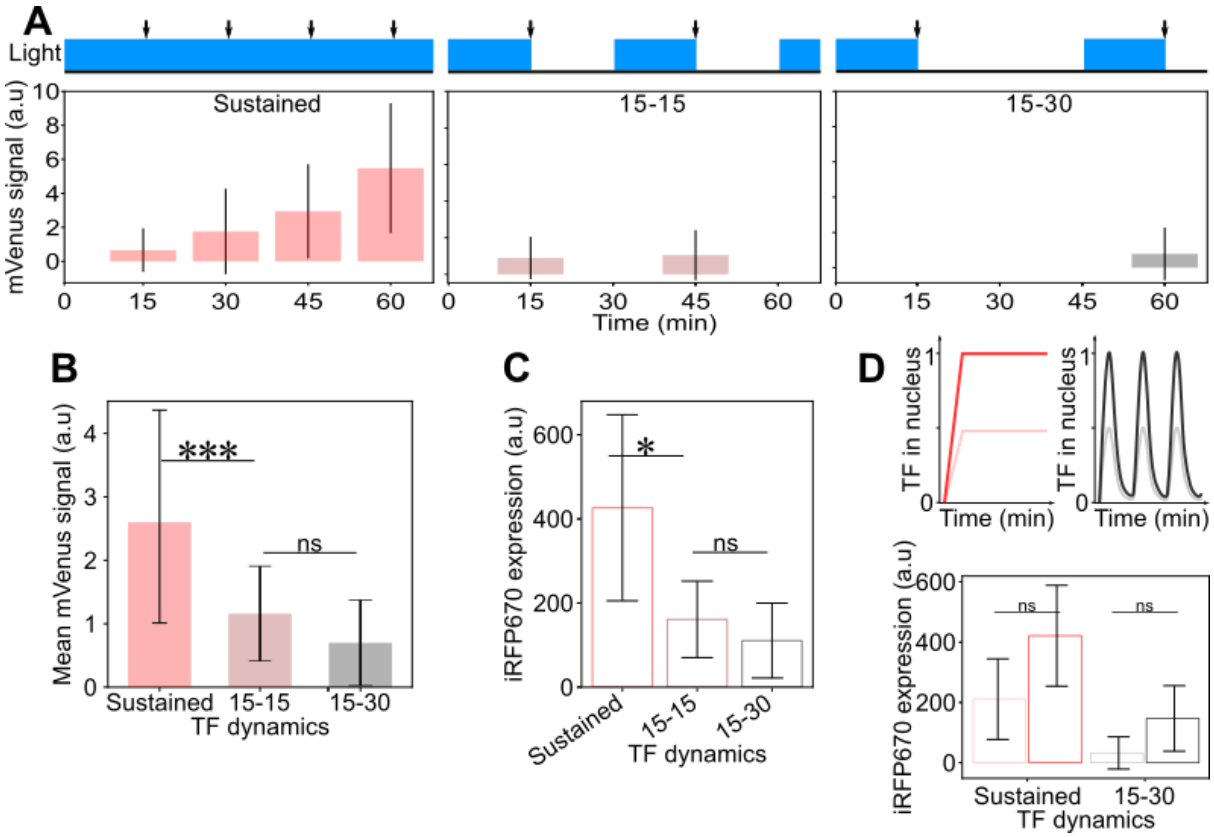

Figure 2.1.13 | Promoter with a low affinity REs and TATA-box induces lower reporter expression under pulsatile synTF dynamics. A) Mean mRNA foci volume at different time points of the corresponding synTF dynamics. The top blue panel indicates blue light activation. Reporter mRNA expression foci data shown in the barplots were taken at the end of each 15 minutes of blue light activation indicated by the black arrows on the top panel. Left, middle and right barplots mRNA expression data for sustained, 15-15 and 15-30 pulsatile synTF dynamics. B) Comparison of mean mRNA expression under the different synTF dynamics. C) Comparison of reporter protein expression at 5hrs post blue light activations which generated the different synTF dynamics. D) Effect of nuclear synTF amplitude on gene expression. Schemes of high and low sustained synTF dynamics (top-left) and 15-30 pulsatile dynamics (top-right). Reporter protein expressions at 5hrs post blue light activation induced by the different amplitudes of sustained and pulsatile dynamics (bottom). Error bars in all plots represents 95% confidence interval of the mean measurement of at least 20 individual cells. ns , * and *** represent p-value > 0.05, p-value < 0.05 and p-value < 0.001 respectively.

2.1.10 Mathematical model of gene expression

The data gathered so far suggests that promoters with high affinity REs (K_D ~ 1.161nM) cannot distinguish different synTF dynamics. Perhaps that is due to the long dwelling time of the TF on the promoter or low amplitude threshold required to induce reporter expression. LexA has been shown to bind stably to its consensus sequence for ~5 minutes before dissociating from it [334]. Such stable binding increases the probability that each looping incidence results in a successful transcription initiation when the TB is not in a refractory state. A TF that weaker affinity for the cognate REs, on the other hand, may have a shorter dwelling time and thus allows a promoter to be able to distinguish different TF dynamics since not all loops formed will have a TF-bound RE even if the TB is active. A promoter with a strong TB efficiently

nucleates the PIC thus allowing for transcription initiation when loops are formed with TF-bound REs.

Based on this understanding, I developed a mathematical model to describe a TF-promoter system, where sub-optimal RE and TB affinities are needed to distinguish different synTF dynamics. The species used in the model and their localization are depicted (Figure 2.1.14A). I modelled the change in concentration of every species over time using ordinary differential equations (ODEs). To model the accumulation of synTF in and out of the nucleus I used different import and export rates for cells in the dark or exposed to blue light. This assumption is based on the fact that nuclear export and import rates for a shuttling proteins are not exclusive and localization depends on which of the two processes dominate [335]. This is especially true for synTF, for which the competition between the NES and the NLS determines the subcellular localization [247]. The different rates were incorporated in the functions liAct(t) for blue light activation and drkrev(t) for dark incubation. All other parameters used in this model are described in table 5.1. The equation describing the change over time of the concentration of cytosolic synTF (TF) contains therefore two functions together with a terms for the production (K_{TF}) and degradation (d_{TF}) of synTF:

$$\frac{dTF}{dt} = K_{TF} + drkrev(t) - liAct(t) - d_{TF} * [TF] \qquad \text{Equation1}$$

The equation describing the change over time of the concentration of nuclear TF (NTF) also contains the functions describing import and export, plus a term for the degradation of synTF in the nucleus:

$$\frac{dNTF}{dt} = liAct(t) - drkrev(t) - d_{NTF} * [NTF] \qquad \text{Equation2}$$

Three-state promoter model has been shown to be the simplest model that can account for refractoriness that is exhibited in the expression of some genes [200]. I therefore implemented a three-state promoter model given the observed refractory behaviour exhibited by some promoters with weaker TB. The nuclear synTF (NTF) binds cooperatively to the unbound promoter P_{r1} and converts it reversibly to an active promoter P_{r2}, with maximum rate K_{on}. P_{r2} and P_{r3} revert to P_{r1} with a rates K_{off} and $d1_{rf}$ respectively.

$$\frac{dP_{r1}}{dt} = K_{off} * P_{r2} + d1_{rf} * P_{r3} - K_{on} * \left(\frac{NTF^m}{{kd_1}^m + NTF^m}\right) * P_{r1} \qquad \text{Equation3}$$

When synTF falls off the promoter, P_{r2} becomes unbound again reverting to P_{r1} or switching to a refractory state P_{r3} that is inactive with maximum rate K_{rf}.

$$\frac{dPr_2}{dt} = K_{on} * \left(\frac{NTF^m}{{kd_1}^m + NTF^m}\right) * P_{r1} + d2_{rf} * P_{r3}$$

$$-K_{off} * P_{r2} - K_{rf} * \left(\frac{NTF^m}{{kd_1}^m + NTF^m}\right) * P_{r2} \qquad \text{Equation4}$$

All promoter states are reversible but with different rates. The refractory promoter state can switch to unbound or directly to active state (P_{r2}) with rate $d2_{rf}$.

$$\frac{dP_{r3}}{dt} = K_{rf} * \left(\frac{NTF^m}{{kd_1}^m + NTF^m}\right) * P_{r2} + \left(d1_{rf} + d2_{rf}\right) * P_{r3} \qquad \text{Equation5}$$

Nucleation of the transcription PIC is dependent on the nuclear synTF but not directly dependent on promoter state. However, the nucleated PIC is not sufficient to initiate transcription since no mRNA nor protein expression was observed for any of the promoters in the absence of the transcription factor.

$$\frac{dPIC}{dt} = K_{pic} * \left(\frac{NTF^m}{{kd_1}^m + NTF^m}\right) - d_{pic} * PIC \qquad \text{Equation6}$$

I further assumed that the probability that a nucleated PIC leads to transcription initiation depends on the stability of the nucleated complex, which in turn depends on the strength of the TATA-box (that is, how high is the affinity of the TBP for it in terms of K_D) [336,337]. Upon TF binding to the RE(s), the promoter needs to loop to bring the TBP bound TATA-box and the TF-bound RE(s) into close proximity to enable transcription initiation [338,339]. The efficiency of DNA looping between two points depends on the distance between them [114,120,340–342]. The nascent mRNA (nRNA) expression term was therefore divided by the factor j_m (equation 11) to account for the distance between the REs and the TB. I modelled the concentration of nRNA as a function of the concentrations of active promoter P_{r2} and PIC.

$$\frac{dnRNA}{dt} = \frac{K_{Nrna}}{j_m} * \left(\frac{[PIC]^m}{{kd_2}^n + [PIC]^n}\right) * P_{r2} - d_{nRNA} * nRNA \qquad \text{Equation7}$$

I considered a constant rate of mRNA maturation (K_{mRNA}) from nascent mRNA and a constant degradation term (d_{mRNA}):

$$\frac{dmRNA}{dt} = K_{mRNA} * nRNA - d_{mRNA} * mRNA \qquad \text{Equation8}$$

Due to the long maturation time of the iRFP670 reporter, I divided the protein expression into two processes. The first is the production of the unfolded protein (P_{uf}), which does not contribute to the quantified protein fluorescence. Translation of mRNA into the unfolded protein happened at a rate K_p, whereas maturation and degradation happened at rates R_p and d_p respectively.

$$\frac{dP_{uf}}{dt} = K_p * mRNA - \left(R_p + d_p\right) * P_{uf} \qquad \text{Equation9}$$

The second process describes how the unfolded protein matures into a fluorescent protein ($_m$P). This term takes into consideration the maturation half-time of iRFP670 which gives the rate R_p. The d_{mP} term describes the degradation term of the matured protein:

$$\frac{dmP}{dt} = R_p * P_{uf} - d_{mP} * mP \qquad \text{Equation10}$$

The factor j_m, which describes DNA looping efficiency between the TB and the REs, was calculated as previously described [200]. L and P represent distance (between TB and RE in bp) and persistent length of the chromatin (in nm) respectively:

$$j_m = \left(\frac{1.25e^5}{P^3}\right) * \left(\frac{4*P}{L*10^4}\right)^{\frac{3}{2}} * e^{-(510*P^2)/(6.25*L^2+50*P^2)} \qquad \text{Equation11}$$

Maturation of nascent mRNA was later observed to be better modelled with a Michaelis-Menten function due to the many enzymatic steps involved in pre-mRNA processing. Equation 8 was therefore updated as shown below:

$$\frac{dmRNA}{dt} = K_{mRNA1} * \left(\frac{[nRNA]}{kd_m + [nRNA]}\right) - d_{mRNA} * mRNA \qquad \text{Equation12}$$

Similarly, protein production as described in equation 9 was updated to account for the enzymatic translation processes and the fact that the population of mRNAs being translated at any time point is dependent on the concentration of mRNA.

$$\frac{dP_{uf}}{dt} = K_{p1} * \left(\frac{[mRNA]^q}{kd_2{}^q + [mRNA]^q}\right) - \left(R_p + d_p\right) * P_{uf} \qquad \text{Equation13}$$

2.1.11 Low DNA looping efficiency leads to low protein expression but allows for differential expression in response to different TF dynamics

Nucleosome positioning on/between a responsive elements and the core promoter has been shown to affect the ability of a promoter to distinguish different TF dynamics [201]. DNA looping efficiency of the DNA stretch between the RE(s) and the TB has also been shown to be a function of the distance between the two elements. The longer the distance the lower the looping efficiency [114]. To test how nucleosome positioning and lower looping efficiency impact differential gene expression in response to different TF dynamics, I inserted the nucleosome positioning sequence 601 [343] between the REs and the TB (Figure 2.1.14B). Plasmid DNA has been shown to possess nucleosome-like structures [344–346], thus I expected the 601 nucleosome positioning sequence to allow for the formation of nucleosomes on the plasmids in living cells. If nucleosome positioning is a strong determinant of whether a promoter can distinguish between different TF dynamics, a promoter with the 601 sequence should differentiate TF dynamics better than one with a random sequence. On the other hand, if DNA looping efficiency determines whether a promoter can distinguish between different TF dynamics, the presence or absence of a nucleosome between the REs and the TB should make no difference.

Using the mathematical model for promoter p2, I updated the DNA length to account for the inefficient looping as previously reported [114,200]. Since we did not know the exact impact of nucleosome positioning on the expression of the reporters in this experimental set up, no specific updates were made to account for nucleosome positioning. Hence, same reporter expression levels were predicted by the mathematical model for both promoters. Experimental p7 reporter protein expression data under sustained synTF dynamics was similar to the levels predicted by the mathematical model (Figure 2.1.14D & Figure 2.1.15E). Experimental p8 reporter protein expression data was however, lower than that of p7 and obviously lower than the mathematical prediction (Figure 2.1.15E). This was counter intuitive in light of previously reported observations [200,201], because promoter p7 was expected to have the most differential expression in response to different dynamics and perhaps lower expression since it had the nucleosome positioning in addition to the distance. Computational modelling of the DNA sequences revealed that the 601 sequence has an inherent curvature which was not observed in the random sequence (Figure 2.1.14C). I concluded that the inherent curvature of 601 made looping in promoter p7 more efficient compared to promoter p8 thus, explaining the difference in gene expression [119]. The mathematical model of promoter p8 was further updated to account for the increased inefficiency of DNA looping due to the rigid nature of the random sequence. The updated model showed low protein expression induced under promoter p8 (Figure 2.1.14D).

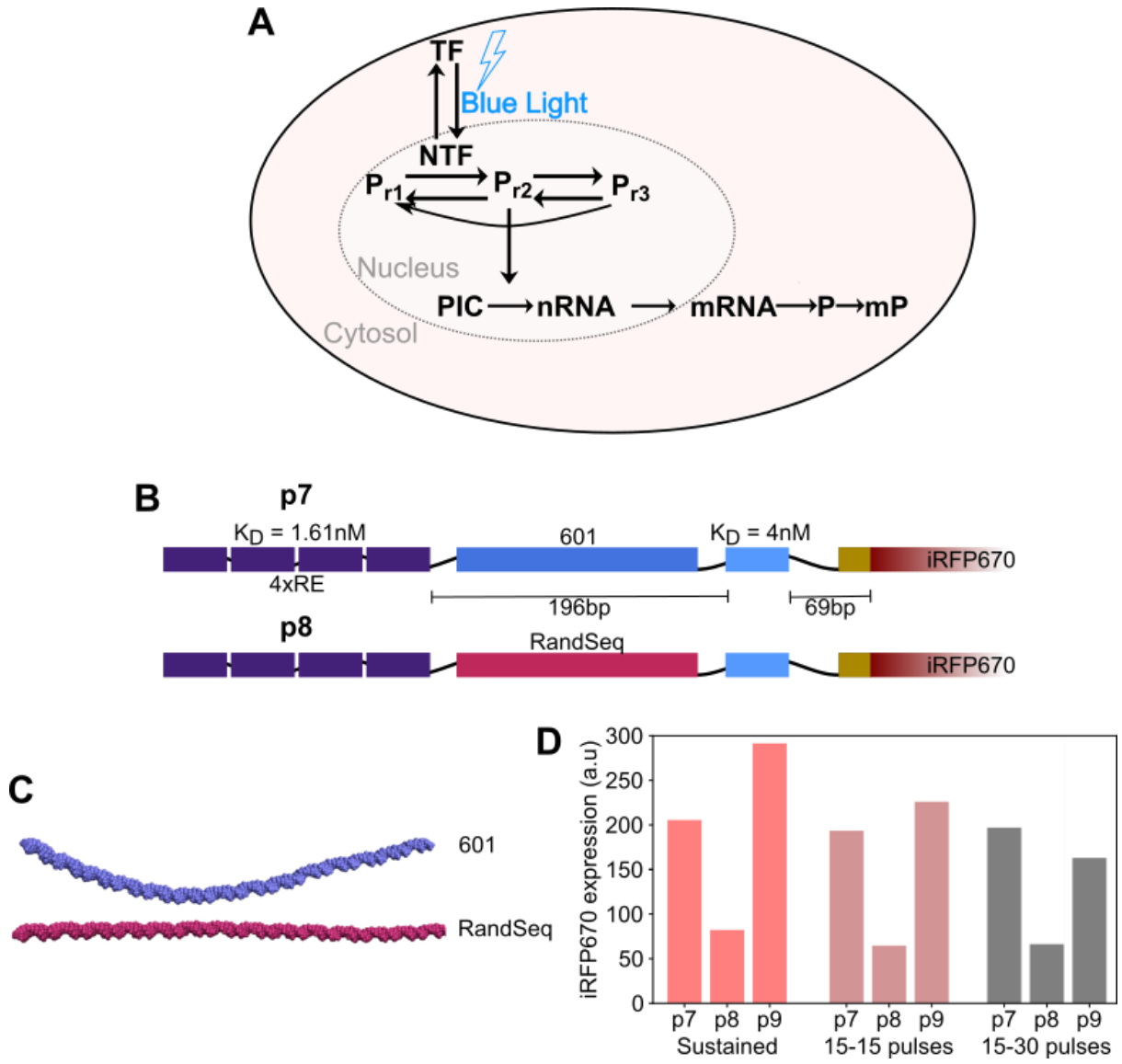

Figure 2.1.14 | Low DNA looping efficiency helps to distinguish different synTF dynamics. A) Subcellular localization of state elements used in the mathematical modelling of reporter expression in response to synTF dynamics. Arrows indicate the direction of transitioning of state elements. TF and NTF represent cytosolic and nuclear TF levels respectively. P_{r1}, P_{r2} and P_{r3} represent unbound, bound/active and refractory promoter states. PIC, nRNA, mRNA, P and mP denote transcription PIC, nascent mRNA, matured mRNA, unfolded reporter proteins and matured reporter proteins respectively. B) Depiction of reporters showing the insertion of an extra 147 bp of either nucleosome positioning (601) sequence (top) or randomly generated (RandSeq) sequence (bottom) between the responsive elements and TATA-box of promoter p2. C) DNA bending models showing the inherent bend of the 601 nucleosome positioning sequence compared to the randomly generated sequence. D) Mathematical model predictions of protein expression levels of reporters under promoters p7 (601) and p8 (RandSeq) as well as promoter p9 (RandSeq plus K_D(TB) = 2nM) when induced by the different synTF dynamics. Promoter p9 addition was necessary because of low expression induced by p8.

A

	#RE	K_D(RE)	λ	K_D(TB	δ
p1	4	1.61nM	49 bp	2nM	69 bp
p2	4	1.61nM	49 bp	4nM	69 bp
p3	4	5.64nM	49 bp	2nM	69 bp
p4	4	5.64nM	49 bp	4nM	69 bp
p5	2	1.61nM	49 bp	2nM	69 bp
p6	2	1.61nM	49 bp	4nM	69 bp
p7	4	1.61nM	196[a] bp	4nM	69 bp
p8	4	1.61nM	196[b] bp	4nM	69 bp
p9	4	1.61nM	196[b] bp	2nM	69 bp

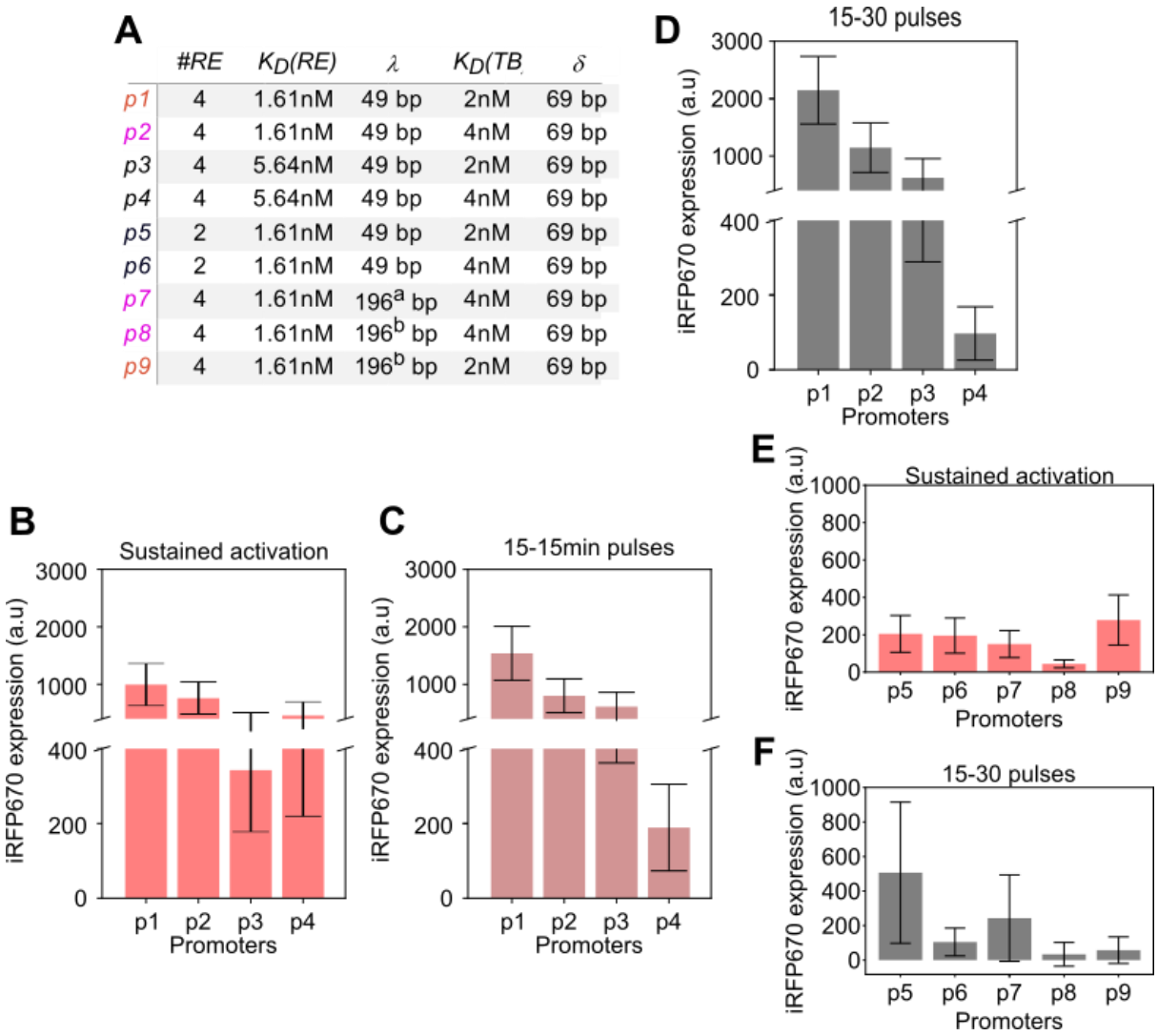

Figure 2.1.15 | Promoters with lower probability of interaction between TF-bound RE and TBP-bound TATA-box show differential expression in response to different synTF dynamics. A) List of promoters and their corresponding promoter elements and parameters. Orange and purple colours highlight promoters that differ only by the value of λ. B) Reporter protein expression induced by promoters with four responsive elements, λ equals 49bp and δ equals 69bp under sustained synTF dynamics. [a] and [b] connotes 601 and RandSeq insertion respectively. C&D) Reporter protein expression induced by the promoters in (B) under 15-15 and 15-30 pulsatile synTF dynamics respectively. E&F) Expression of reporter protein induced by p5 to p9 promoters with the parameters details in (A) under sustained and 15-30 pulsatile nuclear TF dynamics. All error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

I updated the mathematical model by replacing the TATA-box affinity term of p8 (K_D of 4nM) with 2nM K_D TATA-box term. This was done because reporter protein expression under promoter p8 was very low. This model predicted higher protein expression under sustained compared to pulsatile dynamics (Figure 2.1.14D). Interestingly, it was experimentally confirmed that promoter p9 exhibits differential response under different synTF dynamics despite being similar to promoter p1 except the insertion of random sequence (Figure 2.1.15E&F). Protein levels were reduced by ~80% and ~94% for promoters p7 and p8 compared to promoter p2, respectively (Figure 2.1.15B&E) whereas there was about 72%

reduction in protein levels for promoter p9 compared to promoter p1 (Figure 2.1.15B&E). Promoter p4, which has low affinity RE and TB, best distinguished different TF dynamics in terms of protein expression despite having a similar value and sequence of λ as promoters p1-6 (Figure 2.1.15B-D). Due to low expression levels, nascent mRNA visualization was not possible when the number of REs was halved as in the case of promoters p5 and p6. These promoters did not distinguish between different synTF dynamics (Figure 2.1.15E&F). Promoter p5 however, showed similar expression trends as p1 by inducing higher expression under 15-30 pulsatile dynamics. Interestingly, promoter p9 showed a stronger difference for sustained versus 15-30 pulses in the experiments than what was predicted by the model (compare Figure 2.1.14D Figure 2.1.15E&F).

2.1.12 Reduced mRNA translatability improves the ability of a promoter to distinguish different TF dynamics

Protein expression levels can be regulated at both the transcription and translation levels [347,348]. So far I have shown that transcription initiation can be regulated to allow for differential expression in response to different synTF dynamics. I then asked if translation initiation also plays a role in determining if a promoter responds differently to different TF dynamics. To address this with my model, I first adjusted equations 8 and 9 assuming that both, maturation of nascent mRNA and translation of the matured mRNA are concentration-dependent, thus justifying modelling the first process with a Michaelis-Menten equation (equation 12) and the second with a hill function (equation 13). Pre-mRNA processing is a multi-step process, which is dependent on the efficiency of 5' capping and the type of capping proteins [349,350]. Polyadenylation and efficiency of mRNA release from RNAP II can also affect how the quantified nascent mRNA foci volume correlates with mRNA concentrations. Similarly, secondary structures in the 5' UTR, mRNA scanning to find the first translation codon and the efficiency of translation initiation also affects translation efficiency and in effect the protein expression levels [170,177,351–353]. This implies that, for an mRNA that is inefficiently translated, higher mRNA concentration is required to improve the probability of successful translation initiation before the mRNA is degraded. Stability of mRNAs is also affected by translatability. With the updated promoter p2 model, I predicted that low protein translation rate results in differential gene expression when mRNA a less stable (Figure 2.1.16C). Low translation initiation rate can therefore impact differential protein expression in response to different TF dynamics, which may induce different mRNA levels per unit time.

To validate this experimentally, I shortened the distance between the TB and the translation start site to reduce the length of the 5' UTR [177,178]. This was done to eliminate any structures in the region that may either negatively or positively affect protein translation [178]. A shorter 5' UTR led to an increase in reporter protein expression (promoter p10, Figure 2.1.16A&B), likely because of increased translation initiation efficiency since the translation machinery takes less time to scan the mRNA in search of the translation start site. This also reduces the chance that the ribosome falls off before finding the start site [178]. I reasoned that I should increase protein expression as much as possible, to compensate for lower reporter protein expression from those constructs with lower translation efficiency. Therefore, I replaced the adenine at position -6 from the translation start site to a guanine (promoter p11, Figure 2.1.16A&B) to ensure higher protein translation [354,355]. When the Kozak sequence [356] was replaced with a

random sequence (promoter p12, Figure 2.1.16A&B) there was ~ 60% reduction in reporter protein levels under sustained TF dynamics (compare promoters 11 and 12, Figure 2.1.16B). Promoter p12 showed differential protein expression in response to different synTF dynamics (Figure 2.1.16C&D).

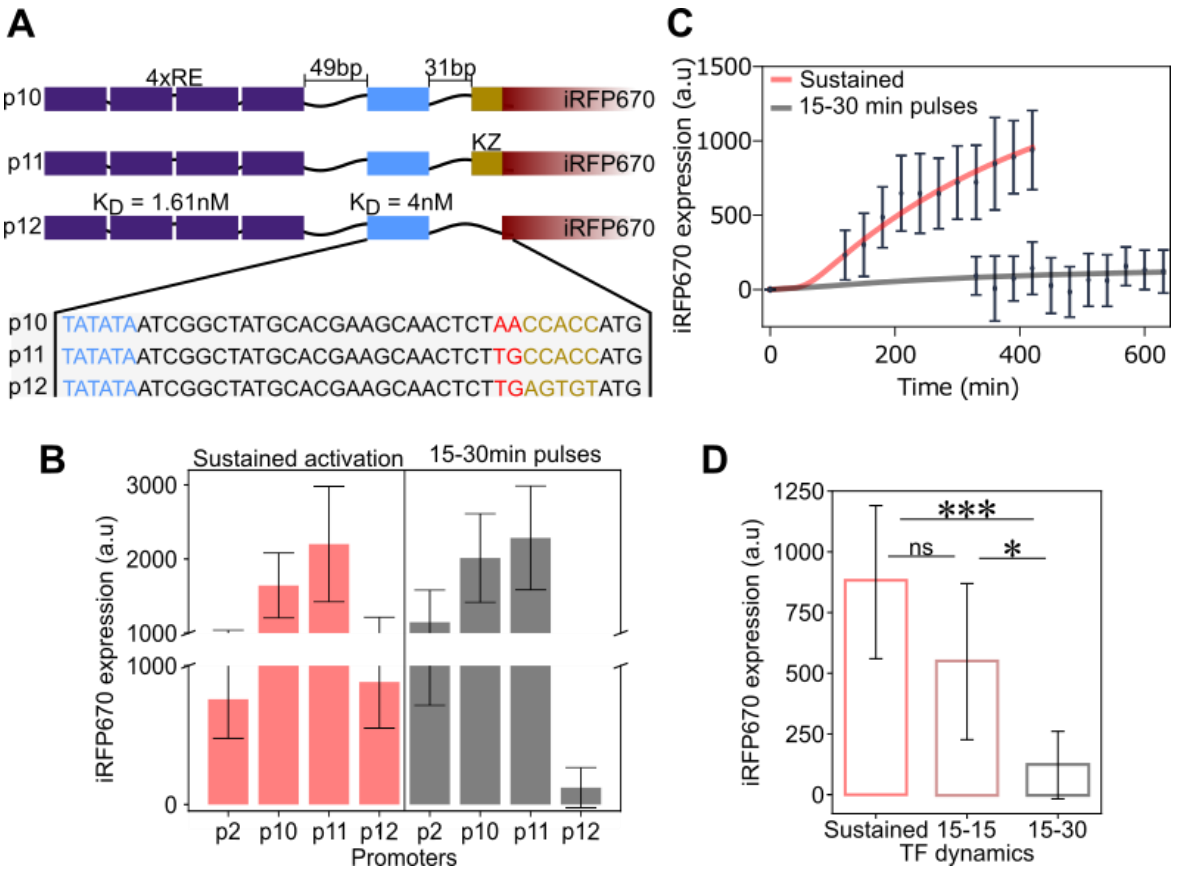

Figure 2.1.16 | Reporter with lower affinity TATA-box and translation rate effectively distinguishes different synTF dynamics. A) Depiction of promoters that generate transcripts with shorter 5' UTR with or without an intact Kozak sequence. Promoter p10 has a kozak sequence and adenine at position -6 from the ATG. Promoter p11 has a Kozak sequence and a guanine at position -6 from the ATG. Also, promoter p12 has a guanine at -6 position from the ATG and no Kozak sequence. B) Expression of reporter protein under promoters p2, p10, p11 and p12 induced by sustained (red) and 15-30 pulsatile (grey) synTF dynamics. C) Experimental and mathematical modelling data agree on the reporter protein expression levels under promoter p12. Red and grey trend lines represent modelled protein expression data induced by sustained and 15-30 min pulsatile synTF dynamics respectively. The dots with error bars represent the experimentally quantified reporter protein levels. D) Comparison of the p12 reporter protein expression at 5 hrs post blue light activation induced by the different TF dynamics. All error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements. Welch's t test was used to test for significance; ns, *, and *** represent p-values > 0.05, < 0.05 and < 0.001 respectively.

2.2 Transcription factor architecture and how it impacts gene expression

2.2.1 Design and construction of synTEAD/synYAP transcription factor and characterization of synYAP nuclear accumulation dynamics

I constructed the TEAD-like DNA binding protein (synTEAD) using the same LexA protein used in synTF to make it compatible with the library of promoters I previously generated (see section 2.1). To localize it to the nucleus, I fused an NLS to it (Figure 2.2.1A). Moreover, to make it interact with the TAD-bearing protein (synYAP) I fused a subunit of a heterodimer pair (ZIPA) to the N-terminus of the NLS-LexA construct (Figure 2.2.1A). The YAP-like synYAP construct was designed as synTF but with LexA replaced with the second subunit of the heterodimer pair (ZIPB). Being under optogenetic control, synYAP could be translocated in and out of the nucleus upon blue light activation (Figure 2.2.1A). In the nucleus, synYAP is targeted to the promoter via interaction with synTEAD, which specifically binds to the DNA at the LexA operator sequence. Once formed, the synTEAD/synYAP complex should be able to induce gene expression. I cloned the synTEAD and synYAP constructs on a single plasmid. I reasoned that I should have higher levels of synYAP than synTEAD since not all synYAP proteins accumulate into the nucleus upon blue light activation. Consequently, I made two constructs: in one, I used the strong CMV promoter to drive both synYAP and synTEAD, with this latter one driven by an internal ribosomal entry site (IRES); in the other, I selected the weak UBC promoter to drive synTEAD (Figure 2.2.1B).

As first heterodimerizer system, I selected the ssrA peptide from *C. crecentus* and its SspB binding protein from *E. coli* [357]. Unfortunately, I did not detect any reporter gene expression, likely because ssrA-SspB do not interact strongly enough (K_D > 240nM; [357]. I then tested the Magnet system [358], but this also did not lead to reporter gene expression. This is not too surprising, given that pMag and nMagHigh interact with weak affinity (K_D > 41μM; [359]. Finally, I tested a heterodimerizing synthetic zipper pair, SYNZIP1 and SYNZIP2, which are reported to interact with high affinity (K_D < 10 nM; [360]. This pair enabled synTEAD/synYAP to induce reporter gene expression (Figure 2.2.1F). Another heterodimer synthetic zipper pair SYNZIP3 and SYNZIP4, which are reported to also bind with high affinity; K_D < 30nM; [360], did not support activity of the TF. This explained why the use of the Magnet system and ssRA/SsPB heterodimer pair did not lead to reporter expression due to high affinity required to induce gene expression. One of the highest affinity ssRA/SsPB heterodimer pairs which was used in the construction of iLID nano has K_D >100nM [361]. Indicating that none of the ssRA/SsPB pairs would have help induce reporter expression. These data show how important it is for synTEAD and synYAP to bind with very high affinity to be able to form a functional TF. In

Table 6.4 I list all the heterodimers I tested. SynYAP constructed with SYNZIP1 showed ~3 fold reversible nuclear accumulation upon blue light illumination, but had relatively high background in the dark (Figure 2.2.1C). I refer to this synthetic TF as y1.

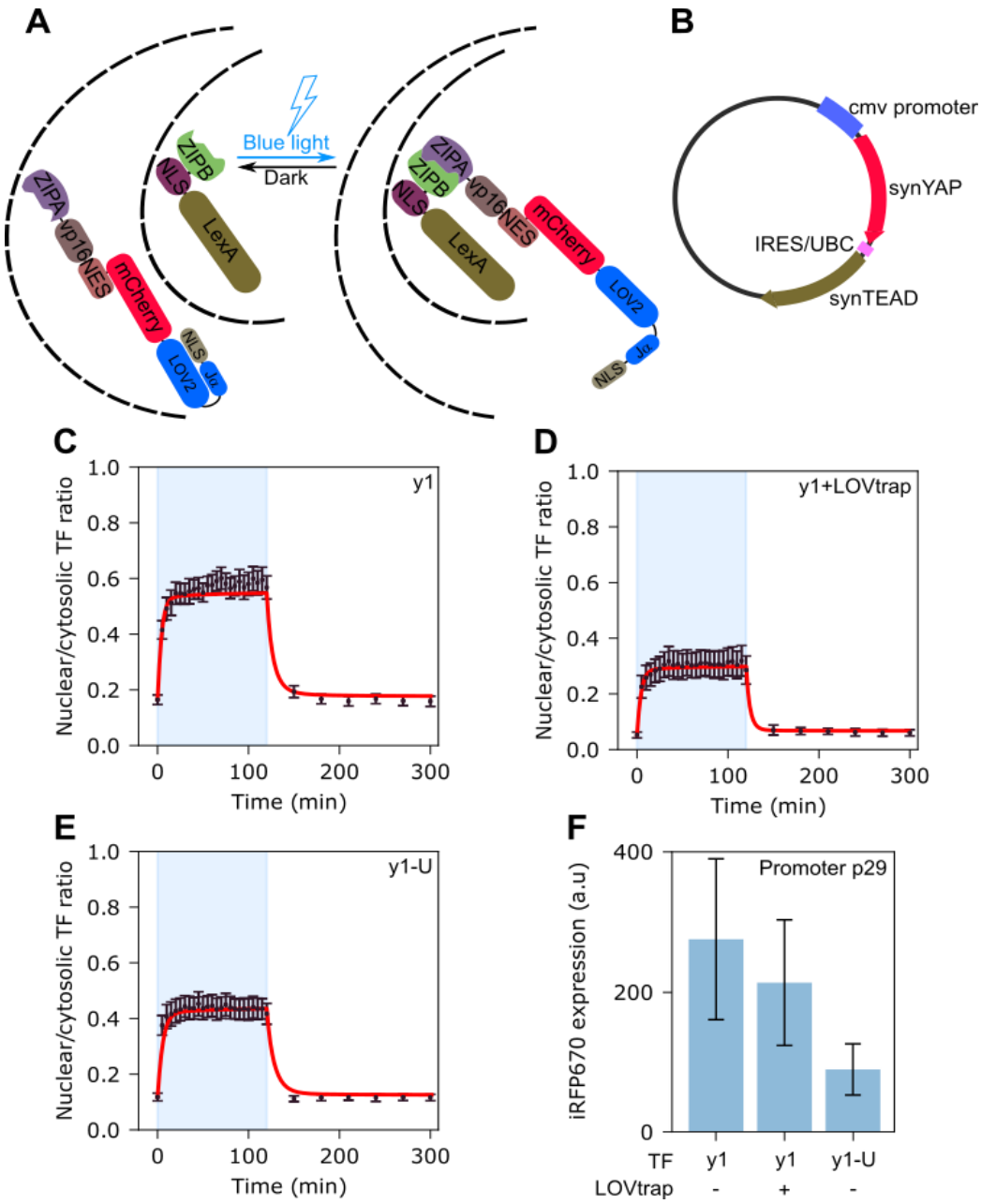

Figure 2.2.1 | Construction of synTEAD/synYAP TF and the characterization of synYAP nuclear accumulation. A) Illustration of synTEAD/synYAP detailing the nuclear localization of synTEAD and the cytosolic retention of synYAP prior to activation with external stimulus. SynYAP translocated to the nucleus upon blue light activation and reverted to the cytosol in the dark. ZIPA and ZIPB represent the two units of a heterodimer pair. B) Depiction of the plasmid construct detailing the tandem arrangement of synYAP and synTEAD. The TAD-bearing synYAP was under the CMV promoter while synTEAD bearing the DBD was expressed under the IRES or the weak UBC promoter. SynTEAD was expressed under weaker promoters to ensure higher expression ratio of synYAP. C) Quantified nuclear accumulation of the synYAP of y1 construct (ZIPA = SYNZIP1 and ZIPB = SYNZIP2) upon constant blue light activation. Due to the high initiation nuclear levels of y1 in the dark, D) the LOVtrap fused to NES was deployed to retain synYAP of y1 more in the cytosol. E) Initial nuclear accumulation and accumulation after blue light activation of the synYAP was reduced when synTEAD was expressed under the UBC promoter (y1-U). Expression under UBC promoter yielded lower synTEAD protein levels. F) Reporter protein expression under promoter p29 triggered by sustained nuclear accumulation of synYAP in the different experimental setups. Trend lines represent model fitted data while the black dots with error bars represent experimental data. All error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

To reduce the nuclear accumulation the synYAP component of y1 in the dark, LOVtrap was used to retained synYAP more in the cytosol. When I co-expressed LOVtrap with synYAP and synTEAD (y1+LOVtrap), I observed significantly lower nuclear levels of synYAP in the dark, however this was also true under blue (Figure 2.2.1D). Interestingly, lowering the expression levels of the synTEAD component of y1 using the UBC promoter (y1-U) also reduced the nuclear accumulation of synYAP in the dark (Figure 2.2.1E). As observed for y1+LOVtrap, also in this case the nuclear accumulation of the synYAP component of y1-U upon blue light activation was reduced (Figure 2.2.1E). These data suggest that synYAP is retained in the nucleus via interaction with synTEAD, whose levels, therefore, become a critical parameter. Not surprisingly, the expression of the reporter under promoter p29 was higher for y1 than for y1+LOVtrap and y1-U, which are characterized by lower nuclear synYAP (Figure 2.2.1F). Promoter p29 is one of the strongest synthetic promoters in my reporter library and it is similar to promoter p1 except δ is shortened to 31bp (Figure 6.4 and Table 6.2).

2.2.2 SynTF has lower nuclear accumulation background and higher dynamic range than synYAP

I expressed both synTF and the different versions of synYAP (different SYNZIPs) under the CMV promoter. To clarify the impact of synTEAD levels on the nuclear import of synYAP, I compared different constructs: (I) synTEAD expressed under the weak UBC promoter, (II) synTEAD transcribed under CMV but translation initiated from an IRES, and (III) synTEAD transcribed CMV but translated at the C-terminus of the T2A self-cleaving peptide which triggers ribosomal skipping to generate two separate proteins. The mean synTEAD expression level for the IRES construct was about 60% of that measured for the T2A construct, which I considered to indirectly reflect the level that would be obtained by driving the expression directly from the CMV promoter (Figure 2.2.2B). Expression under the UBC promoter was ~10% of the expression achieved under IRES (Figure 2.2.2B). To compare different constructs, I calculated the nuclear-to-cytosolic signal ratio for cells in the dark and after blue light illumination. For synTF, the ratio increased by ~3 fold upon blue light activation and reverted to initial levels in the dark (Figure 2.2.2A).

Nuclear import of synYAP constructed with SYNZIP2 was significantly higher than synTF nuclear import when synTEAD consisting SYNZIP1 was expressed under the IRES (y2-I) (Figure 2.2.2C). Nuclear import ratio of y2-I synYAP even in the dark was higher than synTF nuclear accumulation under constant blue light activation. The accumulation of synYAP of y2-I in the nucleus was also significantly higher than of y1 (which has the same design setup except the SYNZIPs were swapped). When synTEAD of y2-I was expressed under UBC (y2-U), the background nuclear accumulation of synYAP was drastically reduced yet the nuclear import was higher than that of synTF (Figure 2.2.2D). Nuclear import ratio of synYAP of y2-U was similar to y1 (synTEAD expressed under IRES). This suggests that the arrangement of the zippers in y2 consents a better interaction between SYNZIP2 and SYNZIP1 compared to the arrangement in y1. The synthetic zippers require a parallel arrangement to effectively interact [360]; this is true in both arrangements, though perhaps the binding interfaces may be differently positioned in y1 and y2. I observed a similar trend of nuclear accumulation for synYAP constructed with SYNZIP4 when synTEAD was constructed with SYNZIP3 (y4). The nuclear import was higher for the construct with the IRES (y4-I) than for that with the UBC promoter (y4-U) (Figure 2.2.2E&F).

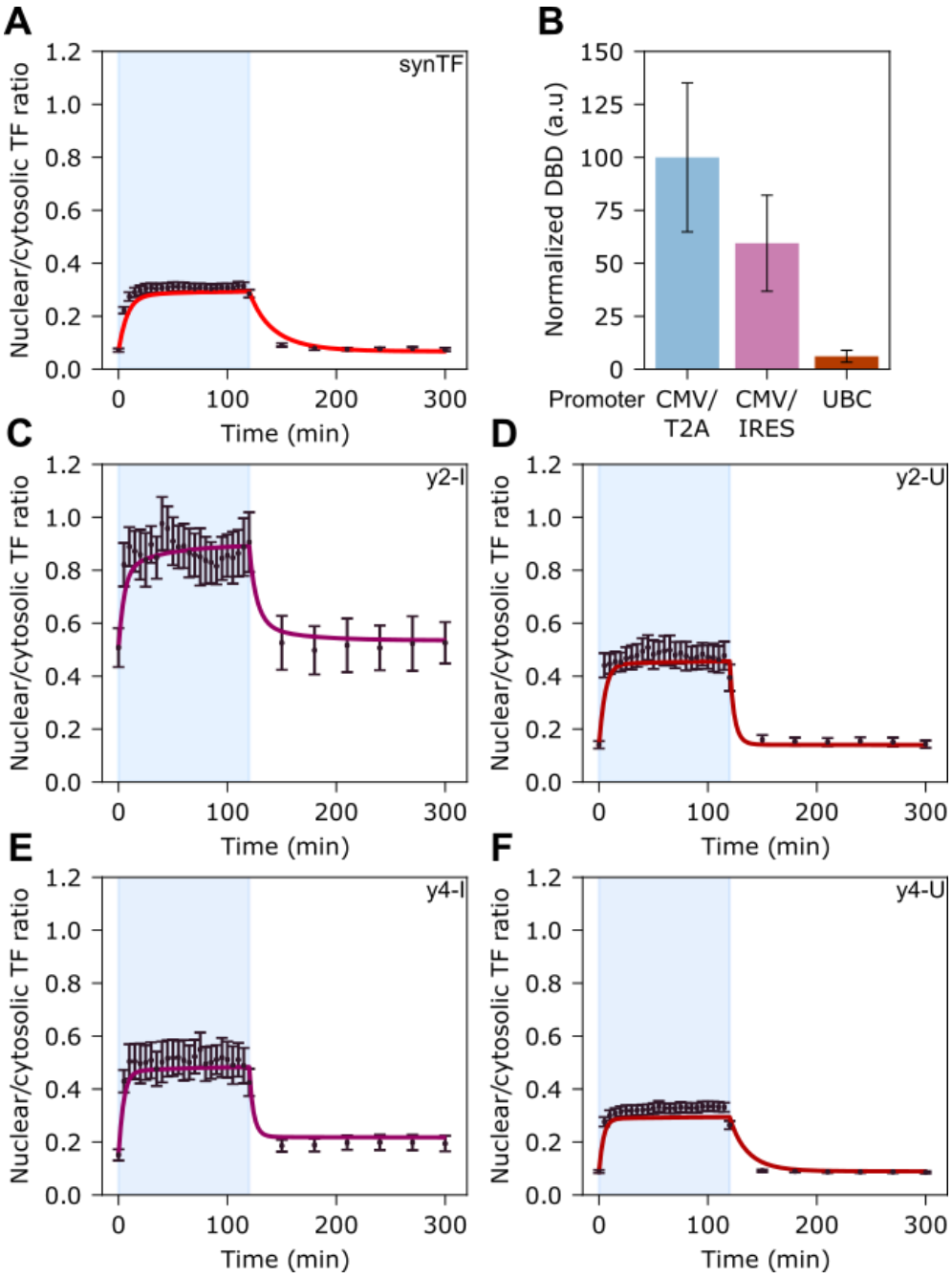

Figure 2.2.2 | Nuclear accumulation of the synYAP is enhanced by stronger interaction with synTEAD and higher nuclear synTEAD levels. A) Nuclear accumulation of synTF upon blue light activation expressed in terms of nuclear to cytosolic signal ratio. B) Expression of the synTEAD under T2A, IRES or UBC promoter. C) Nuclear import of synYAP of y2-I TF (ZIPA = SYNZIP2, ZIPB = SYNZIP1, K_D < 10nM, and synTEAD expressed under IRES). D) The nuclear import was approximately halved when the synTEAD of the y2 TF setup was expressed under the UBC promoter. E) Nuclear accumulation of synYAP of y4-I (ZIPA = SYNZIP4, ZIPB = SYNZIP3, K_D < ~30nM and synTEAD under IRES). Nuclear accumulation upon blue light activation was low due to lower interaction affinity. F) Nuclear import was further decreased when the synTEAD was expressed under the UBC promoter (y4-U). The trend lines represent model fitted data while the black dots with error bars represent experimental data (purple and red denote synTEAD expression under IRES and UBC respectively). All error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

Regardless of the specific setup used for expression of synTEAD, nuclear import of y4 synYAP constructs was lower compared to their corresponding y2 constructs, thus, confirming that synYAP/synTEAD affinity improves nuclear import of synYAP. Also, the nuclear import for y1 constructs was slightly higher than that of the y4 constructs under similar experimental conditions. From these data, I would infer that affinity between synTEAD and synYAP follows this order: y2>y1>y4>y3 which is consistent with the measured K_D of the SYNZIPs [360].

2.2.3 Higher synYAP/synTEAD affinity and lower synTEAD concentration improve the functionality of the TF

The levels of synTEAD was observed to be crucial, since reporter expression after sustained nuclear accumulation of synYAP was 4.7 fold higher for y4-U than y4-I (Figure 2.2.3A). I observed the same trend for y2 (Figure 2.2.3A). I speculated that the reason why the construct with the UBC promoter performs better than the one with the IRES is because there is competition between transcriptionally active synTEAD/synYAP heterodimers and the free transcriptionally inactive synTEAD (Figure 2.2.3B). A higher synYAP to synTEAD ratio in the nucleus leads to a higher concentration of synTEAD/synYAP heterodimers, which can activate reporter gene expression (Figure 2.2.3B top). On the contrary, a lower synYAP to synTEAD ratio in the nucleus leads to a higher concentration of synTEAD not in complex with synYAP, which competes with the synTEAD/synYAP heterodimer population. As a consequence, reporter gene expression is reduced (Figure 2.2.3B bottom). It should however, be noted that there is a lower limit to the level of nuclear synTEAD required for reporter expression (Figure 6.4). Also, increasing synTEAD levels in the nucleus is not the only way to increase the competition for the REs at the promoter. Indeed, I could reduce reporter gene expression by reducing the nuclear import of synYAP by adding the LOVtrap to y1 (Figure 2.2.3C).

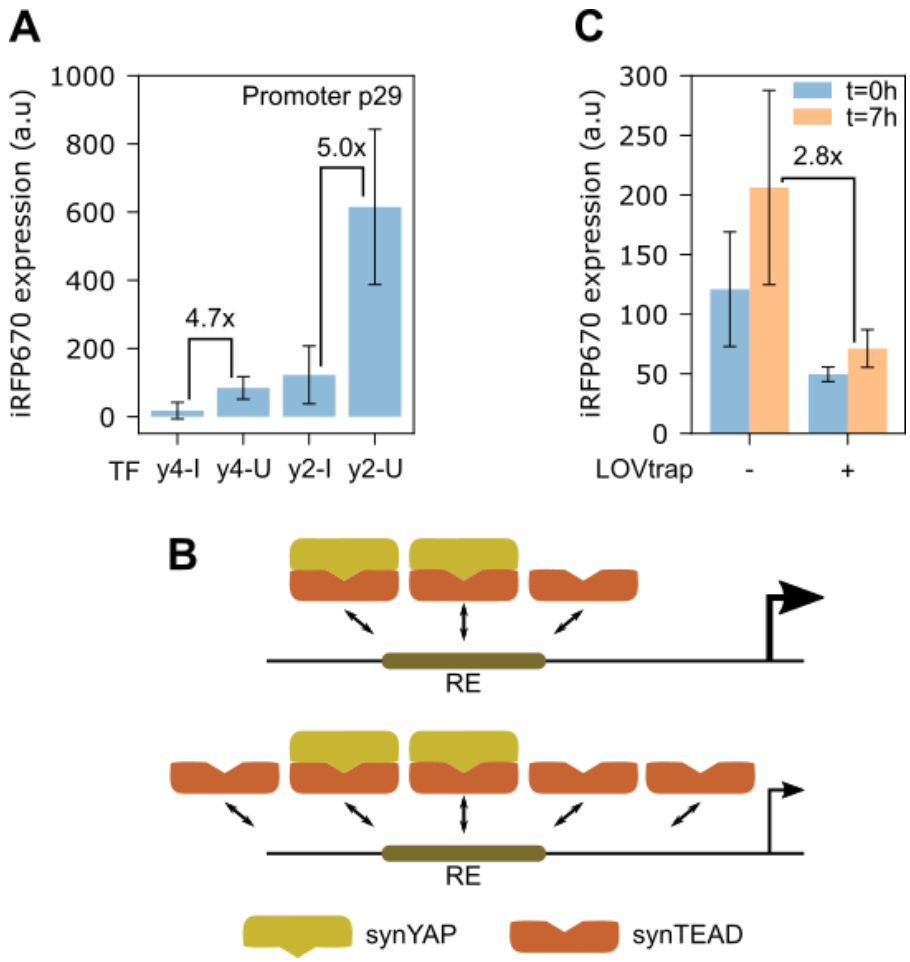

Figure 2.2.3 | Higher interaction affinity and lower synTEAD levels yield higher reporter protein expression. A) Reporter protein expression under the p29 promoter 5hrs post sustained nuclear synYAP accumulation. Reporter protein expression was consistently higher in setups in which synTEAD was expressed under the UBC promoter (y4-U and y2-U) compared to when it was expressed under the IRES (y4-I and y2-I). B) Scheme depicting the competition between transcriptionally active synTEAD/synYAP hetero-dimer and the transcriptionally inactive synTEAD to bind the responsive element. C) Similar competition was observed when synTEAD levels were kept same while varying the synYAP nuclear accumulation levels. The y1 synYAP accumulated less in the nucleus in the presence of the LOVtrap. Error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

2.2.4 synTF outperforms any synTEAD/synYAP TF construct

When I compared their activity, I realized that synTF (see Chapter 2.1) outperformed y1. For instance, with promoter p2, reporter levels were ~7.86 fold higher for synTF than y1 (Figure 2.2.4A). I observed the same with promoters p4, p22 and p29 (Figure 2.2.4A-C). Using the best performing construct for the synTEAD/synYAP TF, namely y2-U, the difference decreases, but visible (Figure 2.2.4A). These results indicate that a strong affinity between the TAD- and DBD-bearing proteins is paramount to the functionality of the TF. Clearly, a TF with both domains on the same protein will function optimally. A weaker affinity between the TAD- and DBD-bearing proteins reduces the effective dwelling time of the transcriptionally active TF at the promoter, thus leading to lower reporter expression.

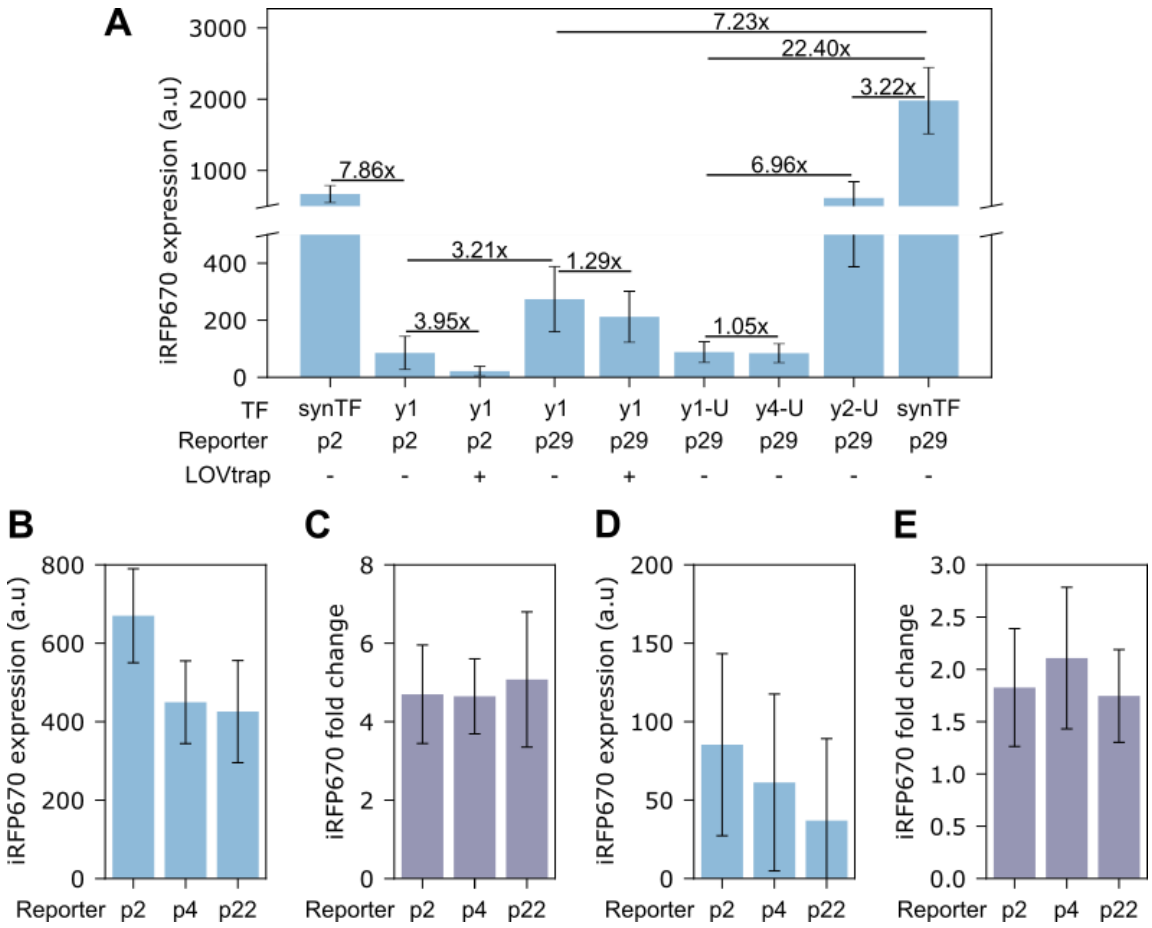

Figure 2.2.4 | Reporter expression induced by synTF is significantly higher than induction by synTEAD/synYAP TFs. A) Under sustained nuclear accumulation, synTF induced the highest reporter gene expression even under the relatively weaker p2 promoter. B) Expression of iRFP670 reporter protein under different promoters induced by synTF. C) Reporter protein expression fold change for (B). D) Expression of iRFP670 reporter protein under promoters induced by y1 synTEAD/synYAP TF setup. E) Expression fold change of reporters induced by the y1 TF. All error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

I also noticed that reporter expression was noisier for y1 than synTF (Figure 2.2.4B&C). Promoter p4, for example, is built using REs that are bound by LexA with weaker affinity, it is therefore not surprising that any further decrease of dwelling time of transcriptionally active TF causes significantly lower reporter expression. I also reasoned that the increase in reporter expression noise could be a reflection of the noise in synTEAD expression. Translation of synTEAD under IRES yields a more variable expression than the CMV promoter which is the case for synTF. The combined variability in synYAP and synTEAD expression therefore impact the observed variability in reporter expression under y1.

2.2.5 SynYAP can be translocated in and out of the nucleus by interacting with a partner protein

Endogenous mammalian YAP is sequestered in the cytosol by interaction with the 14-3-3 cytosolic protein upon phosphorylation at serine 127 (S127) and accumulates in the nucleus when S127 is dephosphorylated [251–254]. To mimic this scenario in my synthetic TEAD/YAP-like constructs, I replaced LINuS with the wild type AsLOV2 domain, which interacts with cytosolic Zdk2 (NES-mVenus-Zdk2) thus, retaining it being retained in the cytosol (Figure 2.2.5A). Zdk2 was used despite the wild AsLOV2 domain because the P2A self-cleaving peptide (Figure 2.2.5B) left extra amino acids on the C-terminus of the Jα helix. The NES-mVenus-Zdk2 therefore acts for synYAP like 14-3-3 for YAP. Upon blue light activation, which mimics the dephosphorylation of YAP at S127, synYAP is released from NES-mVenus-Zdk2 and passively diffuses into the nucleus, where it is retained if there is an interacting partner (synTEAD in my case). Without a nuclear interacting partner, I expect synYAP to equilibrate between the cytoplasm and the nucleus, as seen for GFP alone for instance. In the situation of nuclear retention, synYAP accumulation in the nucleus should proceed gradually, since nuclear import is based on passive diffusion rather than active import. In the dark, the synYAP fraction that diffuses out of the nucleus will be retained in the cytosol due to the interaction with NES-mVenus-Zdk2 (Figure 2.2.5A).

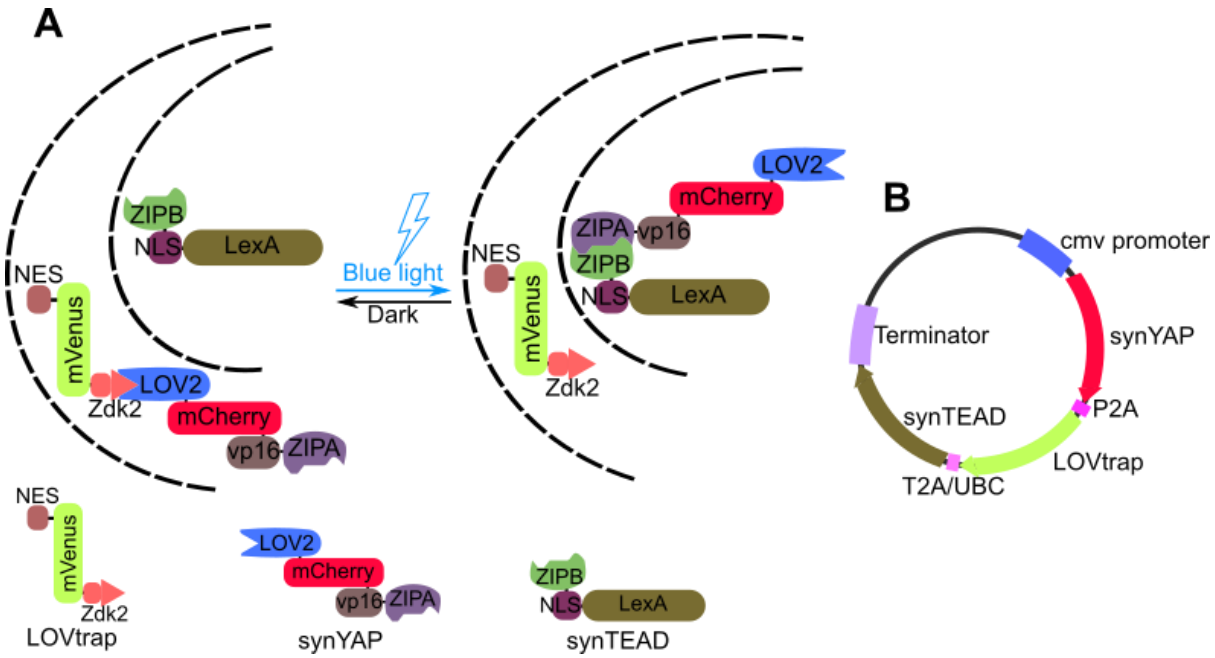

Figure 2.2.5 | Nucleocytoplasmic translocation of synYAP based on interaction with proteins localized to different subcellular compartments. The synTEAD/synYAP TF was constructed and effectively translocated between the nucleus and cytosol without LINuS. A) The TAD bearing co-activator is retained in the cytosol in the absence of external stimulus by interacting with the cytosolic LOVtrap construct. Upon exposure to blue light, synYAP is released and diffuses to the nucleus where it is retained by interacting with the nuclear synTEAD. B) The three fragments together with the reporter were arranged on a single plasmid. Equal levels of the synYAP and LOVtrap was obtained by expressing them under the CMV promoter. Different levels of the synTEAD was achieved by expressing it under the CMV promoter, IRES or UBC promoter.

To experimentally test this system, I cloned all three components on a single plasmid together with the reporter (Figure 2.2.5B). As a proof of concept, I used the synTEAD construct with the T2A peptide to ensure high levels of synTEAD, which should facilitate nuclear retention of synYAP. I also cloned the system with synTEAD driven by the UBC promoter to later assess reporter expression.

I constructed synTEAD with SYNZIP4 and synYAP with SYNZIP3 and wild type AsLOV2 domain which interacted with Zdk2 in the dark (NL3). The synYAP of NL3 was mostly cytosolic in dark when all interacting partner were present. It was however, nuclear when Zdk2 was removed (Figure 2.2.6A) and remained mostly cytosolic when the SYNZIP3 heterodimer subunit was removed. Quantification of nuclear accumulation of synYAP of NL3 with or without synZIP3 were however similar (Figure 2.2.6B&C). This suggested that the nuclear accumulation of synYAP was mainly as a result of diffusion and not retention via the interaction with synTEAD. This is consistent with our earlier observation synYAP constructed with SYNZIP3 had little or no interaction affinity for synTEAD constructed with SYNZIP4. Nuclear accumulation of synYAP was however higher both in the dark and upon blue light activation when synYAP was constructed with SYNZIP4 and synTEAD was constructed with SYNZIP3 (NL4) (Figure 2.2.6A&D). As previously explained (section 2.2.2), this is perhaps due to better interaction conformation of the SYNZIPs. Nuclear accumulation of synYAP of NL4 increased steadily from the beginning of blue light activation to the end due to the gradual diffusion of

synYAP from the cytosol to the nucleus and subsequent nuclear retention by interacting with synTEAD (Figure 2.2.6D).

I then constructed synYAP with SYNZIP2 and synTEAD with SYNZIP1 (NL2) with the expectation that synYAP of NL2 will accumulate better in the nucleus. I however, observed that synYAP of NL2 had low nuclear accumulation both in the dark and upon blue light activation (Figure 2.2.6A&F). Further, probing revealed that the synYAP of NL2 was less stable and that its protein expression level was about 40-50% of synYAP of NL4 (Figure 2.2.6E). This implied that there was about twice Zdk2 fusion protein as the synYAP of NL2 resulting in an efficient retention of synYAP in the cytosol even in the presence of blue light. This explained the lower nuclear accumulation of synYAP of NL2 compared to NL4 (Figure 2.2.6D&F). Expressing synTEAD of NL2 under UBC promoter (NL2-U) resulted in further decrease in the nuclear accumulation of synYAP (Figure 2.2.6G). In addition, despite the high interaction between the units of NL2-U, nuclear import of its synYAP was below that of synYAP of NL3 in the dark when there was no SYNZIP3.

Consistent with observations made on the system with LINuS, constructs with very high synTEAD expression (expressed under T2A) showed no or very low reporter expression both in the dark and after blue light activation. The only construct that was able to induce some level of reporter expression was NL2-U in which the nuclear accumulation of synYAP was the lowest (Figure 2.2.7).

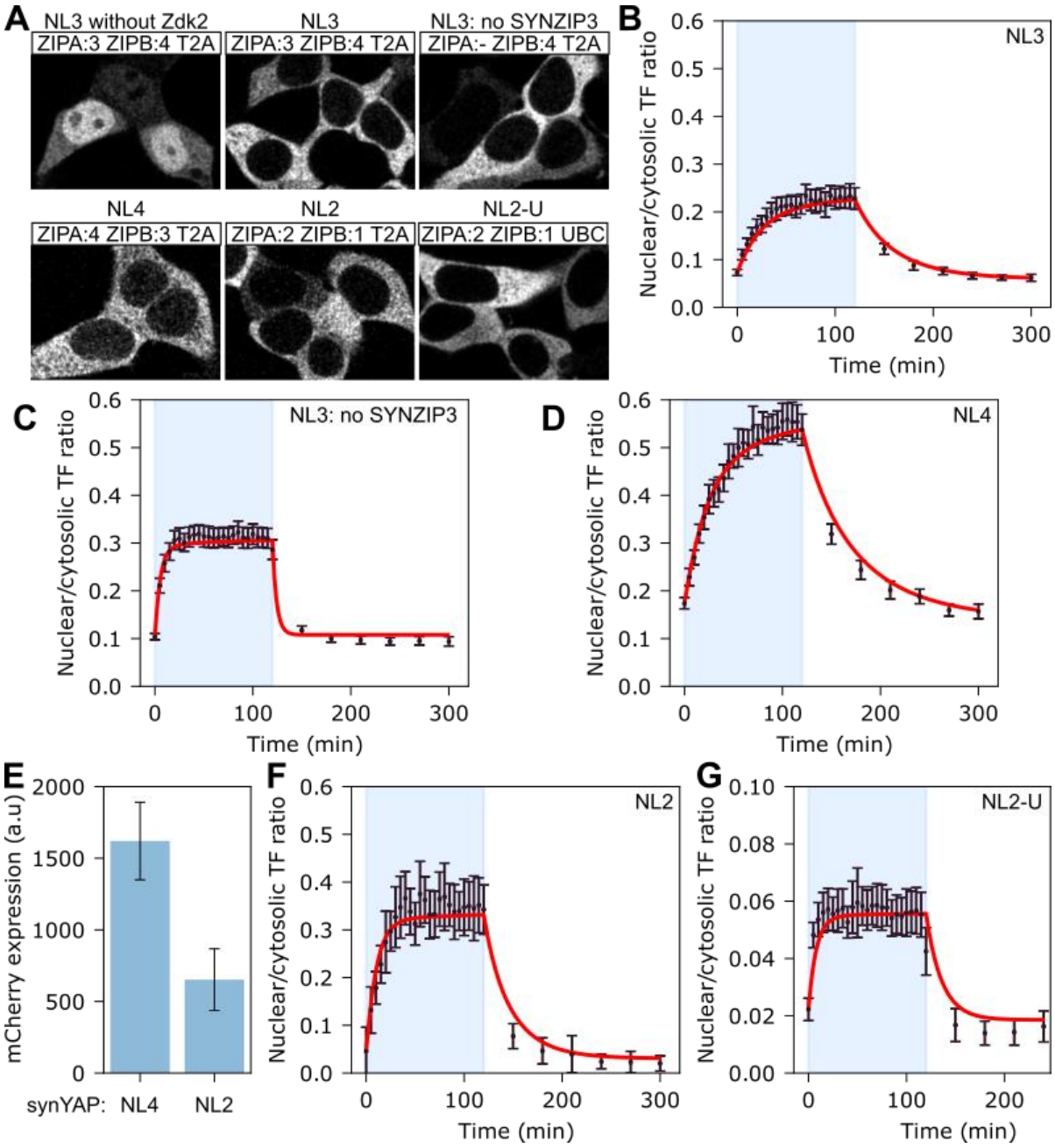

Figure 2.2.6 | Higher ZIPA and ZIPB affinity facilitates nuclear accumulation of synYAP in the absence of any NLS. A) Representative microscopy images showing the level of nuclear accumulation of the synYAP prior to blue light exposure. The synYAP construct accumulates in the nucleus in the absence of cytosolic interacting partners (NL3 without Zdk2). B) With very weak interaction with synTEAD, synYAP showed a decent accumulation in the nucleus. C) Similar level of nuclear accumulation was however observed when the SYNZIP3 on synYAP was deleted thus, preventing any interaction with the nuclear synTEAD. D) Nuclear accumulation of synYAP increased by about two folds when SYNZIP4 and SYNZIP3 in the place of ZIPA and ZIPB respectively were used. E) Expression of synYAP of NL2 (whose ZIPA and ZIPB were SYNZIP2 and SYNZIP1 respectively) was about 2 folds lower than the expression of NL4 synYAP. In effect, the synYAP of NL2 was about half the level of LOVtrap. F) Nuclear accumulation of synYAP of NL2 was lower compared to that of NL4 despite having higher affinity to synTEAD. G) Nuclear accumulation of NL2 synYAP was further decreased when the synTEAD was expressed under the UBC promoter (NL2-U). Very low nuclear accumulation level of synYAP of NL2-U was achieved due to the high LOVtrap to synYAP ratio and low synTEAD expression.

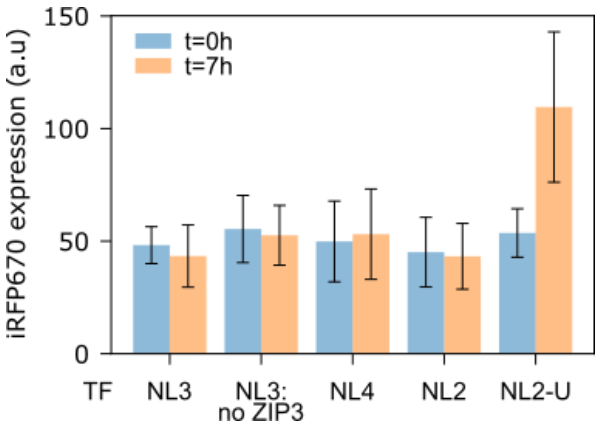

Figure 2.2.7 | Reporter expression induced by synTEAD/synYAP was significantly lower without nuclear localization signal. High interaction affinity between NL2 TF units induced low reporter gene expression when synTEAD expression was lowly expressed.

3 DISCUSSION AND OUTLOOK

This study assessed how information encoded in TF dynamics is decoded by promoters of different target genes and how the elements in the promoter contribute to differential expression in response to different dynamics. Gene expression regulation by a TF bearing both DBD and transactivation domain in a single protein such as p53 and a TF that bears only the DBD and as such requires interaction with a co-activator as in the case of TEAD/YAP were compared. To be able to study such systems with the least complexity arising from regulation by other activated pathways, synthetic expression circuits consisting of a TF and a reporter were generated for both studies. Data from the TF dynamics studies was combined with mathematical modelling to better under the role of promoter elements and their impact on transcription initiation and how elements in the 5' UTR affect translation initiation.

The results and their interpretation in the context of other TF dynamics studies and how different promoter elements impact transcription initiation rate and thus expression of different target genes are discussed in this section. I will also discussion how different TF designs regulation of expression of their target genes.

3.1 Transcription factor dynamics

3.1.1 Choice of experimental setup for studying TF dynamics impacts interpretation of data

Most previous studies of TF dynamics have used external stimuli or chemical agents that modulate the activity of a molecule upstream of the TF of interest to either trigger certain natural dynamics [210,211,214,218,242], modify the natural dynamics into something else (for instance pulses into sustained signal; [218]), or impose desired dynamics ([201,202]). However, when using chemicals or natural ligands, it is not clear which other pathways may be activated, moreover post-translational modifications of the TF take place, contributing to the activity of the TF in ways that may be hard to dissect and disentangle from pure dynamics. For example, by administering the Mdm2 inhibitor Nutlin-3 [218] at the right time points and doses to cells previously subjected to γ-radiation, it is possible to 'transform' the natural pulsatile p53 dynamics in these cells into a sustained one [218]. The complete side effects of both γ-radiation and Nutlin-3 on the global translation and transcription rates are not known. For instance, Nutlin-3 has been treatment has been shown downregulate genes that are not even targets of p53 [362]. Optogenetics is expected to help overcome non-specific activation of other molecules or pathways. However, using optogenetics tools such as Opto-SOS to induce molecules upstream of the TF rather than the TF itself [158,245], does not allow studying solely the impact of TF dynamics. In my opinion, the best set up so far used to study TF dynamics using optogenetics was implemented using the CLASP optogenetic tool that regulates the nuclear localization of the yeast transcription factor Crz1 [248]. This set up very closely resembles the one I used in my study. In this case, the translocation of the TF of interest was directly regulated without triggering the activation of other pathways. The focus of this study, however, was not to specifically understand promoter decoding of TF dynamics.

The study that has so far focused on understanding the mechanistic details of how a promoter decodes TF dynamics has used the small molecule 1-NM-PP1 to induce PKA-mediated phosphorylation of the yeast TF Mns2, which results in its nuclear accumulation [201,202,244]. The mechanistic insights gathered in yeast using this system are extremely interesting [201,223],

however it is fair to ask whether modifying PKA affects global gene expression considering how important it is in regulating cellular processes [363,364]. This may have a plethora of effects that could affect transcription of individual genes independent of Msn2 dynamics. To avoid such complexity of modulating a molecule upstream of a TF, I decided to use optogenetics to directly regulate the nuclear translocation of my TF (Figure 2.1.1). In the end, I decided to use a synthetic TF and reporter circuit to understand how TF dynamics are decoded by mammalian promoters without having to consider any confounding factors that may not be fully understood. I considered this to be a good system because, aside chromatin remodelling, transcription of a reporter gene from a plasmid requires the same cellular machineries as transcription of an endogenous gene from the chromosome. This experimental setup also permitted the direct testing of different promoter elements and parameters without having to construct tens of cell lines. With this approach, I could more easily test the role of individual promoter components without worrying about the contribution of other factors, such as distant enhancers for instance.

Naturally occurring TF dynamics do not necessarily have the same area under the curve of nuclear TF nor do they have the same activation duration [198]. However, to compare two TF dynamics for better understanding of how they affect expression of target genes, parameters such as nuclear TF amplitude, cumulative TF levels and how long experiments are run have to be properly calibrated. A promoter might indeed rather sense the cumulative TF levels and not dynamics per se. For most studies where TF dynamics have been generated using natural stimuli, data collected from experiments with the same duration have been compared [198,201,202,208,209,223,235]. I find that, in such cases, the interpretation of the results might be misleading since the differences in the amplitude and cumulative levels of the TF in the various dynamics were not considered. Comparing data from same amplitude dynamics at same experimental duration suffer from the fact that cumulative levels are different [201]. In the Crz1 optogenetics study, the authors used same area under the curve and same experimental duration [248]. This implies for a promoter in which both sustained and pulsatile dynamics induced transcription, mRNAs expressed at the end of the shorter sustained dynamics will have more time to produce proteins, which in turn may have longer time to mature compared to mRNAs and proteins produced at the end of the longer pulsatile dynamics.

I constructed my experiments to have the same area under the curve for the TF in the different dynamics and ended them all at a fixed waiting time not to give unfair expression advantage to any specific dynamics (Figure 2.1.8). Like this, I am sure to observe differences triggered solely by the dynamics.

3.1.2 Promoters requiring high TF amplitudes or long TF dwelling time distinguish well between different TF dynamics

I found that sustained and pulsatile TF dynamics are distinguished better by promoters that require high TF amplitude and/or long dwelling time to get activated than by promoters that are activated already at low TF levels or have fast response. The presence of an amplitude threshold allows effectively filtering out low amplitude signals while slow response filters out short bursts of nuclear TF.

3.1.2.1 Responsive element affinity dictates the threshold of nuclear TF amplitude a promoter responds to

At a fixed promoter response time, the affinity of the TF for the RE dictates how much transcriptionally active nuclear TF is needed to induce reporter expression. A high-affinity RE can lead to induction of the downstream gene even from low-amplitude TF signals [365]. Consequently, fast responding promoters with very strong REs cannot efficiently distinguish TF dynamics when the area under the curve of all dynamics remain the same (Figure 2.1.9B-D). For some promoters, the low TF signal present in the nucleus in the dark phase prior to the start of the experiment (what I call background signal) is already capable of eliciting a strong response. Assuming REs in such a promoter are already saturated at the low TF levels, further increase in the nuclear TF amplitude becomes redundant [366–368].

In contrast, promoters with low-affinity REs require higher nuclear TF concentration to be able to induce significant expression of the reporter gene [202,235,369]. The amount of information decoded by such promoters is however limited by how efficient the coupling between TF binding and transcription initiation is. With a low transcription initiation rate, these promoters are very noisy unless the signal amplitude is significantly higher than the threshold. The gene expression noise in such promoters then causes information loss during the decoding process [201,365].

Promoters with low affinity REs displayed differential reporter expression in response to different TF amplitudes but only when the transcription initiation rate as inferred from the strength of the TATA-box was high (Figure 2.1.11D, Figure 2.1.13D). This is in accordance with the observation that the residence time of a TF at the promoter (which is defined by its affinity for the RE) has no effect on expression noise at high transcription initiation rate [365]. Beyond amplifying expression noise, low transcription initiation rate combined with low-affinity RE leads to delayed expression in addition to increased sensitivity to amplitude [201,367]. This is because a promoter with low-affinity RE and low transcription initiation rate requires even higher nuclear TF to be able to increase the probability of transcription initiation.

3.1.2.2 TATA-box strength determines transcription initiation efficiency and refractoriness of response

The transcription initiation rate or response time has been found to be a property of the core promoter [200]. In this study, the core promoter is defined by the TATA-box. By inference, the transcription initiation rate is dependent on the affinity of the transcription PIC for the nucleator at the core promoter. A stronger nucleator, therefore, leads to higher activation rate or short response since it positively influences the recruitment of the transcription machinery [90,370,371] and subsequently transcription initiation. Efficiency of PIC nucleation determines how effectively TF binding to the RE is converted to transcription initiation. Consequently, a high affinity of PIC subunits for a nucleator should lead to higher coupling of RE binding and transcription initiation. Higher coupling efficiency means high nuclear amplitude and/or sustained TF signal will generate higher mRNA transcription. Similarly, longer duration of dynamics should lead to higher reporter expression (Figure 2.1.9B).

I observed that reducing the affinity of the nucleator introduces refractoriness into the reporter expression kinetics (Figure 2.1.12B). I therefore reasoned that, a promoter with lower transcription initiation rate may also have longer refractory time. Conversely, promoters with shorter response time should exhibit less refractory behaviour (Figure 6.2). Fast responding

promoters with short refractory time should therefore effectively integrate the cumulative TF levels.

If low affinity PIC nucleators lead to longer refractory times, then it is expected that a promoter with a weak nucleator shows an initial overshoot of transcription before decreasing to steady state during sustained activation (Figure 2.1.12). The decrease of expression from the initial peak to the steady state may then be explained by the fact that some promoters get into the refractory state. Since slow response promoter have longer refractory time, the total number of cells transcribing starts to decrease till it reaches equilibrium where promoters going into refractory and those leaving the refractory state are equal [200,372]. In contrast, promoters with high affinity nucleators will spend less time in the refractory state making the steady state and the peak expression value more similar [372]. Refractoriness of promoters with low affinity nucleator can also mean that there is a slow recycling of some components of the PIC and that the strong initial transcription burst leads to decrease in the local concentration of such components to levels that do not permit the formation of transcriptionally competent PIC on the weaker TATA-box. PIC may also readily dissociate after the initial round of transcription initiation which will then take some time to be reconstituted thus impeding fast re-initiation. This might explain the lower re-initiation rate observed for the p53 target gene – p21. The p21 gene contains a non-consensus TATA-box in its core promoter [234].

3.1.2.3 The length and sequence of the DNA stretch between the TATA-box and the responsive elements dictates looping efficiency

As I wrote above, TF amplitude is sensed by the REs. The strength of the core promoter influences instead the response time of a promoter. For a fast response and/or low amplitude threshold promoter, the efficiency of looping between the TF-bound RE and the core promoter determines whether the promoter will distinguish between different TF dynamics. Varying the length of the DNA stretch between the core promoter and the RE reduces the probability of converting TF binding into transcription initiation. Low efficiency of DNA looping therefore leads to less coupling between TF binding and transcription initiation. Gene expression mediated by a TF requires the looping of the DNA between the RE(s), at which the TF is bound, and the core promoter, to successfully get started [112,113]. Hansen and O'Shea, and C. Li et al. observed that the distance between the RE and the core promoter contributes to differential expression in response to TF dynamics [200,223]. In my experiments, I saw that increasing this distance by 147 bp (length of DNA to wind a nucleosome) led to over 80% decrease of reporter expression (Figure 2.1.15E&F), which confirms the impact of looping efficiency on gene expression. This can be explained by the fact that looping efficiency between two points on a stretch of DNA is inversely proportional to the distance (in bp) between them [114].

Promoter accessibility and nucleosome occupancy of the DNA stretch between the RE(s) and the core promoter have also been shown to be key for differential expression of target genes in response to different TF dynamics [201]. Nucleosomes on their own serve to inhibit gene expression. Although I tested nucleosome occupancy on plasmid and not within the chromosomal context, I am sure my observation is relevant since nucleosome positioning by the 601 sequence on a plasmid has been observed in cells in different studies [373,374]. I expected that the presence of the nucleosome positioning sequence would yield lower gene expression compared to a random sequence. This is because in addition to ability to position

nucleosome, it had the same length as the random sequence. I however observed the opposite (Figure 2.1.14D, Figure 2.1.15E&F). These data cast doubt on the contribution of nucleosome positioning to TF dynamics decoding. Earlier work on the 601 sequence has shown that it is more bendable, thus causing it to have higher looping efficiency compared to a random sequence [119]. Although this may explain why the nucleosome positioning sequence I used does not efficiently decouple TF binding and transcription initiation, it does not exclude that nucleosome positioning contribute to decoding TF dynamics. I did not directly prove the presence of a nucleosome on my plasmids. That notwithstanding, high and efficient nucleosome positioning is generally associated with highly flexible DNA, which can readily form curved structures [375–377], which is consistent with what is known about the 601 sequence. The fact that the 601-bearing promoter did not show any differential reporter expression in response to the different dynamics (Figure 2.1.15E&F) perhaps because its bendability skews loop formation thus reducing the randomness in the looping process. The decrease in looping efficiency therefore caused inefficient conversion of TF binding to transcription initiation but without increase in the randomness. This would imply that the probability of transcription initiation although small remains more or less fixed for all the dynamics.

I reason therefore that DNA looping efficiency must be below a fixed threshold to effectively contribute to TF dynamics decoding. However, since decreasing DNA looping efficiency significantly impacts gene expression, TF binding affinity and transcription initiation rate must be high for meaningful decoding of TF dynamics.

3.1.3 Translation initiation and mRNA structure affect target gene expression in response to different TF dynamics

TF dynamics studies on p53 [241] and NF-κB [242] have shown that transcripts of some early response genes are very unstable. The instability of these mRNAs has been shown to be intrinsic to the gene sequence and not due to some extra layer of regulation. My data suggest that, indeed, the low stability is intrinsic to the sequence, but may be a consequence of low translation rate (Figure 2.1.16B-D). Stability and translatability of mRNAs are intertwined processes. Due to competition between translation initiation and mRNA degradation, high translation initiation rates protect mRNA from decaying thus conferring stability to otherwise unstable mRNAs [195,378]. In contrast, mRNAs that are translated at low rate are prone to degradation, and produce less protein during their life-time. Definitely, mRNA stability is an intrinsic property to gene sequences as observed for some target genes of NF-κB and for stress response genes in yeast cells [242,243]. I argue that instability of transcripts may result from lack of protection due to too low translation rate. I do acknowledge that more experiments need to be done to confirm the low translatability of the unstable transcripts of these NF-κB response genes.

That notwithstanding, continuous expression of mRNAs that are transcribed at low rate is required to ensure higher accumulation of the final protein products. Continuous mRNA production increases the chance of translation initiation on some transcripts before they are degraded. Under pulsatile dynamics, lower mRNA production over a long period reduces the number of mRNAs being translated at any given time. For p53, the short mRNA half-life of some early response genes has been shown to produce pulses of gene expression at the transcript level [241]. When combined with low protein stability, low translation rate can enhance

small differences in mRNA production rate. The finding that translation can modulate differential gene expression of a target gene at the protein level suggests that two target genes of the same TF with different translation rates can show differential response to different TF dynamics regardless of the fact that they may have identical promoter elements (Figure 2.1.16 and Figure 2.1.10). Transcript levels have been shown repeatedly not to directly correlate with protein levels [379–382]. For two target genes under the same promoter, differential gene expression can arise not only from different translation rate but also from different mRNA processing [383,384].

In conclusion, targets genes with high mRNA expression may be coupled with low translation rate to be able to achieve differential protein expression in response to different TF dynamics. Since translation is an amplification step, small differences in transcript levels arising from different TF dynamics can be amplified at the protein level if the transcripts are not efficiently translated. Finally, it is possible to predict whether a promoter will respond differentially to different TF dynamics by evaluating the strengths of the REs and the core promoter and the distance between them. Yet at the protein level, the translation efficiency must also be considered.

3.1.4 Higher reporter expression by pulsatile TF dynamics: is it possible?

In the absence of any extra layer of gene expression regulation, I claim that a promoter that responds to pulsatile signals will equally or better respond to a sustained one. This is because a promoter that responds to pulses has either fast activation rate or low amplitude threshold and these same parameters will also favour target gene expression under sustained dynamics. As observed in studies of ERK dynamics, there is no known gene that responds better on it's own to ERK pulses than it does for sustained ERK activity [158]. However, in the presence of the negative feedback regulation of ERK, some immediate early genes (IEGs) show higher expression under pulsatile compared sustained nuclear ERK dynamics. The continuous expression of the negative regulator of ERK causes dephosphorylation of nuclear ERK thus reducing the activity of ERK (adaptation). The higher expression observed under pulsatile dynamics can therefore be attributed to the low sensitivity of such IEGs to ERK. For IEGs with high ERK sensitivity, even the low activity of ERK after adaptation could still induce a good level of expression thus leading to better expression under sustained dynamics.

Using a strong promoter, I observed a higher expression under pulsatile compared to sustained TF dynamics (Figure 2.1.9). I reason that such observation was due to the sensitivity of the promoter to the TF. In the absence of any adaptation, a promoter that is highly sensitive to the TF will still have a high expression even in the presence of low TF concentration. This therefore leads to good reporter expression even during the phases when the TF gets exported out of the nucleus. During the longer experimental time I used for the pulses, therefore, more protein was expressed at the end compared to what is made during the shorter experiment used for the sustained signal. Although this may seem like a drawback of our experimental setup, it makes any differential expression observed for any promoter under sustained and pulsatile dynamics more impressive.

Similarly, the GYP7 promoter was seen to induce higher reporter expression under pulsatile compared to sustained Crz1 dynamics [248]. This experiment was also performed to have the

same area under the curve for the different dynamics and, as such, it also implied a longer activation phase for the pulses. I speculate that Crz1 has high affinity for the GYP7 promoter, which causes the promoter to be active even in the phases during which Crz1 is exported out of the nucleus. One may ask if GYP7 has higher sensitivity to Crz1, why then does it show lower expression than other tested promoters that showed better expression under sustained dynamics? I argue that the affinity of the TF for the RE determines the threshold of TF sensitivity, but the number of REs can limit the achievable expression. In my library, promoter p5, which has half the number of REs as promoter p1, nicely exhibits similar property to GYP7: it corresponds to higher reporter expression under pulses compared to sustained dynamics, but leads to lower expression than promoter p4 for example.

3.1.5 A theoretical exercise: modifying the mathematical model to obtain higher reporter expression under pulsatile dynamics

I concluded from my experimental and mathematical modelling data and literature searched that higher reporter expression under pulsatile TF dynamics is not readily achievable under physiological conditions without extra layers of regulation [385]. However, it is theoretically possible to achieve higher mRNA and protein expression under pulsatile TF dynamics when considering TF, promoter, TF binding affinity and transcription initiation rate as the only parameters without extra layers of regulation. What is needed is a TF that inhibits gene expression once above a certain concentration. The promoter should efficiently induce gene expression at low TF concentrations and have a fast response so that expression of the downstream gene occurs before there is too high TF that acts as inhibitor. To prove this idea, I modelled such a system using a four-state promoter model. The four states are: unbound, bound/active, refractory and inhibited state triggered by higher TF concentration (Figure 3.1A).

I found that, for such system to efficiently decode pulsatile activation, the inhibited state should have an intermediate stability to allow for strong promoter inhibition while quickly reversing under low concentrations of TF (Figure 3.1B). This type of promoter will however fail to completely filter out sustained activations, but will rather show some response that declines over time (Figure 3.1C). Nonetheless, this TF/promoter setup would induce significantly higher gene expression under pulsatile activation compared to sustained TF dynamics (Figure 3.1D&E), given the same cumulated stimulus.

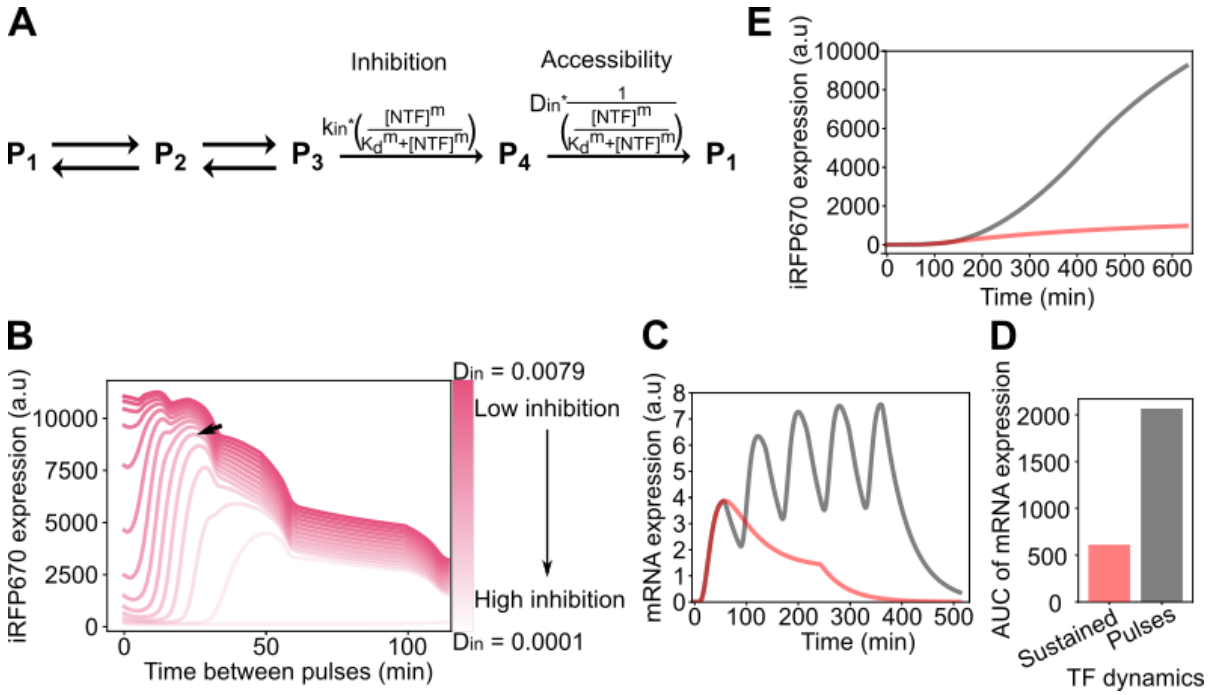

Figure 3.1 | Theoretical prediction of a promoter that induces higher reporter expression under pulsatile TF dynamics. Promoter with a fourth but inhibited state can theoretically induce higher expression under pulsatile activation compared to sustained TF dynamics. A) Scheme of a promoter with an inhibited state which is induced by higher TF concentration. B) Mathematically predicted reporter protein expression as a function of the time between pulses (dark or 'OFF' duration between two blue light pulses). Different dissociation rates of the inhibiting complexes (D_{in}) leads to differential reporter expression under pulsatile activations with varying periods. Period of pulsatile activation increases with increasing time between pulses. C) Nascent mRNA expression. D) Area under the curve of nascent mRNA expression induced by the different dynamics. E) Reporter protein expression under the different TF dynamics. Red and grey lines in C and E represent model predicted data for sustained and 55-55 min pulsatile nuclear TF dynamics.

3.1.6 Outlook on TF dynamics

My synthetic circuits do not completely recapitulates endogenous systems. Since it does not embrace the full complexity of the mammalian gene expression, it is more or less an isolated system. It can therefore be argued that this system is not very representative of mammalian gene expression response. That notwithstanding, the simplicity of this system makes it ideal to study the intricate details of decoding transcription factor dynamics since one can focus on the little details that may be overlooked when studying a very complex system. I believe that observations made in this study are applicable in explaining many observations on expression of endogenous genes in response to TF dynamics as demonstrated. Aside, my contradictory observation on nucleosome positioning which is not in direct agreement with earlier studies, all other observations can applied in explaining observations made in yeast and mammalian expression systems. TF dynamics affect cell fate and can therefore be implicated in human diseases. Detailed mechanistic understanding of TF dynamics is therefore required. I propose for further studies, to assess how the translatability of endogenous target genes contribute to TF dynamics. I deem that necessary since proteins of target genes are mostly the effector molecules that contribute to cell fate.

3.2 TF architecture

3.2.1 Strong TEAD/YAP interaction is necessary for effective reconstitution of a functional TF

Most heterodimers in the synthetic biology toolbox have very low interaction affinity (K_D > 100 nM) (Chien et al., 2007; Kawano et al., 2015; [386]. This makes their use for the construction of the TEAD/YAP-like system which can induce reporter expression comparable to synTF impossible. The ability of the TF to induce expression of its target genes is dependent on the dwelling time and the potency of the TAD. The DBD-bearing TEAD already has a dissociation constant, so the weakly interacting coactivator decreases the effective dwelling time of the TF complex. The combination of two dissociation events therefore requires that either the promoter of the target gene or the interaction between the two domains be very strong. Mammalian TEAD and YAP interact strongly with a K_D between 16-38 nM [277,387], which supports the notion that the interaction between the DBD-bearing TF like TEAD and its TAD-bearing coactivator like YAP needs to be strong to induce gene expression. The strong interaction between the two proteins ensures that the complex is stable enough to permit induction of gene expression.

3.2.2 Diffusion and protein-protein interaction drive YAP nuclear translocation

The subcellular localization of YAP is generally regulated by association with other proteins such as 14-3-3 in the cytosol or TEAD in the nucleus [251–254]. I could nicely recapitulate this mechanism of retention into a particular subcellular compartment with my completely synthetic synTEAD/synYAP TF (Figure 2.2.2). This makes my setup relevant for studying the properties of the natural TEAD/YAP, with the advantage of being a much simpler and easier to manipulate system. My setup allowed me to draw another conclusion, which I did not see reported in the current literature: the level of nuclear YAP may be directly dependent on the level of TEAD.

YAP S94A mutant lacks the ability to bind to TEAD and thereby has reduced nuclear accumulation [279]. In contrast, the YAP S127A mutant cannot be phosphorylated at the site that permits interaction with 14-3-3 leading to decreased cytosolic retention [388]. In these studies, compartments in which interaction were inhibited were still not completely devoid of the YAP co-activator. Of course, YAP interacts with other proteins other than its main interacting partners and that easily explains why compartments were not devoid it. I observed that, in the absence of any other interacting partner, synYAP is not completely absent in a compartment after interaction with its partner there is inhibited (Figure 2.2.2C&E). I speculate that, despite the rather large size (~80kDa), YAP passively diffuses in and out of the nucleus. This was clearly demonstrated by constructing the synTEAD/synYAP system, in which synYAP lacks any targeting sequences but still showed nuclear translocation solely dependent on its interaction with synTEAD. The nuclear accumulation in the absence of an NLS and a nuclear interacting partner can only be explained by passive diffusion. This explanation should also hold true for the observations made for the YAP S94A mutant [279]. Both synYAP and native YAP are composed of about 500 amino acids therefore making our observations relevant in explaining the mechanism of nucleocytoplasmic translocation of YAP protein in mammalian cells.

3.2.3 Intermediate TEAD levels drive better expression of target genes

I observed that, reducing the concentration of synTEAD helps to improve expression of the target genes. To the best of my knowledge, such observation has not yet been described in the literature for the natural TEAD/YAP TF. The observation actually contradicts the fact that increasing the nuclear synTEAD expression would enhance the nuclear accumulation of synYAP, thus increasing the population of the transcriptionally active synTEAD/synYAP complex. I therefore reason that the high synTEAD expression also increases the ratio between free synTEAD and synTEAD/synYAP heterodimers in the nucleus. The high free synTEAD outcompetes the synTEAD/synYAP population in binding to the REs thus inhibiting expression of the reporter (Figure 2.2.3C). Since this phenomenon was observed for more than one promoter, I reasoned that this is not an artefact of the synthetic system. Moreover, I observed that there is a limit to how low synTEAD expression can be to induce sufficient reporter expression (Figure 6.6). It has been reported that YAP concentration limits the activity of TEAD [20]. I confirmed this claim by modulating synYAP nuclear import. Lowering synYAP nuclear import negatively affected reporter expression confirming the importance of free versus YAP-bound TEAD ratio. YAP has not been reported to directly modulate TEAD binding to DNA; it is therefore logical that free TEAD molecules will compete with TEAD/YAP heterodimer for binding to the TEAD binding site. It has been shown that competitive binding of different TFs to the same DNA sequence is useful in regulating cellular processes such as erythroid/megakaryocyte switching [389]. It is therefore not farfetched to assume that it may have a role in regulating cell proliferation or differentiation.

To assess the relevance of this observation, it would be best to test this using endogenous TEAD/YAP. The YAP concentration can be kept constant while that of TEAD should be increased to see the effect on the expression of target genes. It will be interesting to know how the cells respond to high TEAD expression without increase in YAP.

3.2.4 Relevance and outlook of synthetic TEAD/YAP studies

Due to synthetic nature of the setup, the cell might regulated their abundance differently since they serve no physiological purpose. Furthermore, the protein were overexpressed despite efforts to reduce their expression. That notwithstanding, confirming whether higher TEAD concentration is a viable way to reduce expression of YAP target genes should be tested using endogenous proteins. Elucidating how the TEAD/YAP TF architecture affects gene expression is required to fill in the missing details of how TEAD/YAP regulate their target genes. How does this TF architecture contribute to gene expression noise? Are the target genes of this TF design capable of sensing TF dynamics? Can mammalian cells encode information in the dynamics of this TF in the first place? These are questions that remain to be answered using this system.

4 REFERENCES

1. Zhu, J., He, F., Hu, S. & Yu, J. On the nature of human housekeeping genes. *Trends in Genetics* **24**, 481–484 (2008).
2. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes are compact. *Trends in Genetics* **19**, 362–365 (2003).
3. Smale, S. T. Selective Transcription in Response to an Inflammatory Stimulus. *Cell* **140**, 833–844 (2010).
4. Hamilton, T. *et al.* Cell type- and stimulus-specific mechanisms for post-transcriptional control of neutrophil chemokine gene expression. *J. Leukoc. Biol.* **91**, 377–383 (2012).
5. Molina, N. *et al.* Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20563–20568 (2013).
6. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
7. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: Function, expression and evolution. *Nature Reviews Genetics* **10**, 252–263 (2009).
8. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
9. Weake, V. M. & Workman, J. L. Inducible gene expression: Diverse regulatory mechanisms. *Nature Reviews Genetics* **11**, 426–437 (2010).
10. Courey, A. J. & Jia, S. Transcriptional repression: The long and the short of it. *Genes and Development* **15**, 2786–2796 (2001).
11. Latchman, D. S. Transcription factors: Bound to activate or repress. *Trends in Biochemical Sciences* **26**, 211–213 (2001).
12. Gaston, K. & Jayaraman, P. S. Transcriptional repression in eukaryotes: Repressors and repression mechanisms. *Cellular and Molecular Life Sciences* **60**, 721–741 (2003).
13. Mapp, A. K., Ansari, A. Z., Ptashne, M. & Dervan, P. B. Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 3930–3935 (2000).
14. Thiel, G., Lietz, M. & Hohl, M. How mammalian transcriptional repressors work. *European Journal of Biochemistry* **271**, 2855–2862 (2004).
15. Herschbach, B. M. & Johnson, A. D. Transcriptional repression in eukaryotes. *Annual Review of Cell Biology* **9**, 479–509 (1993).
16. Wu, J. & Grunstein, M. 25 Years after the nucleosome model: Chromatin modifications. *Trends in Biochemical Sciences* **25**, 619–623 (2000).
17. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Structural and Molecular Biology* **11**, 394–403 (2004).
18. Latchman, D. S. *Eukaryotic transcription factors*. (Academic press, 2010).
19. Hirai, H., Tani, T. & Kikyo, N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *International Journal of Developmental Biology* **54**, 1589–1596 (2011).
20. Vassilev, A., Kaneko, K. J., Shu, H., Zhao, Y. & DePamphilis, M. L. TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. *Genes Dev.* **15**, 1229–1241 (2001).
21. Jiang, B. H., Rue, E., Wang, G. L., Roe, R. & Semenza, G. L. Dimerization, DNA binding, and transactivation properties of hypoxia- inducible factor 1. *J. Biol. Chem.* **271**, 17771–17778 (1996).
22. Arnold, C. D. *et al.* A high- throughput method to identify trans- activation domains within transcription factor sequences. *EMBO J.* **37**, (2018).
23. Weirauch, M. T. & Hughes, T. R. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell. Biochem.* **52**, 25–73 (2014).
24. Najafabadi, H. S. *et al.* C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* **33**, 555–562 (2015).
25. Bobola, N. & Merabet, S. Homeodomain proteins in action: similar DNA binding

preferences, highly variable connectivity. *Current Opinion in Genetics and Development* **43**, 1–8 (2017).
26. Svingen, T. & Tonissen, K. F. Hox transcription factors and their elusive mammalian gene targets. *Heredity* **97**, 88–96 (2006).
27. Gehring, W. J. *et al.* The structure of the homeodomain and its functional implications. *Trends Genet.* **6**, 323–329 (1990).
28. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* **29**, 231–262 (2005).
29. Jones, S. An overview of the basic helix-loop-helix proteins. *Genome Biology* **5**, 226 (2004).
30. Fairman, R. *et al.* Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10429–10433 (1993).
31. Murre, C., McCaw, P. S. & Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**, 777–783 (1989).
32. Srivastava, D. & Mahony, S. Sequence and chromatin determinants of transcription factor binding and the establishment of cell type-specific binding patterns. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1863**, 194443 (2020).
33. Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16–23 (2000).
34. Stormo, G. D. & Zhao, Y. Determining the specificity of protein-DNA interactions. *Nature Reviews Genetics* **11**, 751–760 (2010).
35. Levo, M. *et al.* Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* **25**, 1018–1029 (2015).
36. Slattery, M. *et al.* Absence of a simple code: How transcription factors read the genome. *Trends in Biochemical Sciences* **39**, 381–399 (2014).
37. Kitayner, M. *et al.* Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **17**, 423–429 (2010).
38. Machado, A. C. D. *et al.* Proteopedia: 3D visualization and annotation of transcription factor-DNA readout modes. *Biochem. Mol. Biol. Educ.* **40**, 400–401 (2012).
39. Chen, Y. *et al.* Structure of p53 binding to the BAX response element reveals DNA unwinding and compression to accommodate base-pair insertion. *Nucleic Acids Res.* **41**, 8368–8376 (2013).
40. Johnson, A. 37 A Combinatorial Regulatory Circuit in Budding Yeast. *Cold Spring Harb. Monogr. Arch.* **22**, 975–1006 (1992).
41. Wolberger, C. Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 29–56 (1999).
42. Kim, S. *et al.* Probing allostery through DNA. *Science (80-.).* **339**, 816–819 (2013).
43. Dai, Z., Dai, X., Xiang, Q. & Feng, J. Nucleosomal Context of Binding Sites Influences Transcription Factor Binding Affinity and Gene Regulation. *Genomics, Proteomics Bioinforma.* **7**, 155–162 (2009).
44. Fogh, R. H. *et al.* Solution structure of the LexA repressor DNA binding domain determined by 1H NMR spectroscopy. *EMBO J.* **13**, 3936–3944 (1994).
45. Butala, M., Žgur-Bertok, D. & Busby, S. J. W. The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences* **66**, 82–93 (2009).
46. Zhang, A. P. P., Pigli, Y. Z. & Rice, P. A. Structure of the LexA–DNA complex and implications for SOS box measurement. *Nature* **466**, 883–886 (2010).
47. Lewis, L. K., Harlow, G. R., Gregg-Jolly, L. A. & Mount, D. W. Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in Escherichia coli. *J. Mol. Biol.* **241**, 507–523 (1994).
48. Minezaki, Y., Homma, K., Kinjo, A. R. & Nishikawa, K. Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation. *J. Mol. Biol.* **359**, 1137–1149 (2006).

49. Frietze, S. & Farnham, P. J. Transcription factor effector domains. *Subcell. Biochem.* **52**, 261–277 (2011).
50. Tanese, N., Pugh, B. F. & Tjian, R. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev.* **5**, 2212–2224 (1991).
51. Kim, T. K. & Roeder, R. G. Proline-rich activator CTF1 targets the TFIIB assembly step during transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4170–4174 (1994).
52. Blair, W. S., Bogerd, H. P., Madore, S. J. & Cullen, B. R. Mutational analysis of the transcription activation domain of RelA: identification of a highly synergistic minimal acidic activation module. *Mol. Cell. Biol.* **14**, 7226–7234 (1994).
53. Lin, J., Chen, J., Elenbaas, B. & Levine, A. J. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. *Genes Dev.* **8**, 1235–1246 (1994).
54. Blau, J. *et al.* Three functional classes of transcriptional activation domain. *Mol. Cell. Biol.* **16**, 2044–2055 (1996).
55. Rahl, P. B. *et al.* C-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
56. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192 (2009).
57. Yankulov, K., Blau, J., Purton, T., Roberts, S. & Bentley, D. L. Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell* **77**, 749–759 (1994).
58. Krumm, A., Hickey, L. B. & Groudine, M. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes Dev.* **9**, 559–572 (1995).
59. Walker, S., Greaves, R. & O'Hare, P. Transcriptional activation by the acidic domain of Vmw65 requires the integrity of the domain and involves additional determinants distinct from those necessary for TFIIB binding. *Mol. Cell. Biol.* **13**, 5233–5244 (1993).
60. Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J. & Verdine, G. L. Induced α helix in the VP16 activation domain upon binding to a human TAF. *Science (80-.).* **277**, 1310–1313 (1997).
61. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D. & Knutson, J. R. Transcriptional activation domain of the herpesvirus protein VP16 becomes conformationally constrained upon interaction with basal transcription factors. *J. Biol. Chem.* **271**, 4827–4837 (1996).
62. Hall, D. B. & Struhl, K. The VP16 Activation Domain Interacts with Multiple Transcriptional Components as Determined by Protein-Protein Cross-linking in Vivo *. **277**, 46043–46050 (2002).
63. Ikeda, K., Stuehler, T. & Meisterernst, M. The H1 and H2 regions of the activation domain of herpes simplex virion protein 16 stimulate transcription through distinct molecular mechanisms. *Genes to Cells* **7**, 49–58 (2002).
64. Hampsey, M. Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503 (1998).
65. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews: Developmental Biology* **1**, 40–51 (2012).
66. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics* **15**, 272–286 (2014).
67. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
68. Spitz, F. & Furlong, E. E. M. Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics* **13**, 613–626 (2012).
69. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology* **19**, 621–637 (2018).
70. Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 6453–6457 (1982).
71. Weller, P., Jeffreys, A. J., Wilson, V. & Blanchetot, A. Organization of the human

myoglobin gene. *EMBO J.* **3**, 439–446 (1984).
72. Wefald, F. C., Devlin, B. H. & Williams, R. S. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* **344**, 260–262 (1990).
73. Xi, H. *et al.* Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res.* **17**, 798–806 (2007).
74. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. & Ebright, R. H. New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**, 34–44 (1998).
75. Ohler, U., Liao, G. chun, Niemann, H. & Rubin, G. M. Computational analysis of core promoters in the Drosophila genome. *Genome Biol.* **3**, (2002).
76. Burke, T. W. & Kadonaga, J. T. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**, 711–724 (1996).
77. Lewis, B. A., Kim, T. K. & Orkin, S. H. A downstream element in the human β-globin promoter: Evidence of extended sequence-specific transcription factor IID contacts. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7172–7177 (2000).
78. Roy, A. L. & Singer, D. S. Core Promoters in Transcription: Old Problem, New Insights. *Trends Biochem. Sci.* **40**, 165 (2015).
79. Shi, W. & Zhou, W. Frequency distribution of TATA Box and extension sequences on human promoters. *BMC Bioinformatics* **7**, S2 (2006).
80. Albert, T. K., Grote, K., Boeing, S. & Meisterernst, M. Basal core promoters control the equilibrium between negative cofactor 2 and preinitiation complexes in human cells. *Genome Biol.* **11**, (2010).
81. Ponjavic, J. *et al.* Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.* **7**, R78 (2006).
82. FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of Drosophila and human core promoters. *Genome Biol.* **7**, R53 (2006).
83. Patikoglou, G. A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217–3230 (1999).
84. Burley, S. K. & Roeder, R. G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review of Biochemistry* **65**, 769–799 (1996).
85. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
86. Smale, S. T. & Baltimore, D. The 'initiator' as a transcription control element. *Cell* **57**, 103–113 (1989).
87. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
88. Ngoc, L. V., Cassidy, C. J., Huang, C. Y., Duttke, S. H. C. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev.* **31**, 6–11 (2017).
89. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAF(II)60 of Drosophila. *Genes Dev.* **11**, 3020–3031 (1997).
90. Deng, W. & Roberts, S. G. E. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**, 2418–2423 (2005).
91. Lee, D.-H. *et al.* Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. *Mol. Cell. Biol.* **25**, 9674–9686 (2005).
92. Wu, S. Y. & Chiang, C. M. TATA-binding Protein-associated Factors Enhance the Recruitment of RNA Polymerase II by Transcriptional Activators. *J. Biol. Chem.* **276**, 34235–34243 (2001).
93. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched

in human TATA-less core promoters. *Gene* **389**, 52–65 (2007).
94. Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095–1107 (1992).
95. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annual Review of Biochemistry* **72**, 449–479 (2003).
96. Yamashita, R., Suzuki, Y., Sugano, S. & Nakai, K. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* **350**, 129–136 (2005).
97. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
98. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
99. Murakami, K. *et al.* Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J. Biol. Chem.* **288**, 6325–6332 (2013).
100. Bushnell, D. A., Bamdad, C. & Kornberg, R. D. A minimal set of RNA polymerase II transcription protein interactions. *J. Biol. Chem.* **271**, 20170–20174 (1996).
101. Ohkuma, Y., Hashimoto, S., Wang, C. K., Horikoshi, M. & Roeder, R. G. Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. *Mol. Cell. Biol.* **15**, 4856–4866 (1995).
102. Maxon, M. E., Goodrich, J. A. & Tjian, R. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: A model for promoter clearance. *Genes Dev.* **8**, 515–524 (1994).
103. Holstege, F. C. P., Tantin, D., Carey, M., Van Der Vliet, P. C. & Timmers, H. T. M. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* **14**, 810–819 (1995).
104. Timmers, H. T. M. Transcription initiation by RNA polymerase II does not require hydrolysis of the β-γ phosphoanhydride bond of ATP. *EMBO J.* **13**, 391–399 (1994).
105. Sopta, M., Carthew, R. W. & Greenblatt, J. Isolation of three proteins that bind to mammalian RNA polymerase II. *J. Biol. Chem.* **260**, 10353–10360 (1985).
106. Čabart, P., Újvári, A., Pal, M. & Luse, D. S. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15786–15791 (2011).
107. Fishburn, J. & Hahn, S. Architecture of the Yeast RNA Polymerase II Open Complex and Regulation of Activity by TFIIF. *Mol. Cell. Biol.* **32**, 12–25 (2012).
108. Cianfrocco, M. A. *et al.* Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152**, 120–131 (2013).
109. Blair, R. H., Goodrich, J. A. & Kugel, J. F. Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics. *Biochemistry* **51**, 7444–7455 (2012).
110. Hieb, A. R. *et al.* TFIIA Changes the Conformation of the DNA in TBP/TATA Complexes and Increases their Kinetic Stability. *J. Mol. Biol.* **372**, 619–632 (2007).
111. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
112. Su, W., Jackson, S., Tjian, R. & Echols, H. DNA looping between sites for transcriptional activation: Self-association of DNA-bound Sp1. *Genes Dev.* **5**, 820–826 (1991).
113. Hensel, Z., Weng, X., Lagda, A. C. & Xiao, J. Transcription-Factor-Mediated DNA Looping Probed by High-Resolution, Single-Molecule Imaging in Live E. coli Cells. *PLoS Biol.* **11**, 1001591 (2013).
114. Ringrose, L., Chabanis, S., Angrand, P. O., Woodroofe, C. & Stewart, A. F. Quantitative comparison of DNA looping in vitro and in vivo: Chromatin increases effective DNA flexibility at short distances. *EMBO J.* **18**, 6630–6641 (1999).
115. Swigon, D., Coleman, B. D. & Olson, W. K. Modeling the Lac repressor-operator assembly: The influence of DNA looping on Lac repressor conformation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9879–9884 (2006).

116. Towles, K. B., Beausang, J. F., Garcia, H. G., Phillips, R. & Nelson, P. C. First-principles calculation of DNA looping in tethered particle experiments. *Phys. Biol.* **6**, (2009).
117. Zhang, Y., McEwen, A. E., Crothers, D. M. & Levene, S. D. Analysis of In-Vivo LacR-Mediated Gene Repression Based on the Mechanics of DNA Looping. *PLoS One* **1**, e136 (2006).
118. Shimada, J. & Yamakawa, H. Ring-Closure Probabilities for Twisted Wormlike Chains. Application to DNA. *Macromolecules* **17**, 689–698 (1984).
119. Johnson, S., Lindén, M. & Phillips, R. Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res.* **40**, 7728–7738 (2012).
120. Nolis, I. K. *et al.* Transcription factors mediate long-range enhancer-promoter interactions. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20222–20227 (2009).
121. Cloutier, T. E. & Widom, J. Spontaneous sharp bending of double-stranded DNA. *Mol. Cell* **14**, 355–362 (2004).
122. Cloutier, T. E. & Widom, J. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3645–3650 (2005).
123. Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
124. Kornberg, R. D. Chromatin structure: A repeating unit of histones and DNA. *Science (80-.).* **184**, 868–871 (1974).
125. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).
126. Cosma, M. P. Ordered recruitment: Gene-specific mechanism of transcription activation. *Molecular Cell* **10**, 227–236 (2002).
127. Robinson, P. J. J., Bushnell, D. A., Trnka, M. J., Burlingame, A. L. & Kornberg, R. D. Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17931–17935 (2012).
128. Robinson, P. J. *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1422.e16 (2016).
129. Lu, H., Flores, O., Weinmann, R. & Reinberg, D. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10004–10008 (1991).
130. Myers, L. C. *et al.* The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12**, 45–54 (1998).
131. Hahn, S. & Young, E. T. Transcriptional regulation in saccharomyces cerevisiae: Transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705–736 (2011).
132. Tirode, F., Busso, D., Coin, F. & Egly, J. M. Reconstitution of the transcription factor TFIIH: Assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* **3**, 87–95 (1999).
133. Luse, D. S. The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled? *Transcription* **5**, (2013).
134. Wang, W., Carey, M. & Gralla, J. D. Polymerase II promoter activation: Closed complex formation and ATP-driven start site opening. *Science (80-.).* **255**, 450–453 (1992).
135. Payne, J. M., Laybourn, P. J. & Dahmus, M. E. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J. Biol. Chem.* **264**, 19621–19629 (1989).
136. Søgaard, T. M. M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282**, 14113–14120 (2007).
137. Jeronimo, C. & Robert, F. Kin28 regulates the transient association of Mediator with core promoters. *Nat. Struct. Mol. Biol.* **21**, 449–455 (2014).
138. Wong, K. H., Jin, Y. & Struhl, K. TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol. Cell* **54**, 601–612 (2014).

139. Harlen, K. M. & Churchman, L. S. The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology* **18**, 263–273 (2017).
140. Cadena, D. L. & Dahmus, M. E. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.* **262**, 12468–12474 (1987).
141. Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
142. Yamaguchi, Y. *et al.* NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).
143. Marshall, N. F. & Price, D. H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol. Cell. Biol.* **12**, 2078–2090 (1992).
144. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**, 12335–12338 (1995).
145. Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D. & Handa, H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* **17**, 7395–7403 (1998).
146. Czudnochowski, N., Bösken, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3**, (2012).
147. Peterlin, B. M. & Price, D. H. Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell* **23**, 297–305 (2006).
148. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, 1045–1051 (2017).
149. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes and Development* **33**, 960–982 (2019).
150. Beach, D. L., Salmon, E. D. & Bloom, K. Localization and anchoring of mRNA in budding yeast. *Curr. Biol.* **9**, 569–578 (1999).
151. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
152. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2014).
153. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science (80-.).* **332**, 475–478 (2011).
154. Wu, B., Chen, J. & Singer, R. H. Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci. Rep.* **4**, 3615 (2014).
155. Garcia, J. F. & Parker, R. MS2 coat proteins bound to yeast mRNAs block 5′ to 3′ degradation and trap mRNA decay products: Implications for the localization of mRNAs by MS2-MCP system. *RNA* **21**, 1393–1395 (2015).
156. Mannack, L. V. J. C., Eising, S. & Rentmeister, A. Current techniques for visualizing RNA in cells. *F1000Research* **5**, (2016).
157. Tutucci, E. *et al.* An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* **15**, 81–89 (2018).
158. Wilson, M. Z., Ravindran, P. T., Lim, W. A. & Toettcher, J. E. Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control. *Mol. Cell* **67**, 757-769.e5 (2017).
159. Shuman, S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Progress in Nucleic Acid Research and Molecular Biology* **66**, 1–40 (2000).
160. Hocine, S., Singer, R. H. & Grünwald, D. RNA processing and export. *Cold Spring Harbor perspectives in biology* **2**, (2010).
161. Yue, Z. *et al.* Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form

of RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12898–12903 (1997).
162. Walther, T. N., Wittop Koning, T. H., Schümperli, D. & Müller, B. A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. *RNA* **4**, 1034–1046 (1998).
163. Hsu, C. L. & Stevens, A. Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.* **13**, 4826–4835 (1993).
164. Proudfoot, N. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Current Opinion in Cell Biology* **16**, 272–278 (2004).
165. Drummond, D. R., Armstrong, J. & Colman, A. The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in Xenopus oocytes. *Nucleic Acids Res.* **13**, 7375–7394 (1985).
166. Zilka, A., Garlapati, S., Dahan, E., Yaolsky, V. & Shapira, M. Developmental regulation of heat shock protein 83 in Leishmania: 3′ processing and mRNA stability control transcript abundance, and translation is directed by a determinant in the 3′-untranslated region. *J. Biol. Chem.* **276**, 47922–47929 (2001).
167. Neugebauer, K. M. On the importance of being co-transcriptional. *J. Cell Sci.* **115**, 3865–3871 (2002).
168. Moore, M. J. & Proudfoot, N. J. Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. *Cell* **136**, 688–700 (2009).
169. Kim, J. & Iyer, V. R. Global Role of TATA Box-Binding Protein Recruitment to Promoters in Mediating Gene Expression Profiles. *Mol. Cell. Biol.* **24**, 8104–8112 (2004).
170. Hinnebusch, A. G. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
171. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* **11**, 113–127 (2010).
172. Pestova, T. V. *et al.* The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403**, 332–335 (2000).
173. Unbehaun, A., Borukhov, S. I., Hellen, C. U. T. & Pestova, T. V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev.* **18**, 3078–3093 (2004).
174. Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J. & McCarthy, J. E. G. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol. Microbiol.* **51**, 987–1001 (2004).
175. Vassilenko, K. S., Alekhina, O. M., Dmitriev, S. E., Shatsky, I. N. & Spirin, A. S. Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Research* **39**, 5555–5567 (2011).
176. Matsuda, D. & Dreher, T. W. Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA* **12**, 1338–1349 (2006).
177. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* **19**, 158–174 (2018).
178. Haimov, O., Sinvani, H. & Dikstein, R. Cap-dependent, scanning-free translation initiation mechanisms. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1849**, 1313–1318 (2015).
179. Hernández, G. & Jagus, R. *Evolution of the protein synthesis machinery and its regulation*. *Evolution of the Protein Synthesis Machinery and Its Regulation* (Springer International Publishing, 2016). doi:10.1007/978-3-319-39468-8
180. Hernández, G., Altmann, M. & Lasko, P. Origins and evolution of the mechanisms regulating translation initiation in eukaryotes. *Trends in Biochemical Sciences* **35**, 63–73 (2010).
181. Montoya, J., Ojala, D. & Attardi, G. Distinctive features of the 5′-terminal sequences of

the human mitochondrial mRNAs. *Nature* **290**, 465–470 (1981).
182. Lynch, M. The Evolution of Transcription-Initiation Sites. *Mol. Biol. Evol.* **22**, 1137–1146 (2005).
183. Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A Red Carpet for Iron Metabolism. *Cell* **168**, 344–361 (2017).
184. Hentze, M. W. *et al.* Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science (80-.).* **238**, 1570–1573 (1987).
185. Kumari, S., Bugaut, A. & Balasubramanian, S. Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5′ UTR of the NRAS proto-oncogene. *Biochemistry* **47**, 12664–12669 (2008).
186. Halder, K., Wieland, M. & Hartig, J. S. Predictable suppression of gene expression by 5′-UTR-based RNA quadruplexes. *Nucleic Acids Res.* **37**, 6811–6817 (2009).
187. Resch, A. M., Ogurtsov, A. Y., Rogozin, I. B., Shabalina, S. A. & Koonin, E. V. Evolution of alternative and constitutive regions of mammalian 5′ UTRs. *BMC Genomics* **10**, 162 (2009).
188. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7507–7512 (2009).
189. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORF s are prevalent translational repressors in vertebrates . *EMBO J.* **35**, 706–723 (2016).
190. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
191. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (80-.).* **324**, 218–223 (2009).
192. Luukkonen, B. G., Tan, W. & Schwartz, S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J. Virol.* **69**, 4086–4094 (1995).
193. Hinnebusch, A. G. Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology* **59**, 407–450 (2005).
194. Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416 (2016).
195. Jia, L. *et al.* Decoding mRNA translatability and stability from the 5′ UTR. *Nat. Struct. Mol. Biol.* **27**, 814–821 (2020).
196. Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational Control through Differential Ribosome Pausing during Amino Acid Limitation in Mammalian Cells. *Mol. Cell* **71**, 229-243.e11 (2018).
197. Joazeiro, C. A. P. Ribosomal stalling during translation: Providing substrates for ribosome-associated protein quality control. *Annual Review of Cell and Developmental Biology* **33**, 343–368 (2017).
198. Purvis, J. E. & Lahav, G. Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956 (2013).
199. Castellanos, M., Mothi, N. & Muñoz, V. Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nat. Commun.* **11**, 1–13 (2020).
200. Li, C., Cesbron, F., Oehler, M., Brunner, M. & Höfer, T. Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. *Cell Syst.* **6**, 409-423.e11 (2018).
201. Hansen, A. S. & O'Shea, E. K. Promoter decoding of transcription factor dynamics involves a trade- off between noise and control of gene expression. *Mol. Syst. Biol.* **9**, 704 (2013).
202. Hao, N. & O'Shea, E. K. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.* **19**, 31–40 (2012).
203. Marshall, C. J. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).

204. Traverse, S., Gomez, N., Paterson, H., Marshall, C. & Cohen, P. Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor. *Biochem. J.* **288**, 351–355 (1992).
205. Nguyens, T. T. *et al. Co-regulation of the mitogen-activated protein kinase, extracellular signal-regulated kinase 1, and the 90-kDa ribosomal S6 kinase in PC12 cells. Distinct effects of the neurotrophic factor, nerve growth factor, and the mitogenic factor, epidermal growth factor*. **268**, (1993).
206. GOTOH, Y. *et al.* Microtubule- associated- protein (MAP) kinase activated by nerve growth factor and epidermal growth factor in PC12 cells: Identity with the mitogen-activated MAP kinase of fibroblastic cells. *Eur. J. Biochem.* **193**, 661–669 (1990).
207. Tay, S. *et al.* Single-cell NF-B dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
208. Sung, M. H. *et al.* Sustained oscillations of NF-κB produce distinct genome scanning and gene expression profiles. *PLoS One* **4**, (2009).
209. Nelson, D. E. *et al.* Oscillations in NF-κB signaling control the dynamics of gene expression. *Science (80-.).* **306**, 704–708 (2004).
210. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IκB-NF-κB signaling module: Temporal control and selective gene activation. *Science (80-.).* **298**, 1241–1245 (2002).
211. Werner, S. L., Barken, D. & Hoffmann, A. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science (80-.).* **309**, 1857–1861 (2005).
212. Lee, T. K. *et al.* A noisy paracrine signal determines the cellular NF-κB response to lipopolysaccharide. *Sci. Signal.* **2**, (2009).
213. Covert, M. W., Leung, T. H., Gaston, J. E. & Baltimore, D. Achieving stability of lipopolysaccharide-induced NF-κB activation. *Science (80-.).* **309**, 1854–1857 (2005).
214. Werner, S. L. *et al.* Encoding NF-κB temporal control in response to TNF: Distinct roles for the negative regulators IκBα and A20. *Genes Dev.* **22**, 2093–2101 (2008).
215. Basak, S., Behar, M. & Hoffmann, A. Lessons from mathematically modeling the NF-κB pathway. *Immunol. Rev.* **246**, 221–238 (2012).
216. Ashall, L. *et al.* Pulsatile stimulation determines timing and specificity of NF-κB-dependent transcription. *Science (80-.).* **324**, 242–246 (2009).
217. Cheong, R. *et al.* Transient IκB kinase activity mediates temporal NF-κB dynamics in response to a wide range of tumor necrosis factor-α doses. *J. Biol. Chem.* **281**, 2945–2950 (2006).
218. Purvis, J. E. *et al.* p53 dynamics control cell fate. *Science (80-.).* **336**, 1440–1444 (2012).
219. Geva-Zatorsky, N. *et al.* Protein Dynamics in Drug Combinations: a Linear Superposition of Individual-Drug Responses. *Cell* **140**, 643–651 (2010).
220. Lahav, G. *et al.* Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* **36**, 147–150 (2004).
221. Batchelor, E., Loewer, A., Mock, C. & Lahav, G. Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* **7**, (2011).
222. Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca2+ response amplitude and duration. *Nature* **386**, 855–858 (1997).
223. Hansen, A. S. & O'Shea, E. K. Cis Determinants of Promoter Threshold and Activation Timescale. *Cell Rep.* **12**, 1226–1233 (2015).
224. Oda, K. *et al.* p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by ser-46-phosphorylated p53. *Cell* **102**, 849–862 (2000).
225. Knights, C. D. *et al.* Distinct p53 acetylation cassettes differentially influence gene-expression patterns and cell fate. *J. Cell Biol.* **173**, 533–544 (2006).
226. Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of

p53. *Mol. Cell* **8**, 781–794 (2001).
227. Bergamaschi, D. *et al.* iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. *Nat. Genet.* **33**, 162–167 (2003).
228. Bergamaschi, D. *et al.* iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53. *Nat. Genet.* **38**, 1133–1141 (2006).
229. Budhram-Mahadeo, V. S. *et al.* Brn-3b enhances the pro-apoptotic effects of p53 but not its induction of cell cycle arrest by cooperating in trans-activation of bax expression. *Nucleic Acids Res.* **34**, 6640–6652 (2006).
230. Budram-Mahadeo, V., Morris, P. J. & Latchman, D. S. The Brn-3a transcription factor inhibits the pro-apoptotic effect of p53 and enhances cell cycle arrest by differentially regulating the activity of the p53 target genes encoding Bax and p21CIP1/Waf1. *Oncogene* **21**, 6123–6131 (2002).
231. Das, S. *et al.* Hzf Determines Cell Survival upon Genotoxic Stress by Modulating p53 Transactivation. *Cell* **130**, 624–637 (2007).
232. Millau, J. F. *et al.* Formation of stress-specific p53 binding patterns is influenced by chromatin but not by modulation of p53 binding affinity to response elements. *Nucleic Acids Res.* **39**, 3053–3063 (2011).
233. Gomes, N. P. & Espinosa, J. M. Differential regulation of p53 target genes: It's (core promoter) elementary. *Genes and Development* **24**, 111–114 (2010).
234. Morachis, J. M., Murawsky, C. M. & Emerson, B. M. Regulation of the p53 transcriptional response by structurally diverse core promoters. *Genes Dev.* **24**, 135–147 (2010).
235. Tay, S. *et al.* Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–71 (2010).
236. Sen, S., Cheng, Z., Sheu, K. M., Chen, Y. H. & Hoffmann Correspondence, A. Gene Regulatory Strategies that Decode the Duration of NFκB Dynamics Contribute to LPS- versus TNF-Specific Gene Expression. *Cell Syst.* **10**, 169-182.e5 (2020).
237. Dvir, S. *et al.* Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2021 (2013).
238. Acevedo, J. M., Hoermann, B., Schlimbach, T. & Teleman, A. A. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci. Rep.* **8**, 4018 (2018).
239. Li, J., Liang, Q., Song, W. & Marchisio, M. A. Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast S. cerevisiae. *J. Biol. Eng.* **11**, 25 (2017).
240. Babendure, J. R., Babendure, J. L., Ding, J. H. & Tsien, R. Y. Control of mammalian translation by mRNA structure near caps. *RNA* **12**, 851–861 (2006).
241. Porter, J. R., Fisher, B. E. & Batchelor, E. P53 Pulses Diversify Target Gene Expression Dynamics in an mRNA Half-Life-Dependent Manner and Delineate Co-regulated Target Gene Subnetworks. *Cell Syst.* **2**, 272–282 (2016).
242. Hao, S. & Baltimore, D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.* **10**, 281–288 (2009).
243. Shalem, O. *et al.* Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* **4**, (2008).
244. Bishop, A. C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000).
245. Toettcher, J. E., Weiner, O. D. & Lim, W. A. Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell* **155**, 1422–1434 (2013).
246. Yumerefendi, H. *et al.* Control of protein activity and cell fate specification via light-mediated nuclear translocation. *PLoS One* **10**, (2015).
247. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* **5**, 4404 (2014).
248. Chen, S. Y. *et al.* Optogenetic Control Reveals Differential Promoter Interpretation of

Transcription Factor Nuclear Translocation Dynamics. *Cell Syst.* **11**, 336-353.e24 (2020).
249. Plouffe, S. W. *et al.* The Hippo pathway effector proteins YAP and TAZ have both distinct and overlapping functions in the cell. *J. Biol. Chem.* **293**, 11230–11240 (2018).
250. Tian, W., Yu, J., Tomchick, D. R., Pan, D. & Luo, X. Structural and functional analysis of the YAP-binding domain of human TEAD2. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7293–7298 (2010).
251. Kanai, F. *et al.* TAZ: A novel transcriptional co-activator regulated by interactions with 14-3-3 and PDZ domain proteins. *EMBO J.* **19**, 6778–6791 (2000).
252. Zhao, C. *et al.* Yes-associated protein (YAP) and transcriptional coactivator with a PDZ-binding motif (TAZ): a nexus between hypoxia and cancer. *Acta Pharmaceutica Sinica B* **10**, 947–960 (2020).
253. Zhao, B. *et al.* Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761 (2007).
254. Zhao, B., Lei, Q. Y. & Guan, K. L. The Hippo-YAP pathway: new connections between regulation of organ size and cancer. *Current Opinion in Cell Biology* **20**, 638–646 (2008).
255. Varelas, X. The hippo pathway effectors TAZ and YAP in development, homeostasis and disease. *Development (Cambridge)* **141**, 1614–1626 (2014).
256. Holden, J. K. & Cunningham, C. N. Targeting the hippo pathway and cancer through the TEAD family of transcription factors. *Cancers* **10**, (2018).
257. Dong, X. *et al.* YAP/TAZ: A promising target for squamous cell carcinoma treatment. *Cancer Management and Research* **11**, 6245–6252 (2019).
258. Strano, S. *et al.* Physical Interaction with Yes-associated Protein Enhances p73 Transcriptional Activity. *J. Biol. Chem.* **276**, 15164–15173 (2001).
259. Du, W. *et al.* Insights into the Regulation of Yap/Taz from Cellular Systems and Mouse Models. *Curr. Stem Cell Res. Ther.* **13**, (2018).
260. Kim, M. K., Jang, J. W. & Bae, S. C. DNA binding partners of YAP/TAZ. *BMB Reports* **51**, 126–133 (2018).
261. Wu, S., Liu, Y., Zheng, Y., Dong, J. & Pan, D. The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Dev. Cell* **14**, 388–398 (2008).
262. Sun, Y. *et al.* Hippo/YAP-mediated rigidity-dependent motor neuron differentiation of human pluripotent stem cells. *Nat. Mater.* **13**, 599–604 (2014).
263. Dupont, S. *et al.* Role of YAP/TAZ in mechanotransduction. *Nature* **474**, 179–184 (2011).
264. Aragona, M. *et al.* A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell* **154**, 1047–1059 (2013).
265. Varelas, X. *et al.* Developmental Cell The Crumbs Complex Couples Cell Density Sensing to Hippo-Dependent Control of the TGF-b-SMAD Pathway. (2010). doi:10.1016/j.devcel.2010.11.012
266. Seo, J. & Kim, J. Regulation of Hippo signaling by actin remodeling. *BMB Reports* **51**, 151–156 (2018).
267. Wada, K. I., Itoga, K., Okano, T., Yonemura, S. & Sasaki, H. Hippo pathway regulation by cell morphology and stress fibers. *Development* **138**, 3907–3914 (2011).
268. Das, A., Fischer, R. S., Pan, D. & Waterman, C. M. YAP nuclear localization in the absence of cell-cell contact is mediated by a filamentous actin-dependent, Myosin IIand Phospho-YAP-independent pathway during extracellular matrix mechanosensing. *J. Biol. Chem.* **291**, 6096–6110 (2016).
269. Shi, J., Farzaneh, M. & Khoshnam, S. E. Yes-associated protein and PDZ binding motif: A critical signaling pathway in the control of human pluripotent stem cells self-renewal and differentiation. *Cell. Reprogram.* **22**, 55–61 (2020).
270. Jacquemin, P., Hwang, J. J., Martial, J. A., Dollé, P. & Davidson, I. A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS

DNA binding domain. *J. Biol. Chem.* **271**, 21775–21785 (1996).
271. Anbanandam, A. *et al.* Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17225–17230 (2006).
272. Azakie, A., Larkin, S. B., Farrance, I. K., Grenningloh, G. & Ordahl, C. P. DTEF-1, a novel member of the transcription enhancer factor-1 (TEF-1) multigene family. *J. Biol. Chem.* **271**, 8260–8265 (1996).
273. Farrance, I. K. G., Mar, J. H. & Ordahl, C. P. M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1. *J. Biol. Chem.* **267**, 17234–17240 (1992).
274. Xiao, J. H. *et al.* One cell-specific and three ubiquitous nuclear proteins bind in vitro to overlapping motifs in the domain B1 of the SV40 enhancer. *EMBO J.* **6**, 3005–3013 (1987).
275. Zanconato, F. *et al.* Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat. Cell Biol.* **17**, 1218–1227 (2015).
276. Benhaddou, A. *et al.* Transcription factor TEAD4 regulates expression of Myogenin and the unfolded protein response genes during C2C12 cell differentiation. *Cell Death Differ.* **19**, 220–231 (2012).
277. Kaneko, K. J. & Depamphilis, M. L. Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. *Dev. Genet.* **22**, 43–55 (1998).
278. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Current Opinion in Structural Biology* **47**, 1–8 (2017).
279. Zhao, B. *et al.* TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22**, 1962–1971 (2008).
280. Shi, Z. *et al.* DNA-binding mechanism of the Hippo pathway transcription factor TEAD4. *Oncogene* **36**, 4362–4369 (2017).
281. Li, Z. *et al.* Structural insights into the YAP and TEAD complex. *Genes Dev.* **24**, 235–240 (2010).
282. Chen, L. *et al.* Structural basis of YAP recognition by TEAD4 in the Hippo pathway. *Genes Dev.* **24**, 290–300 (2010).
283. Chen, H. H., Mullett, S. J. & Stewart, A. F. R. Vgl-4, a novel member of the vestigial-like family of transcription cofactors, regulates α1-adrenergic activation of gene expression in cardiac myocytes. *J. Biol. Chem.* **279**, 30800–30806 (2004).
284. Chen, H. H., Maeda, T., Mullett, S. J. & Stewart, A. F. R. Transcription cofactor Vgl-2 is required for skeletal muscle differentiation. *Genesis* **39**, 273–279 (2004).
285. Pobbati, A. V., Chan, S. W., Lee, I., Song, H. & Hong, W. Structural and functional similarity between the Vgll1-TEAD and the YAP-TEAD complexes. *Structure* **20**, 1135–1140 (2012).
286. Jiao, S. *et al.* A Peptide Mimicking VGLL4 Function Acts as a YAP Antagonist Therapy against Gastric Cancer. *Cancer Cell* **25**, 166–180 (2014).
287. Zhang, W. *et al.* VGLL4 functions as a new tumor suppressor in lung cancer by negatively regulating the YAP-TEAD transcriptional complex. *Cell Res.* **24**, 331–343 (2014).
288. Butler, A. J. & Ordahl, C. P. Poly(ADP-Ribose) Polymerase Binds with Transcription Enhancer Factor 1 to MCAT1 Elements To Regulate Muscle-Specific Transcription. *Mol. Cell. Biol.* **19**, 296–306 (1999).
289. Liu, X. *et al.* Tead and AP1 Coordinate Transcription and Motility. *Cell Rep.* **14**, 1169–1180 (2016).
290. Gupta, M. P., Amin, C. S., Gupta, M., Hay, N. & Zak, R. Transcription enhancer factor 1 interacts with a basic helix-loop-helix zipper protein, Max, for positive regulation of cardiac alpha-myosin heavy-chain gene expression. *Mol. Cell. Biol.* **17**, 3924–3936 (1997).
291. Maeda, T., Chapman, D. L. & Stewart, A. F. R. Mammalian vestigial-like 2, a cofactor of TEF-1 and MEF2 transcription factors that promotes skeletal muscle differentiation. *J. Biol. Chem.* **277**, 48889–48898 (2002).

292. MacLellan, W. R., Lee, T. C., Schwartz, R. J. & Schneider, M. D. Transforming growth factor-β response elements of the skeletal α-actin gene. Combinatorial action of serum response factor, YY1, and the SV40 enhancer-binding protein, TEF-1. *J. Biol. Chem.* **269**, 16754–16760 (1994).
293. Schank, J. C. The development of locomotor kinematics in neonatal rats: An agent-based modeling analysis in group and individual contexts. *J. Theor. Biol.* **254**, 826–842 (2008).
294. Hübner, K., Sahle, S. & Kummer, U. Applications and trends in systems biology in biochemistry. *FEBS J.* **278**, 2767–2857 (2011).
295. ElKalaawy, N. & Wassal, A. Methodologies for the modeling and simulation of biochemical networks, illustrated for signal transduction pathways: A primer. *BioSystems* **129**, 1–18 (2015).
296. Vera, J. *et al.* A systems biology approach to analyse amplification in the JAK2-STAT5 signalling pathway. *BMC Syst. Biol.* **2**, (2008).
297. Vera, J., Curto, R., Cascante, M. & Torres, N. V. Detection of potential enzyme targets by metabolic modelling and optimization: Application to a simple enzymopathy. *Bioinformatics* **23**, 2281–2289 (2007).
298. Voit, E. O. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. (Cambridge University Press, 2000).
299. Park, L. J., Park, C. H., Park, C. & Lee, T. Application of genetic algorithms to parameter estimation of bioprocesses. *Med. Biol. Eng. Comput.* **35**, 47–49 (1997).
300. Banga, J. R. Optimization in computational systems biology. *BMC Systems Biology* **2**, (2008).
301. Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A. & Blom, J. G. Systems biology: Parameter estimation for biochemical models. *FEBS Journal* **276**, 886–902 (2009).
302. van Riel, N. A. W. Dynamic modelling and analysis of biochemical networks: Mechanism-based models and model-based experiments. *Briefings in Bioinformatics* **7**, 364–374 (2006).
303. Polisetty, P. K., Voit, E. O. & Gatzke, E. P. Identification of metabolic system parameters using global optimization methods. *Theor. Biol. Med. Model.* **3**, (2006).
304. Savageau, M. A. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* **229**, 542–544 (1971).
305. Wu, W. H., Wang, F. S. & Chang, M. S. Dynamic sensitivity analysis of biological systems. in *BMC Bioinformatics* **9**, (BMC Bioinformatics, 2008).
306. Ni, T. C. & Savageau, M. A. Model assessment and refinement using strategies from biochemical systems theory: Application to metabolism in human red blood cells. *J. Theor. Biol.* **179**, 329–368 (1996).
307. Ni, T. C. & Savageau, M. A. Application of biochemical systems theory to metabolism in human red blood cells: Signal propagation and accuracy of representation. *J. Biol. Chem.* **271**, 7927–7941 (1996).
308. Okamoto, M. & Savageau, M. A. Integrated Function of a Kinetic Proofreading Mechanism: Double-Stage Proofreading by Isoleucyl-tRNA Synthetase. *Biochemistry* **25**, 1969–1975 (1986).
309. Okamoto, M. & Savageau, M. A. Integrated Function of a Kinetic Proofreading Mechanism: Steady-State Analysis Testing Internal Consistency of Data Obtained in Vivo and in Vitro and Predicting Parameter Values. *Biochemistry* **23**, 1701–1709 (1984).
310. Maiwald, T. *et al.* Combining theoretical analysis and experimental data generation reveals IRF9 as a crucial factor for accelerating interferon α-induced early antiviral signalling. *FEBS J.* **277**, 4741–4754 (2010).
311. Torres, N. V. & Santos, G. The (Mathematical) Modeling Process in Biosciences. *Front. Genet.* **6**, 354 (2015).
312. Kepler, T. B. & Elston, T. C. Stochasticity in transcriptional regulation: Origins,

consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
313. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science (80-.).* **304**, 1811–1814 (2004).
314. Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.* **9**, 1–10 (2018).
315. Rullan, M., Benzinger, D., Schmidt, G. W., Milias-Argeitis, A. & Khammash, M. An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation. *Mol. Cell* **70**, 745-756.e6 (2018).
316. Harper, C. V. *et al.* Dynamic Analysis of Stochastic Transcription Cycles. *PLoS Biol.* **9**, e1000607 (2011).
317. Suter, D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science (80-.).* **332**, 472–474 (2011).
318. Cesbron, F., Oehler, M., Ha, N., Sancar, G. & Brunner, M. Transcriptional refractoriness is dependent on core promoter architecture. *Nat. Commun.* **6**, 1–11 (2015).
319. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* **11**, 823 (2015).
320. Howard-Flanders, P., Boyce, R. P. & Theriot, L. Three loci in Escherichia coli K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA. *Genetics* **53**, 1119–1136 (1966).
321. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **2**, 718–729 (1988).
322. Yao, X., Rosen, M. K. & Gardner, K. H. Estimation of the available free energy in a LOV2-Jα photoswitch. *Nat. Chem. Biol.* **4**, 491–497 (2008).
323. Wang, H. *et al.* LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat. Methods* **13**, 755–758 (2016).
324. Hancock, J. F., Cadwallader, K., Paterson, H. & Marshall, C. J. A CAAX or a CAAL motif and a second signal are sufficient for plasma membrane targeting of ras proteins. *EMBO J.* **10**, 4033–4039 (1991).
325. Hancock, J. F., Paterson, H. & Marshall, C. J. A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell* **63**, 133–139 (1990).
326. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5718–5722 (1989).
327. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128**, 707–719 (2007).
328. Hassan, A. H., Neely, K. E., Vignali, M., Reese, J. C. & Workman, J. L. Promoter targeting of chromatin-modifying complexes. *Frontiers in bioscience : a journal and virtual library* **6**, (2001).
329. Hassan, A. H., Neely, K. E. & Workman, J. L. Histone acetyltransferase complexes stabilize SWI/SNF binding to promoter nucleosomes. *Cell* **104**, 817–827 (2001).
330. Green, M. R. Eukaryotic transcription activation: Right on target. *Molecular Cell* **18**, 399–402 (2005).
331. Çiçek, Ö., Marrakchi, Y., Boasiako Antwi, E., Di Ventura, B. & Brox, T. Recovering the Imperfect: Cell Segmentation in the Presence of Dynamically Localized Proteins. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **12446 LNCS**, 85–93 (Springer Science and Business Media Deutschland GmbH, 2020).
332. Filonov, G. S. *et al.* Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* **29**, 757–761 (2011).
333. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nat. Methods* **10**, 751–754 (2013).
334. Luo, Y., North, J. A., Rose, S. D. & Poirier, M. G. Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.* **42**, 3017–3027 (2014).

335. Abu-Shaar, M., Ryoo, H. D. & Mann, R. S. Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. *Genes Dev.* **13**, 935–945 (1999).
336. Meyers, R. E. & Sharp, P. A. TATA-binding protein and associated factors in polymerase II and polymerase III transcription. *Mol. Cell. Biol.* **13**, 7953–7960 (1993).
337. Akhtar, W. & Veenstra, G. J. C. TBP-related factors: A paradigm of diversity in transcription initiation. *Cell and Bioscience* **1**, 1–12 (2011).
338. Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A. & Blobel, G. A. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell* **62**, 237–247 (2016).
339. Petrascheck, M. *et al.* DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res.* **33**, 3743–3750 (2005).
340. Su, W., Jackson, S., Tjian, R. & Echols, H. *DNA looping between sites for transcriptional activation: self-association of DNA-bound Spl*. (1991).
341. Kulaeva, O. I., Nizovtseva, E. V., Polikanov, Y. S., Ulianov, S. V. & Studitsky, V. M. Distant Activation of Transcription: Mechanisms of Enhancer Action. *Mol. Cell. Biol.* **32**, 4892–4897 (2012).
342. Mora, A., Sandve, G. K., Gabrielsen, O. S. & Eskeland, R. In the loop: promoter-enhancer interactions and bioinformatics. *Briefings in bioinformatics* **17**, 980–995 (2016).
343. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
344. Nakagawa, T., Bulger, M., Muramatsu, M. & Ito, T. Multistep Chromatin Assembly on Supercoiled Plasmid DNA by Nucleosome Assembly Protein-1 and ATP-utilizing Chromatin Assembly and Remodeling Factor. *J. Biol. Chem.* **276**, 27384–27391 (2001).
345. Mladenova, V., Mladenov, E. & Russev, G. Organization of plasmid dna into nucleosome-like structures after transfection in eukaryotic cells. *Biotechnol. Biotechnol. Equip.* **23**, 1044–1047 (2009).
346. Jeong, S., Lauderdale, J. D. & Stein, A. Chromatin assembly on plasmid DNA in vitro. Apparent spreading of nucleosome alignment from one region of pBR327 by histone H5. *J. Mol. Biol.* **222**, 1131–1147 (1991).
347. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
348. Vogel, C. & Marcotte, E. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation Systematic bacterialization of yeast genes identifies a near-universally swappable pathway View project Retinal differentiation View project. (2014). doi:10.1038/nbt1270
349. Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: Biological functions and applications. *Nucleic Acids Research* **44**, 7511–7526 (2016).
350. Zhai, L. T. & Xiang, S. mRNA quality control at the 5′ end. *Journal of Zhejiang University: Science B* **15**, 438–443 (2014).
351. Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* **16**, 2906–2922 (2002).
352. Hinnebusch, A. G. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.* **75**, 434–467 (2011).
353. Araujo, P. R. *et al.* Before It Gets Started: Regulating Translation at the 5′ UTR. *Comp. Funct. Genomics* **2012**, 1–8 (2012).
354. Mohan, R. A. *et al.* A mutation in the Kozak sequence of *GATA4* hampers translation in a family with atrial septal defects. *Am. J. Med. Genet. Part A* **164**, 2732–2738 (2014).
355. De Angioletti, M., Lacerra, G., Sabato, V. & Carestia, C. β+45 G → C: a novel silent β-thalassaemia mutation, the first in the Kozak sequence. *Br. J. Haematol.* **124**, 224–231 (2004).

356. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
357. Chien, P., Grant, R. A., Sauer, R. T. & Baker, T. A. Structure and Substrate Specificity of an SspB Ortholog: Design Implications for AAA+ Adaptors. *Structure* **15**, 1296–1305 (2007).
358. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, 1–8 (2015).
359. Chervyachkova, E. Light-controlled self-assembly and self-sorting of cell-like compartments. (Ruperto Carola University Heidelberg, Germany, 2018).
360. Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synth. Biol.* **1**, 118–129 (2012).
361. Guntas, G. *et al.* Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 112–117 (2015).
362. Dey, A., Wong, E. T., Bist, P., Tergaonkar, V. & Lane, D. P. Nutlin-3 inhibits the NFκB pathway in a p53-dependent manner: Implications in lung cancer therapy. *Cell Cycle* **6**, 2178–2185 (2007).
363. Smith, A., Ward, M. P. & Garrett, S. *Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation*. *The EMBO Journal* **17**, (1998).
364. Thevelein, J. M. & De Winde, J. H. Novel sensing mechanisms and targets for the cAMP-protein kinase A pathway in the yeast Saccharomyces cerevisiae. *Molecular Microbiology* **33**, 904–918 (1999).
365. Azpeitia, E. & Wagner, A. Short Residence Times of DNA-Bound Transcription Factors Can Reduce Gene Expression Noise and Increase the Transmission of Information in a Gene Regulation System. *Front. Mol. Biosci.* **7**, 67 (2020).
366. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annual Review of Cell and Developmental Biology* **35**, 357–379 (2019).
367. Ramos, A. I. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, (2013).
368. Gao, R. & Stock, A. M. Temporal hierarchy of gene expression mediated by transcription factor binding affinity and activation dynamics. *MBio* **6**, (2015).
369. Wu, M. *et al.* p53 dynamics orchestrates with binding affinity to target genes for cell fate decision. *Cell Death Dis.* **8**, e3130 (2017).
370. Donczew, R. & Hahn, S. Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters. *Mol. Cell. Biol.* **38**, (2018).
371. Chen, W. & Struhl, K. Saturation mutagenesis of a yeast his3 'TATA element': genetic evidence for a specific TATA-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2691–2695 (1988).
372. Sanchez, A. & Golding, I. Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**, 1188–1193 (2013).
373. Gracey, L. E. *et al.* An in vitro-identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome in vivo. *Epigenetics and Chromatin* **3**, (2010).
374. Subtil-Rodríguez, A. & Reyes, J. C. BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. *EMBO Rep.* **11**, 751–757 (2010).
375. Ngo, T. T. M., Zhang, Q., Zhou, R., Yodh, J. G. & Ha, T. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* **160**, 1135–1144 (2015).
376. Nizovtseva, E. V. *et al.* Nucleosome-free DNA regions differentially affect distant communication in chromatin. *Nucleic Acids Res.* **45**, 3059–3067 (2017).
377. Roccatano, D., Barthel, A. & Zacharias, M. Structural flexibility of the nucleosome core

particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers* **85**, 407–421 (2007).
378. Chan, L. Y., Mugler, C. F., Heinrich, S., Vallotton, P. & Weis, K. Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *Elife* **7**, (2018).
379. Franks, A., Airoldi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS Comput. Biol.* **13**, (2017).
380. Silva, G. M. & Vogel, C. Quantifying gene expression: the importance of being subtle. *Mol. Syst. Biol.* **12**, 885 (2016).
381. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
382. Mata, J., Marguerat, S. & Bähler, J. Post-transcriptional control of gene expression: A genome-wide perspective. *Trends in Biochemical Sciences* **30**, 506–514 (2005).
383. Vogel, C. *et al.* Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* **6**, 400 (2010).
384. Salovska, B. *et al.* Isoform-resolved correlation analysis between <scp>mRNA</scp> abundance regulation and protein level degradation. *Mol. Syst. Biol.* **16**, (2020).
385. Ravindran, P. T., McFann, S. & Toettcher, J. E. A synthetic gene circuit for imaging-free detection of dynamic cell signaling. *bioRxiv* 2021.01.06.425615 (2021). doi:10.1101/2021.01.06.425615
386. Lebar, T., Lainšček, D., Merljak, E., Aupič, J. & Jerala, R. A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat. Chem. Biol.* **16**, 513–519 (2020).
387. Mesrouze, Y. *et al.* Dissection of the interaction between the intrinsically disordered YAP protein and the transcription factor TEAD. *Elife* **6**, (2017).
388. Camargo, F. D. *et al.* YAP1 Increases Organ Size and Expands Undifferentiated Progenitor Cells. *Curr. Biol.* **17**, 2054–2060 (2007).
389. Norton, L. J. *et al.* Direct competition between DNA binding factors highlights the role of Krüppel-like Factor 1 in the erythroid/megakaryocyte switch. *Sci. Rep.* **7**, 1–10 (2017).
390. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
391. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
392. De Bruin, L. & Maddocks, J. H. CgDNAweb: A web interface to the cgDNA sequence-dependent coarse-grain model of double-stranded DNA. *Nucleic Acids Res.* **46**, W5–W10 (2018).
393. Delano, W. L. The PyMOL Molecular Graphics System. (2002).
394. Çiçek, Ö., Marrakchi, Y., Antwi, E. B., Di Ventura, B. & Brox, T. Recovering the Imperfect: Cell Segmentation in the Presence of Dynamically Localized Proteins. *MICCAI Work. Med. Image Learn. with Less Labels Imperfect Data* (2020). doi:10.1007/978-3-030-61166-8_9
395. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. *ICCV* (2017). doi:10.1109/TPAMI.2018.2844175
396. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI* (2015).
397. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *NIPS* (2017).
398. Ilg, E. *et al.* Uncertainty Estimates and Multi-Hypotheses Networks for Optical Flow. *ECCV* (2018).

5 MATERIALS AND METHODS

5.1 Decoding transcription factor dynamics

5.1.1 Plasmid Construction

All plasmids were designed using the Gibson assembly protocol [390]. The pDN98 plasmid [247] was used as parental plasmid for the construction of our synthetic TF. LexA dimerization domain which was amplified from a lab strain *E.coli* was insert between the LexA DBD and the VP48 transactivation domain to generate synTF. Different combinations of nuclear export signals (weaker PKIt NES and relatively strong IkBα NES) and different nuclear localization signals (biNLS2, biNLS10 [247]) were tested for their nuclear accumulation in the dark and upon blue light activation.

The pDN100 plasmid [247] was used as the parental plasmid for the construction of the initial synthetic reporter. The iRFP670 was amplified from the pNLS-iRFP670 plasmid [333] with a primer containing the CAAX motif. This was inserted in place of the firefly luciferase gene. The synPlasmid1 was constructed by amplifying the synthetic promoter together with reporter gene and bovine growth hormone (BGH) terminator sequences and inserting the amplicon upstream of CMV promoter of the TF construct in order to be able to infer the reporter transfection efficiency from the TF expression levels.

5.1.2 LOVtrap experimental setup

To express LOVtrap (signal peptide fused to Zdk2 with/without mVenus) at a level similar to the synthetic TF, they were both expressed as a single fusion protein separated by T2A sequence. This ensured that there was sufficient LOVtrap molecules to sequester the TFs of interest out of the nucleus in the dark. LOVtrap in Figure 6.4 was however expressed under the CMV promoter while the corresponding synTFs were expressed under the IRES. This ensured better retention of synTF out of the nucleus due to the higher LOVtrap to synTF concentration ratio.

5.1.3 Nascent mRNA visualization construct:

For transcript visualization, MS2-coat protein (MCP) was amplified from addgene plasmid #52985 (ubc-nls-ha-MCP-VenusN-nls-ha-PCP-VenusC) [154]. The IRES-SV40/NLS-MCP-mVenus gene sequence was inserted after the stop codon of the synthetic TF. From the Pcr4-12xMBS-PBS plasmid [154], the 12xMBS-PBS sequence was amplified and was inserted after the stop codon of the reporter gene sequence. This sequence was used for visualization of transcript using the MCP-mVenus fusion protein because it more stable and yields relatively bigger foci due to the longer sequence but only after the BGH promoter was removed to allow for longer 3' UTR.

5.1.4 Mammalian cell culture and transfection:

HEK293T cells were maintained at 37°C and 5% CO_2 in phenol red-free Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal calf serum (Sigma-Aldrich), 2mM L-glutamine (Gibco), 100U/ml penicillin and 10mg/ml streptomycin (Gibco). For confocal microscopy quantification of mCherry-synTF and iRFP670 reporter proteins, $7.5x10^4$ HEK293T cells were seeded into the four-chambered 35mm Quad coverslip bottom dishes (ibidi). The next day, cells were transfected with 500ng of total DNA per chamber using the calcium phosphate transfection protocol. Total DNA comprised of synPlasmid which contained

both reporter and synTF constructs with/without transcript visualization constructs and pBlue-ScriptIIS/K stuffer plasmid in a ratio of 1:200. Microscopy of synTF and reporter protein expressed under different synTF dynamics was performed 24h post transfection.

5.1.5 Imaging of synTF dynamics and reporter expression:

Cell were maintained at 37°C and 5% CO2 in a dark incubation chamber for the duration of microscopy session. Confocal microscopy was done on a Zeiss LSM 800 hybrid microscope with a motorized stage and laser module containing 405, 488, 561 and 640nm lasers and an electronically switchable illumination and detection module. Images of synTF-mCherry were acquired with the following settings: 0.15% of 561nm excitation laser using 37μm pinhole aperture and 700V gain. Images of iRFP670 reporter protein were taken with 0.30% of 640nm excitation laser using 41μm pinhole aperture and 650V gain. This hybrid microscope was also equipped with a colibri consisting of 385, 425, 469, 511, 555 and 631nm LED light sources for widefield epifluorescence microscopy. Blue light activation was performed by exciting cells with 0.5% intensity of 469nm LED light in widefield microscopy mode using the 38 HE GFP filter set from Zeiss. The 0.5% light intensity which corresponded to 6.44Wm^{-2}w as measured with LI-250A light sensor (LI-COR Biosciences) was achieved by blocking 75% of 2% LED intensity. Automated cell focusing was done using mCherry as the reference channel. Constant blue light activation was performed by illuminating cells with blue light for 125ms every 45s for 2h. Post activation, confocal mCherry and iRFP670 images were taken every 30 minutes for 5hrs before experiments were terminated. Both mCherry and iRFP670 confocal images were also taken prior at the beginning of the experiment. For pulsatile activations, cells were illuminated with the same illumination scheme for 15 minutes and then stopped for the desired time (15 or 30 minutes) before further activation. For time-lapse tracking of synTF-mCherry localization, confocal mCherry images were taken every 5min during activation and dark recovery phases. All image acquisitions were done using the ZenBlue software. All reporter protein quantification experiments were done with the plan Apochromat 40x/1.4 numerical aperture oil immersion objective (Zeiss).

5.1.6 Transcript visualization and quantification:

Nascent mRNA foci were visualized on the same microscope using an Axiocam503 camera and 63x/1.4 plan Apochromat oil immersion objective (Zeiss). Blue light activation of cells was performed with the same setup as above except 0.95% 469nm LED light intensity was used to account for the change of objective. This LED intensity corresponded to 6.79Wm^{-2} of light. Transcript foci images were acquired using 5% 511nm LED light in widefield microscopy mode and 46 HE YFP filter set from Zeiss. Transcript foci image acquisition was performed in a Z-stack of 16-18 sections with 0.75μm step size to ensure no foci is missed due to their positioning in the nucleus. Images of the foci were taken every 5min but only during the blue light activation times since the 511nm wavelength can partially activate the LOV domain. Cells were observed to be stressed after long periods of Z-stack imaging. I therefore used six repeats of 15 min blue light activation instead of eight to generate pulsatile dynamics during live cell imaging of mRNA transcription.

The maximum projection of the z-stacks were computed in ImageJ [391] to bring all mRNA foci onto the same plane. Time course images of individual cells were cropped to allow single cell foci quantification over time. We developed a python script that 1) identifies nascent mRNA foci in the maximum intensity projected time course images 2) fits a 2-dimensional Gaussian

to the identified foci and 3) calculate the volume under the fitted Gaussian as mRNA expression intensity. Parameters of the Gaussian fit were limited to exclude background intensity fluctuations and large aggregates of the fluorescent protein.

5.1.7 Flexibility and bendability of 601 and Random sequence

RandSeq was generated using the http://www.faculty.ucr.edu/~mmaduro/random.htm server. Bendability and flexibility of 601 and RandSeq models were generated using the https://cgdnaweb.epfl.ch/ webserver with paramset 4 [392]. The .pdb models were then aligned based on their coordinates using PyMol 2.4.1 [393].

5.1.8 Automated quantification protein signals:

The algorithm for instance segmentation of the nucleus and cytosol and uncertainty-based nuclei propagation were generated by Özgün Çiçek while Yassine Marrakchi implemented measurement of nucleus, cytosol and plasma membrane signals. I provided the manual annotation for training the network and contributed to coding of plasma membrane segmentation.

5.1.8.1 Instance segmentation with Uncertainty estimation

We stick to the same base network as in [394] and use the off-the-shelf Mask R-CNN [395] trained with channel 0 only. To augment the data and create nearly realistic input images, we also use the elastic deformations of U-Net [396] to help improve the generalization capabilities of the network. To detect segmentation errors we use both data uncertainty (aleatoric) and model uncertainty (epistemic). We model the former by learning the noise scale during training and computing the entropy of the class pseudo-probabilities for each pixel at test time as in [397]. For the later, we use the Winner-Takes-All [398] which trains a single network with multiple heads and only updates the head with the best prediction every iteration. We choose this combination since it performs best in [394].

5.1.8.2 Uncertainty-based Nuclei Propagation

To improve the output of Mask R-CNN, we compute the tracks as described in [394]. We stick to the suggested hyperparamters $\alpha = 0.7$ and $\beta = 0.85$ to decide about frames that need to be updated. We consider a simple yet effective warping strategy by estimating the shift and scaling parameters computed between the not yet updated and neighbouring nuclei predictions. Likewise, we implicitly assume that the shape of the nuclei does not change over small time windows and only allow slight deformations to occur. Although flow based methods tend to perform better according to [394], we do not use them to reduce the computational burden. To mitigate this slight drop in performance, we apply a sampling strategy before measuring the fluorescence as discussed in the next section.

5.1.8.3 Measurement

We report the average fluorescence of nucleus, cytosol and membrane per cell and frame. Instead of using the full prediction mask to compute the average, we sample a subset of pixels that have higher chances to belong to the corresponding structure. For nucleus and cytosol, segmentation errors occur mainly on the border. Therefore, we gradually erode the segmentation mask as long as it contains more 2000 pixels. We then superpose the binary mask with channel 0 and compute the average signal. Measuring the fluorescence for the membrane is very challenging since it is a very thin structure. Moreover, touching cells cause interference that amplify the signal. Thus, we use iRFP670 channel and compute a skeleton.

Notice that this skeleton might miss cells because of very low signal and might add artefacts in case of very high signal over surfaces. To avoid these errors, we rely on the cytosol masks and compute candidate pixels for the membrane by dilating the cytosol once. Then we remove intersections between candidate membranes of touching cells and pixels that are very close to the border of the image. Finally, we superpose the skeleton and the candidate membranes and compute the average based on the intersection. If there is no signal in the skeleton, we completely rely on the candidate membrane inferred from the cytosol.

5.1.9 Mathematical model of synthetic promoter response to synTF

5.1.9.1 State variables:

TF: cytosolic transcription factor
NTF: nuclear transcription factor
P_{r1}: unbound promoter
P_{r2}: bound promoter
P_{r3}: refractory promoter
PIC: pre-initiation complex
nRNA: nascent mRNA
mRNA: matured mRNA
P: unfolded protein
mP: matured protein

5.1.9.2 Input:

liAct(t): external stimulus input (blue light induced nuclear accumulation of synTF).

$$
\begin{gathered}
def\ liAct(t, tOn, tOff, uOn, uOff): \\
if\ (t <= ill): \\
if\ (t\%(tOn + tOff)\ <\ tOn): \\
u\ =\ uOn \\
elif\ (t\%(tOn + tOff)\ >=\ tOn): \\
u\ =\ uOff \\
else: \\
u\ =\ uOff \\
return(u)
\end{gathered}
$$

drkrev(t): export of synTF out of the nucleus in the absence of blue light.

$$
\begin{gathered}
def\ drkrec(t, tOn, tOff, k1, k2): \\
if\ (t <= ill): \\
if\ (t\%(tOn + tOff)\ <\ tOn): \\
k = k1 \\
elif\ (t\%(tOn + tOff)\ >=\ tOn): \\
k = k2 \\
else: \\
k = k2 \\
return(k)
\end{gathered}
$$

Where t, tOn, tOff, uOn, uOff, k1 and k2 represent time, blue light 'ON' duration, blue light 'OFF' duration, synTF import rate under blue light, import rate in the dark, export rate under blue light and export rate in the dark respectively.

The functions liAct(t) and drkrev(t) was designed to results in different import and export rates in dark and under blight activation since the two rates were dynamics as a result of the caging and uncaging of the NLS (equations 1&2). SynTF dynamics were simulated based on our experimental set up.

We modelled gene expression regulation with the following three promoter states; unbound, bound/active and refractory states which was necessary to account for the refractory mRNA expression behaviour observed for some of our promoters (equations 3-5). We then expanded mRNA transcription to include PIC nucleation, transcription of nascent mRNA and the maturation of the mRNAs in order to account for the role of DNA looping prior to transcription initiation and the effect of pre-mRNA processing (equations 6-8). Due to the long maturation time of our reporter protein, we modelled protein expression and protein maturation separately (equations 9&10). Equation 11 describes the dependency of DNA looping efficiency on the length of the DNA distance between two points as previously described [200]. Equations 8&9 were updated to equations 12&13 for fitting data in Figure 2.1.16.

5.1.9.3 Parameters

Nuclear import and export of synTF was simulated with the assumption that import and export rates change between dark and light activation phases. The rates were also not zero under any of the experimental conditions due to the competing strengths of the NES and the caged NLS. There were a total of 22 parameters for the final model that was used to describe promoter p12 reporter expression. Parameter R_p was fixed based on the maturation half-life of iRFP670. Four of the remaining parameters were fixed dependent on the RE or TATA-box used. Twelve other parameters were fixed after fitting them for the first promoter since they were not promoter dependent (those are labelled as 'fixed after first fit' in table 5.1). The remaining five parameters kon, koff, Krf, d1rf and d2rf were varied for each combination of promoter elements. It was reasoned that transitioning among promoter states was case specific and dependent on how refractory a promoter was. Parameters Kin and Din were used to describe our fourth promoter as presented in Figure 3.1.

Mathematical modelling was conducted in python v3.8.3.final.0 using Anaconda v2020.07 via jupyter in the anaconda-navigator v1.9.12 from anaconda-project v0.8.4. Numerical simulations were performed using the odeint function in SciPy v1.5.0 scipy.integrate module which is used as a wrapper for the LSODA ordinary differential equation solver for stiff or non-stiff systems from FORTRAN library odepack.

After defining the TF dynamics input functions liAct(t) and drkrev(t), we defined the function which captured the description of the model. Initial conditions, state0 were set according to experimental observations or were fitted. State variables of interest were plotted using matplotlib plotting library.

Table 5.1 | Description of model parameters and how they were varied.

Parameter	Description	Value	Variability
m	Cooperativity of TF binding	2.4588	Fixed for RE
Kd_1	Affinity of RE/TF binding	161.1805	Fixed for RE
n	Cooperativity of TBP binding	3.2224	Fixed for TATA-box
Kd_2	Affinity of TATA box/TBP binding	400.0	Fixed for TATA-box
Kon	Maximum rate of TF binding	1.1422e-1	Variable
Koff	TF dissociation rate	1.1816e-3	Variable
Krf	Maximum rate conversion to refractory promoter	4.8033e-1	Variable
d1rf	Refractory to unbound promoter rate	2.9010e-3	Variable
d2rf	Refractory to bound promoter rate	7.2893e-4	Variable
Kpic	Maximum rate of PIC formation	14.7496	Fixed after first fit
Dpic	Rate of PIC dissociation	5.5526e-1	Fixed after first fit
K_{nRNA}	Maximum rate of transcription	2.7933e-03	Fixed after first fit
d_{nRNA}	Dissociation and degradation of nascent mRNA	2.0238e-02	Fixed after first fit
K_{mRNA}	Maturation of rate of nascent mRNA	96.2841	Fixed after first fit
d_{mRNA}	Degradation of nascent mRNA	6.7317e-02	Fixed after first fit
k_P	Translation rate	56.4272	Fixed after first fit
d_P	Degradation rate of unfolded reporter protein	3.9477e-04	Fixed after first fit
d_{mP}	Degradation of matured reporter protein	7.3672e-06	Fixed after first fit
R_P	iRFP670 maturation rate	2.3105e-03	Fixed
K_m	nRNA concentration to achieve 50% K_{mRNA}	59.8827	Fixed after first fit
j	Cooperativity of ribosomal occupancy	9.9732	Fixed after first fit
Kd_P	mRNA concentration to achieve 50% k_P	102.3959	Fixed after first fit

Parameter values shown above correspond to the parameters used to describe promoter p12. However, only the parameters designated ‘variable’ under the variability column to specific only for promoter p12.

5.1.10 Parameter fitting

Parameter fitting was done with data from the sustained synTF dynamics. The fitted parameters were used to predict the two pulsatile activations except when promoter design does not involve the change of binding affinity (example, promoter p5 where only the #RE was changed). In such cases, predictions were directly made without the need for further fitting. Parameters were fit fitted with the least square function (leastsq) from the lmfit minimize module. The output parameters were then used as the initial parameter guesses for global parameter fit using the adaptive memory programming for global optimization method (ampgo). To improve on the global parameter values, the ampgo fitted output was used as initiation guesses for another round of leastsq function fitting to find the local global minimized parameters.

5.2 TF architecture

5.2.1 Construction of synTEAD/synYAP

The synTEAD was constructed by fusing a subunit of a heterodimer pair to the N-terminus of NLS-LexA (intact) fusion protein. The gene for the fusion protein was placed under IRES or UBC promoter to regulate the level of expression of synTEAD. The synYAP which bears the TAD was constructed by replacing the LexA in synTF with the other subunit of the heterodimer pair. The expression of synYAP was under the CMV promoter. In the case of SYNZIP heterodimer pairs listed in

Table 6.4, they were arranged such that they had parallel arrangement during their interaction. Similar to the arrangement in synPlasmid1, the reporter was inserted tandemly upstream of the CMV promoter. No mRNA visualization constructs were included.

5.2.2 LOVtrap

In all cases, LOVtrap was constructed by fusing Zdk2 to an NES. Zdk2 was either fused to the C-terminus of mVenus which had NES fused to its N-terminus or directly to the C-terminus of NES. Plasma membrane and mitochondria localized LOVtrap made cells look unhealthy and so were not used further.

5.2.3 Blue light illumination of synYAP co-activator

Similar microscopy equipment and setup as described in section 5.1.5 was used for synYAP visualization. Blue light activation was however performed by illuminating cells with blue light for 125ms every 22.5s for 2h. This was done to obtain higher nuclear accumulation for better synTEAD/synYAP reporter expression.

5.2.4 Construction of synTEAD/synYAP without LINuS

To construct a synTEAD/synYAP TF that mimics the nucleocytoplasmic translocation endogenous TEAD/YAP transcription, the NES and NLS in synYAP construct described above were removed. Furthermore, the truncated Jα helix was replaced with an intact one. The synYAP2 construct was then fused to P2A-LOVtrap for a higher expression of the LOVtrap which would be under the direct control of the CMV promoter as synYAP2. The synTEAD was also either put under the T2A or UBC promoter.

5.2.5 QUANTIFICATION AND STATISTICAL ANALYSIS

All protein quantifications were performed using our in-house built neuronal network while transcript quantification were done using a python script. Statistical analysis were also performed using python. Data are typically presented as mean ± 95% confidence interval as specified in the figure legends.

6 APPENDIX

6.1 Supplementary Figures

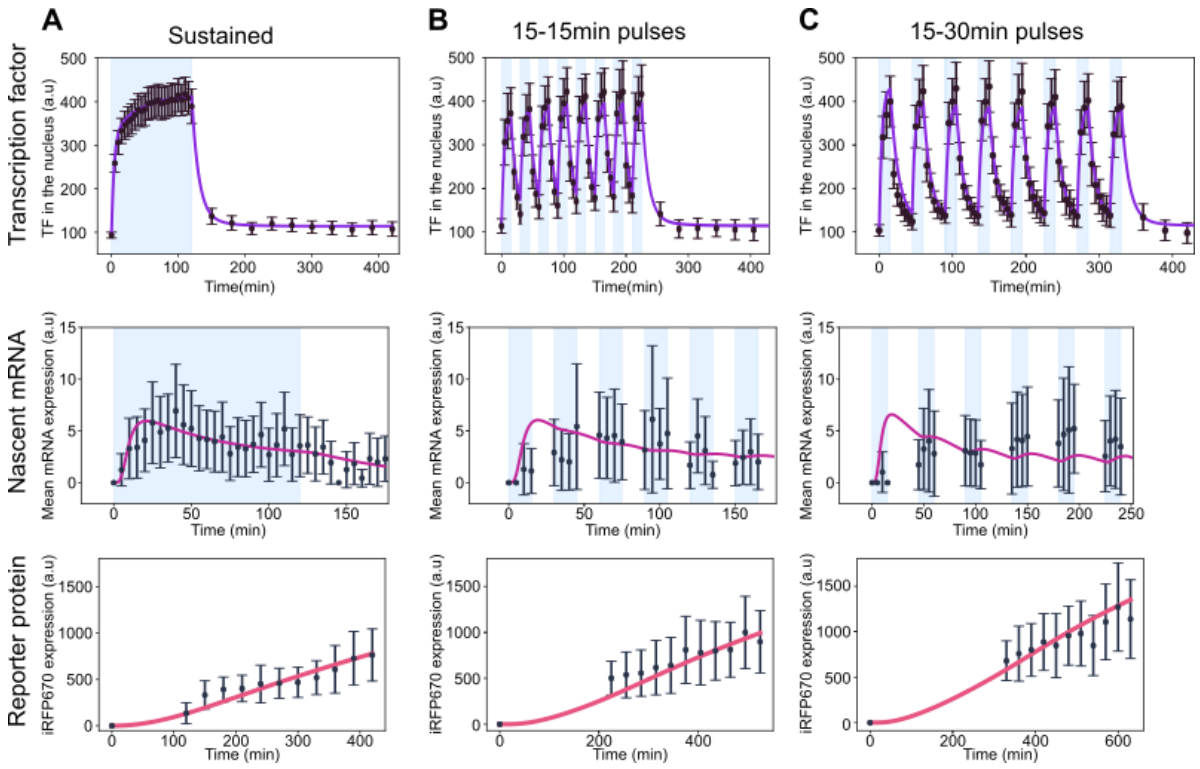

Figure 6.1 | Promoter with low affinity TATA-box and high affinity RE induces high reporter expression. A) The nuclear TF level upon 2h constant blue light activation (top), mean nascent mRNA foci volume (middle) and iRFP670 reporter protein expression (bottom). B) Nuclear TF level upon 15-15 min pulsatile blue light activation (top), mean nascent mRNA foci volume (middle) and iRFP670 protein expression levels (bottom). C) Nuclear TF level upon 15-30 min pulsatile blue light activation (top), mean nascent mRNA foci volume (middle) and iRFP670 protein expression (bottom).

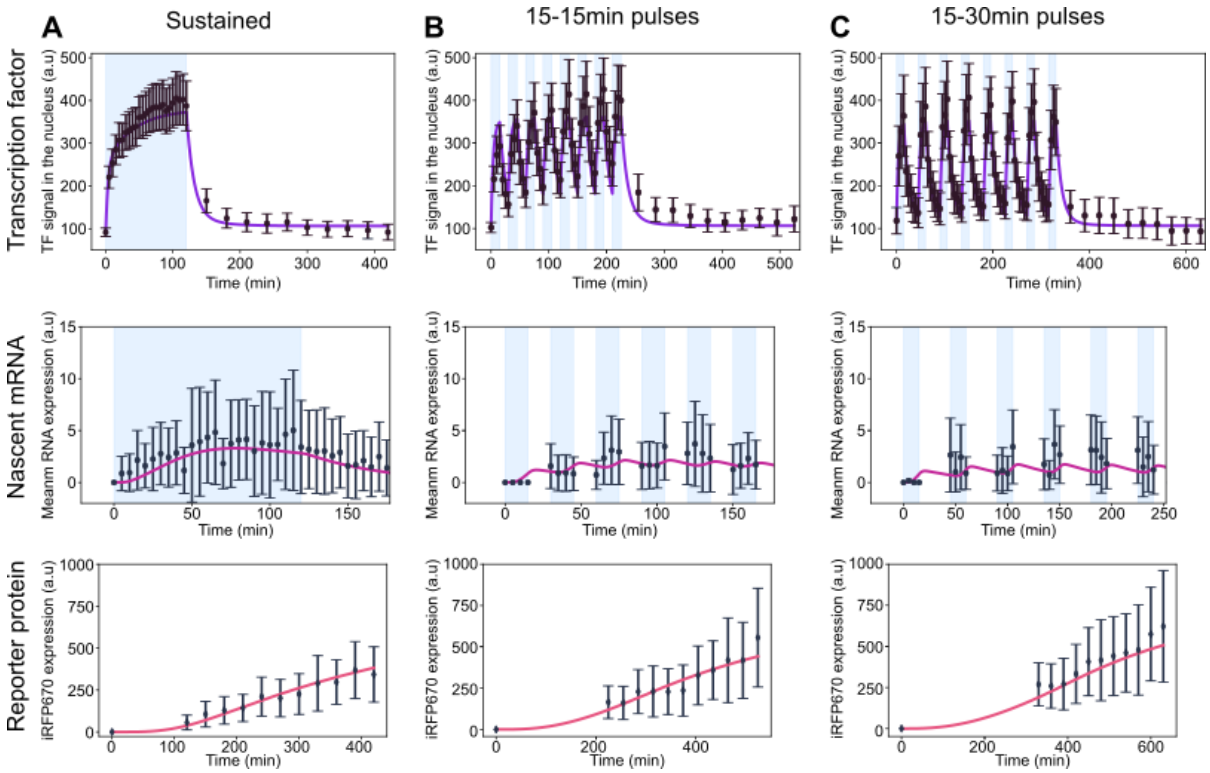

Figure 6.2 | Promoter with low affinity RE and high affinity TATA-box displays no refractory response. A) The nuclear TF level upon 2h constant blue light activation (top), mean nascent mRNA foci volume (middle) and iRFP670 reporter protein expression (bottom). B) Nuclear TF level upon 15-15 min pulsatile blue light activation (top), mean nascent mRNA foci volume (middle) and iRFP670 protein expression levels (bottom). C) Nuclear TF level upon 15-30 min pulsatile blue light activation (top), mean nascent mRNA foci volume (middle) and iRFP670 protein expression (bottom).

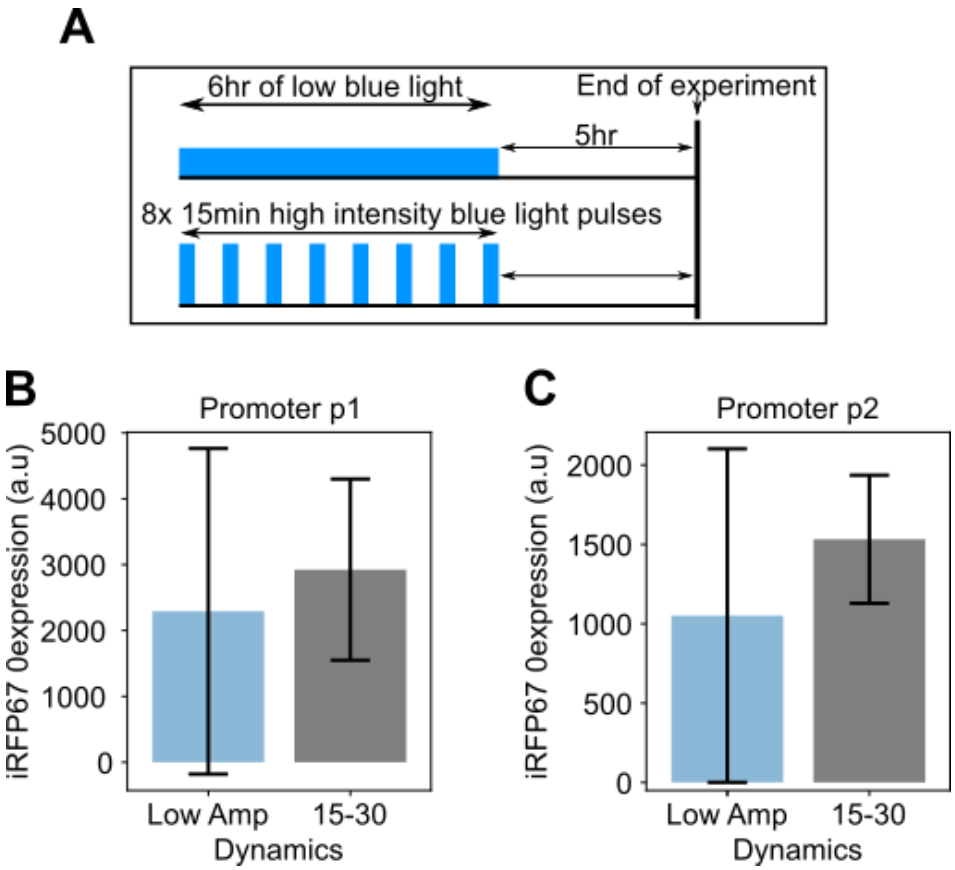

Figure 6.3 | Reporter expression induced by low amplitude sustained and high amplitude pulsatile synTF dynamics. A) Scheme of light activation. B) Reporter protein expression induced by promoter p1 and B) promoter p2. Low-Amp and 15-30 represent low amplitude sustained synTF and high amplitude 15-30 pulsatile synTF dynamics. Errors bars represent 95% confidence interval of mean of at least 20 individual cell measurements.

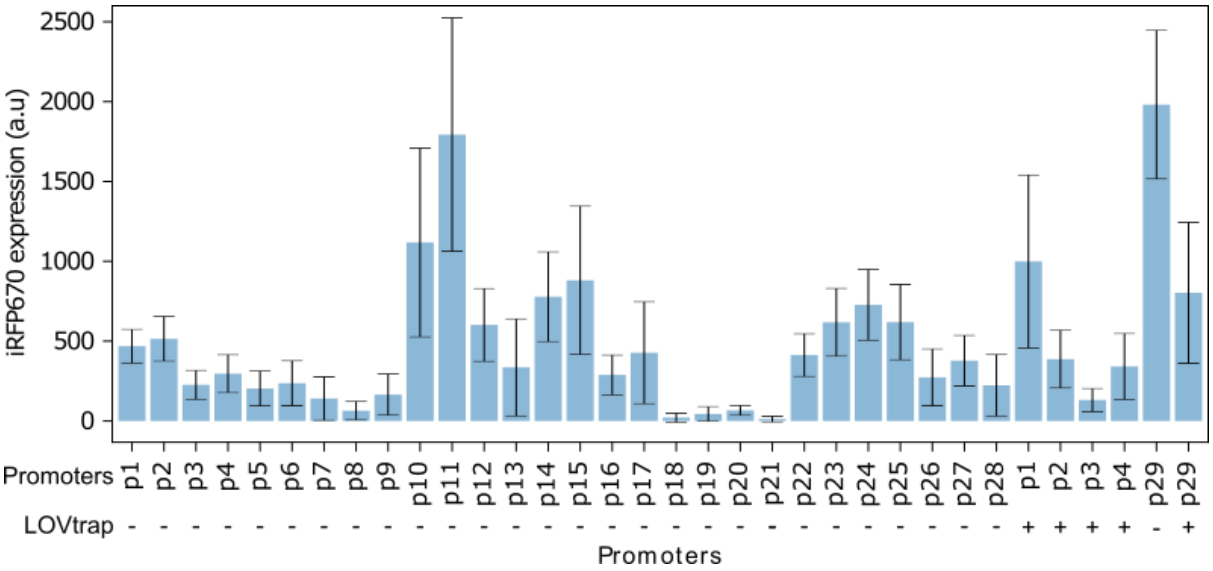

Figure 6.4 | Reporter protein expression induced by different promoter designs under sustained synTF dynamics. Error bars represent 95% confidence interval of the mean of at least 20 individual cell measurements.

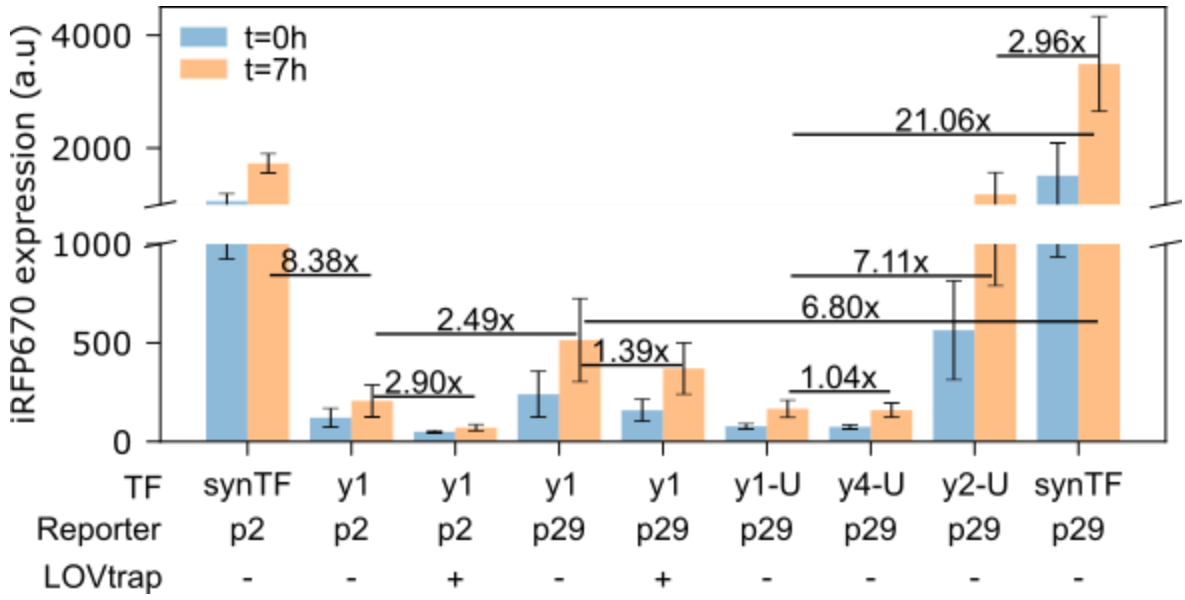

Figure 6.5 | Reporter expression under different TF setups and different interaction affinities. Under sustained nuclear accumulation, synTF induced higher reporter gene expression even under the relatively weaker p2 promoter.

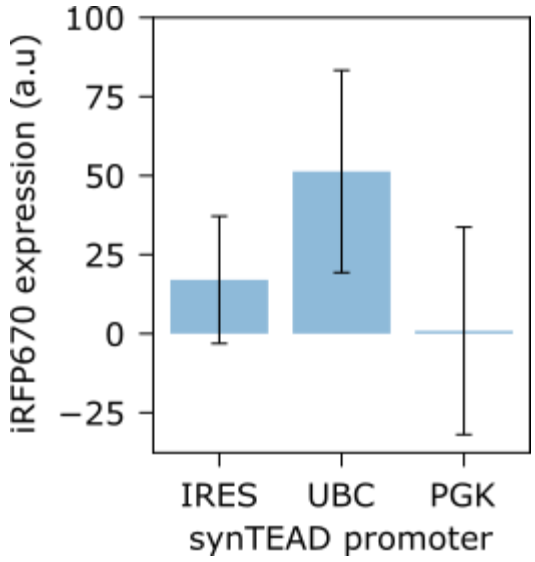

Figure 6.6 | Reporter expression under different levels of synTEAD. There is a limit to how low synTEAD expression can be and still induce good reporter expression. IRES induces the highest synTEAD expression followed by UBC – with PGK inducing the least synTEAD expression. Reporter expression was barely detected when synTEAD was expressed under the PGK promoter. The experiment was run with synYAP consisting of PKIt NES and biNLS2 based LINuS. This was combined with the LOVtrap to ensure better cytosolic retention.

6.2 Supplementary Tables

Table 6.1 | Transcription factor constructs.

	DBD	TAD	NES	NLS	Expression
TF1	LexADBD[a]	VP48[a]	PKIt[a]	biNLS2[a]	Low
TF2	LexA DBD	VP48	IkBα[a]	biNLS10[a]	High
SYNTF	Full LexA[b]	VP48	IkBa	biNLS10	High
TF3	Full LexA	P53 TAD1	IkBa	biNLS10	NA
TF4	Full LexA	FOXM1c TAD1	IkBa	biNLS10	Very low
TF5	Full LexA	vp16	IkBa	biNLS10	Very low
TF CONTROL	Full LexA	-	IkBa	biNLS10	NA

Synthetic TF designed, constructed and tested in this study. Column labelled 'Expression' describe the expression of iRFP670 reporter under promoter p2 when induced by sustained dynamics.

a) [247].

b) Amplified from lab strain *E. coli*.

Table 6.2 | List and composition of synthetic promoters.

Promoter	nRE	K_D(RE)	λ	K_D(TB)	δ	Sensitive
p1	4	1.61nM	49 bp	2nM	69bp	No
p2	4	1.61nM	49 bp	4nM	69 bp	No
p3	4	5.64nM	49 bp	2nM	69 bp	No
p4	4	5.64nM	49 bp	4nM	69 bp	Yes
p5	2	1.61nM	49 bp	2nM	69 bp	No
p6	2	1.61nM	49 bp	4nM	69 bp	
p7	4	1.61nM	196[a]bp	4nM	69 bp	No
p8	4	1.61nM	196[b]bp	4nM	69 bp	
p9	4	1.61nM	196[b]bp	2nM	69 bp	Yes
p10	4	1.61nM	49bp	4nM	31[c]bp	No
p11	4	1.61nM	49bp	4nM	31[d]bp	No
p12	4	1.61nM	49bp	4nM	31[e]bp	Yes
P13	4	5.64nM	49 bp	4nM	31bp	Yes
P14	4	0.80nM	49 bp	4nM	69 bp	No
P15	4	0.80nM	49 bp	2nM	69 bp	No
P16	4	0.80nM	49 bp	2nM	31bp	No
P17	2	0.80nM	49 bp	2nM	69 bp	No

Supplementary Table 6.2 continued

P18	4	1.61nM	343[b]bp	4nM	69 bp	
P19	4	1.61nM	343[f]bp	4nM	69 bp	
P20	4	1.61nM	343[b]bp	2nM	69 bp	
P21	4	0.80nM	343[b]bp	2nM	31bp	
P22	4	1.61nM	49bp	4nM	31[g]bp	No
P23	4	1.61nM	49bp	4nM	31[h]bp	No
P24	4	1.61nM	49bp	4nM	31[i]bp	No
P25	4	1.61nM	49bp	4nM	31[j]bp	No
P26	4	1.61nM	49 bp	4nM	178bp	No
P27	4	5.64nM	49 bp	2nM	176bp	Yes
P28	4	1.67nM	49 bp	4nM	69 bp	No
P29	4	1.61nM	49 bp	2nM	31bp	
P30	4	0.80nM	343[b]bp	4nM	69 bp	
P34	4	1.61nM	196bp	2nM	31bp	
P35	2	1.61nM	196[a]bp	4nM	69 bp	
P36[k]	4	1.61nM	343[a]bp	4nM	69 bp	
P37	8	1.61nM	49 bp	4nM	69 bp	No
P38	8	1.61nM	49 bp	4nM	31bp	No
P39	8	1.61nM	343[b]bp	4nM	69 bp	
P40	1	1.61nM	49 bp	4nM	69 bp	
P41	2	1.61nM	196[a]bp	4nM	69 bp	
P42	4	1.61nM	49 bp	4nM	101bp	
P43	4	1.61nM	49 bp	-	31 bp	
P44[l]	4	1.61nM	49 bp	-	31 bp	
P45[m]	4	1.61nM	49 bp	-	69 bp	
P46[m]	4	1.61nM	49 bp	-	101bp	
P47	7	1.61nM	343[f]bp	4nM	69 bp	
P48	4	1.61nM	49bp	4nM	31[n]bp	No
P49	4	1.61nM	49bp	4nM	31[o]bp	No

The list of synthetic promoters constructed or combination of promoter elements tested this study.

a) Insertion of 601 nucleosome positioning sequence between REs and TATA-box.

b) Insertion of randomly generated sequence between REs and TATA-box.

c) Promoter with A at position -6 of ATG and kozak sequence.

d) Promoter with G at position -6 of ATG and kozak sequence.

e) Promoter with G at position -6 of ATG without a kozak sequence.

f) Promoter with 2x RandSeq sequence flanked by CTCF binding sequence on both 5' and 3' ends.

g) Promoter with 55% GC content between TATA-box and translation start site including TA tract of 6 nucleotides and kozak sequence.

h) Promoter with 62% GC content between TATA-box and translation start site without a canonical kozak sequence.

i) Promoter with 55% GC content between TATA-box and translation start site including kozak sequence without a TA tract of 6 nucleotides.

j) Promoter with 48% GC content between TATA-box and translation start site without a kozak sequence.

k) Promoter with 2x 601 nucleosome positioning sequence.

l) Promoter with 2x initiator sequence in place of the TATA-box.

m) Promoter with 2x initiator sequence in place of the TATA-box.

n) Promoter with 69% GC content between TATA-box and translation start site without a canonical kozak sequence.

o) Promoter with 69% GC content between TATA-box and translation start site including a canonical kozak sequence.

Table 6.3 | Transcription factor dynamics plasmid constructs. All pEA plasmids contained both synTF and reporter constructs arranged in a tandem configuration as in synPlasmid1.

Name	Backbone	Insert	Promoter	Source
pDN98	pmCherry-N1	SynTF	CMV	[247]
pDN100	pFR-Luc	Firefly luciferase	4xLexA operator based promoter	[247]
pEA0	pDN100	iRFP670-CAAX	4xLexA operator based promoter	This study
pEA1	pDN98	Promoter p1 - iRFP670-CAAX	-	This study
pEA002	pEA1	Promoter p2	-	This study
pEA003	pEA1	Promoter p3	-	This study
pEA004	pEA1	Promoter p4	-	This study
pEA005	pEA1	Promoter p5	-	This study
pEA006	pEA1	Promoter p6	-	This study
pEA007	pEA1	Promoter p7	-	This study
pEA008	pEA1	Promoter p8		
pEA009	pEA1	Promoter p9	-	This study
pEA010	pEA1	Promoter p10	-	This study
pEA011	pEA1	Promoter p11	-	This study
pEA012	pEA1	Promoter p12	-	This study
pEA013	pEA1	Promoter p13	-	This study
pEA014	pEA1	Promoter p14	-	This study
pEA015	pEA1	Promoter p15	-	This study
pEA016	pEA1	Promoter p16	-	This study
pEA017	pEA1	Promoter p17	-	This study
pEA018	pEA1	Promoter p18	-	This study

Supplementary Table 6.3 continued

pEA019	pEA1	Promoter p19	-	This study
pEA020	pEA1	Promoter p20		
pEA021	pEA1	Promoter p21	-	This study
pEA022	pEA1	Promoter p22	-	This study
pEA023	pEA1	Promoter p23	-	This study
pEA024	pEA1	Promoter p24	-	This study
pEA025	pEA1	Promoter p25	-	This study
pEA026	pEA1	Promoter p26	-	This study
pEA027	pEA1	Promoter p27	-	This study
pEA028	pEA1	Promoter p28	-	This study
pEA029	pEA1	Promoter p29	-	This study
pEA030	pEA1	Promoter p30	-	This study
pEA031	pEA1	Promoter p31	-	This study
pEA032	pEA1	Promoter p32		
pEA033	pEA1	Promoter p33	-	This study
pEA034	pEA1	Promoter p34	-	This study
pEA035	pEA1	Promoter p35	-	This study
pEA036	pEA1	Promoter p36	-	This study
pEA037	pEA1	Promoter p37	-	This study
pEA038	pEA1	Promoter p38	-	This study
pEA039	pEA1	Promoter p39	-	This study
pEA040	pEA1	Promoter p40	-	This study
pEA041	pEA1	Promoter p41	-	This study
pEA042	pEA1	Promoter p42	-	This study
pEA043	pEA1	Promoter p43	-	This study
pEA044	pEA1	Promoter p44		
pEA045	pEA1	Promoter p45	-	This study
pEA046	pEA1	Promoter p46	-	This study
pEA047	pEA1	Promoter p47	-	This study
pEA048	pEA1	Promoter p48	-	This study
pEA049	pEA1	Promoter p49	-	This study
pEAm	pEA1	MCP-mVenus	-	This study
pEAm0	pEAm	12x MCP-PCP	-	This study
pEAm1	pEAm0	Minus BGH terminator	-	This study
pEAm2	pEAm1	Promoter p2	-	This study
pEAm3	pEAm1	Promoter p3	-	This study
pEAm4	pEAm1	Promoter p4		
pEAm5	pEAm1	Promoter p5	-	This study
pEAm6	pEAm1	Promoter p6	-	This study
pEAm7	pEAm1	Promoter p7	-	This study
pEAm8	pEAm1	Promoter p8	-	This study
pEAm9	pEAm1	Promoter p9	-	This study
pEAm12	pEAm1	Promoter p12	-	This study

Table 6.4 | Heterodimer pairs.

A	B	K_D	Reference
SsRA (*C. crecentus*)	SsPB (*E.coli*)	240nM – 6μM	[357]
pMag	nMagHigh	>41μM [359]	[358]
SynZIP1	SynZIP2	<10nM	[360]
SynZIP3	SynZIP4	<30nM	[360]

Hetero-dimer domains were deployed to facilitate interaction between the two different fusion proteins. This strategy was used in this study to mimic the interaction between TEAD/YAP.

Table 6.5 | synTEAD/synYAP plasmids.

Name	Backbone	Promoter	A	B	Separator	LOVtrap
pEAY0	pEA1	cmv	SsRA-VP48_IKBa-mCherry-AsLOV-biNLS10	SsPB-biNLS2-LexA	IRES	-
pEAY1	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-biNLS2-LexA	IRES	-
pEAY2	pEA1	cmv	SYNZIP3-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP4-biNLS2-LexA	IRES	-
pEAY3	pEA1	cmv	SYNZIP1-VP48_PKIt-mCherry-AsLOV-biNLS2	SYNZIP2-biNLS2-LexA	IRES	-
pEAY4	pEA1	cmv	SYNZIP1-VP48_PKIt-mCherry-AsLOV-biNLS2	SYNZIP2-biNLS2-LexA	IRES	-
pEAY5	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-biNLS2-LexA	P2A	-
pEAY6	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-biNLS2-LexA	T2A	-
pEAY7	pEA1	cmv	SYNZIP1-VP48_PKIt-mCherry-AsLOV-biNLS2	Lyn-tail-NES-PhoCl-SYNZIP2-biNLS2-LexA	IRES	-
pEAY8	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-biNLS2-LexA	B-globin terminator pa/UBC promoter	-
pEAY9	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-SV40 NLS-mClover3-LexA	B-globin terminator pa/UBC promoter	-

Supplementary Table 6.5 continued

pEAY10	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-SV40 NLS-mClover3-LexA	B-globin terminator pa/UBC promoter	PKIt NES_mVenus-Zdk2
pEAY11	pEA1	cmv	SYNZIP1-VP48_PKIt-mCherry-AsLOV-biNLS2	SYNZIP2-SV40 NLS-mClover3-LexA	B-globin terminator pa/UBC promoter	PKIt NES_mVenus-Zdk2
pEAY12	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-SV40 NLS-mClover3-LexA	IRES	PKIt NES_mVenus-Zdk2
pEAY13	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-SV40 NLS-mClover3-LexA	B-globin terminator pa/PGK promoter	PKIt NES_mVenus-Zdk2
pEAY14	pEA1	cmv	SYNZIP1-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP2-SV40 NLS-mClover3-LexA	B-globin terminator pa/UBC promoter	mVenus-Zdk2-CAAX
pEAY15	pEA1	cmv	pMag-VP48_IKBa-mCherry-AsLOV-biNLS10	nMagHigh1-SV40 NLS-LexA	IRES	-
pEAY16	pEA1	cmv	pMag-VP48_PKIt-mCherry-AsLOV-biNLS2	nMagHigh1-SV40 NLS-LexA	IRES	-
pEAY17	pEA1	cmv	SYNZIP2-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP1-biNLS2-LexA	B-globin terminator pa/UBC promoter	-
pEAY18	pEA1	cmv	SYNZIP4-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP1-biNLS2-LexA	B-globin terminator pa/UBC promoter	-
pEAY19	pEA1	cmv	SYNZIP2-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP1-biNLS2-LexA	IRES	-
pEAY20	pEA1	cmv	SYNZIP4-VP48_IKBa-mCherry-AsLOV-biNLS10	SYNZIP1-biNLS2-LexA	IRES	-
pEAY21	pEA1	cmv	SYNZIP3-VP48_IKBa-mCherry-AsLOV	SYNZIP4-SV40 NLS-LexA	IRES	PKIt NES_mVenus-Zdk1
pEAY22	pEA1	cmv	SYNZIP3-VP48_IKBa-mCherry-AsLOV	SYNZIP4-SV40 NLS-LexA	IRES	PKIt NES_mVenus
pEAY23	pEA1	cmv	VP48_IKBa-mCherry-AsLOV	SYNZIP4-SV40 NLS-LexA	IRES	PKIt NES_mVenus-Zdk1

Supplementary Table 6.5 continued

pEAY24	pEA1	cmv	VP48_IKBa-mCherry-AsLOV	SYNZIP4-SV40 NLS-LexA	IRES	PKIt NES_mVenus
pEAY25	pEA1	cmv	SYNZIP3-VP48_IKBa-mCherry-AsLOV	SYNZIP4-SV40 NLS-LexA	IRES	PKIt NES_mVenus-Zdk2
pEAY26	pEA1	cmv	SYNZIP4-VP48_IKBa-mCherry-AsLOV	SYNZIP3-SV40 NLS-LexA	IRES	PKIt NES_mVenus-Zdk2
pEAY27	pEA1	cmv	SYNZIP2-VP48_IKBa-mCherry-AsLOV	SYNZIP1-SV40 NLS-LexA	IRES	PKIt NES_mVenus-Zdk2

Table 6.6 | List of oligonucleotides.

#	**Primer sequence 5'-3'**
1	tcgtgtggctgccggtgaaccacttctggcgcaacagcat
2	gtcggccggcccgccgctttcgttaattaagctggttccgctaccacccagccagtcgccgttgcg
3	gaaagcggcgggccggcc
4	tggttcaccggcagccac
5	aaaagaagaaaaagaagtcaaagacaaagtgtgtaattatgtaggcggccgctcgagcatg
6	ggtggcgcctatttaccaac
7	gttggtaaataggcgccaccatggcgcgtaaggtcgatc
8	ttgtctttgacttctttttcttctttttaccttatagcgttggtggtgggcggcgg
9	atagtaatcaattacggggtcattagttc
10	taataactaatgcatggcggtaatac
11	ccgccatgcattagttattacagacggatcgggagatc
12	accccgtaattgattactatgctggcaagtgtagcggtc
13	atccccgggtaccgagctcgaattccagcttggca
14	gctcggtacccggggatccctttatagcgtctagagtctccgctcggactcg
15	tgatcagacatgtatattggactgtaaaaaaaaacagtggttatatgtacagactagactgtaaaaaaaaacagtggttatatgtacagactagactcgagtccgag
16	tccaatatacatgtctgatcactgtttttttttacagtctagtctgtacatataaccactgtttttttttacagtctagatgcggccgcgaa
17	aattcgcggccgcatctagactgtatataaaaccagtgatcagacatgtatattggactgtatataaaaccagtggttatatgtacagactagactcgagtccgagcg
18	tctagatgcggccgcgaattcggta
19	tgatcagacatgtatattggactgtatataaaaccagtggttatatgtacagactagactgtatataaaaccagtgattcgcggccgcatctagactgt
20	tccaatatacatgtctgatcactggttttatatacagtctagtctgtacatataaccactggttttatatacagtcggtacccggtcacagcttgtctg
21	cgtacgcgctgtcccccgcgttttaaccgccaaggggattactccctagtctccaggcacgtgtcagatatatacatcctgatgagtccgagcggagactcta
22	cgcgggggacagcgcgtacgtgcgtttaagcggtgctagagctgtctacgaccaattgagcggcctgcagaccgggattctccagagtctagtctgtacatata

Supplementary Table 6.6 continued

23	ttaatacaacgaacggtgatgttgtcatagattcggcacatttcccttgtaggtgtgaaatcacttagcttcgcgccgaagtcttatgagtccg agcggagactc
24	tgacaacatcaccgttcgttgtattaatcgatggtggtagcggctgcatattgtcagatgaaggagcgacacccggggaggagtccagag tctagtctgtacata
25	atggcgcgtaaggtcgatctcacctcctgcgatcgcgagccg
26	atcgaccttacgcgccatggtggttagagttgcttcgtgcatagccgattatataccctc
27	atcgaccttacgcgccatggtggcaagagttgcttcgtgcatagccgattatataccctc
28	atcgaccttacgcgccatacactcaagagttgcttcgtgcatagccgattatataccctc
29	tggcgcctatttaccaacagtaccgccctttatagcgtctagagtctccgctcggactc
30	tggcgccccgggatccgagctcgaattccaatggcgcgtaaggtcgatc
31	tggaattcgagctcggatcccggggcgccattatataccctctagagtc
32	gttggtaaataggcgccacctgatagtcagacgctatgtgcatattacgtgaatggcgcgtaaggtcgat
33	agatcgaccttacgcgccatggtggctcacgtaatatgcacatagcgtctgactatcagcctatttaccaacagtaccggaatgccaagct gg
34	tggcgcctcgtgatccgagctcgaaccaccatggcgcgtaaggtcgatctcacc
35	ggtggttcgagctcggatcacgaggcgccattatataccctctagagtctccgc
36	tatatacagtgatcagacatgtatattggactgtatatatatacagtggttatatgtacagactagactgtatatatatacagtggttatatgtac agactagactcgagtccg
37	tccaatatacatgtctgatcactgtatatatatacagtctagtctgtacatataaccactgtatatatatacagtctagatgcggccgcgaattc gg
38	gcgtagctgcgcataagcaaatgacaattaaccactgtgtactcgttataacatctggcagttaaagtcgggagaataggagccgagtcc gagcggagactc
39	ttgcttatgcgcagctacgccatcgcgaggccggtccggcgggcgaagcatataaaagaagctcgtcacatccacatagttgtataaga cttcggcgcgaa
40	gagtctagtctgtacatataacc
41	ggtggcgcctatttaccaacagtaccggaattatataccctctagagtctccgct
42	atggatccccgggtaccgagctcgaattccaatggcgcgtaaggtcgatc
43	tggaattcgagctcggtacccggggatccattatataccctctagagtctccgctc
44	atcgaccttacgcgccatggtggttcgagttgattcgtgcatagccgattatataccctctagagtctccgctcggactcg
45	atcgaccttacgcgccatacactttagagttgcttcgtgcatagccgattatataccctc
46	tgatcagacatgtatattggactgtatgatcatacagtggttatatgtacagactagactgtatgatcatacagtggttatatgtacagactag actcgagtccgag
47	tccaatatacatgtctgatcactgtatgatcatacagtctagtctgtacatataaccactgtatgatcatacagtctagatgcggccgcgaatt cggtacccgg
48	tgatcagacatgtatattggactgtatataaaaccagtggttatatgtacagactagactgtatataaaaccagtgattcgcggccgcatcta gactgt

Supplementary Table 6.6 continued

49	tccaatatacatgtctgatcactggttttatatacagtctagtctgtacatataaccactggttttatatacagtcggtacccggtcacagcttgtctg
50	ggtggcgcctatttaccaacagtaccggaatcgcgcgccctctagagtctccgc
51	gactcttactccctagtcttggatccccgggtaccg
52	aagactagggagtaagagtctccgctcggactc
53	gactcttactccctagtctttccggtactgttggtaaatagg
54	ggaggttcaggtggaagtactagtgatcctacggtacttattgcc
55	ccagtattccagggttcatcagagcatgcatctagaggg
56	acctaaatgctagagctcgctgatcagcctatagtaatcaattacggggtc
57	taataactaatgcatggcggtaatac
58	cagacggatcgggagatctacagac
59	aggctgatcagcgagctc
60	taactcgagcggccgcga
61	ggtggcgaccggtggatc
62	gggatccaccggtcgccaccatggctgctaacgacaactttgctgaagaatttgctgtggctgctggtggtagcggaaccagcgaaagcggcgggccggccgac
63	gtggccatattatcatcgtg
64	ggtggtggcggaggtggtggcggtatggatttgtcacagcta
65	agtcgcggccgctcgagttagtaggcagcttcaggctc
66	cacgatgataatatggccacagcctgcaggatggacgaggccgcaaag
67	accgccaccacctccgccaccaccagccagtcgccgttgcga
68	ttgagcctgaagctgcctactaactcgagcggccgcgact
69	aagttgtcgttagcagccatggtggcgaccggtggatccc
70	gggagggatccatcgatctagtccagctttttcttctttgc
71	caaagaagaaaaagctggactagatcgatggatccctccccccccccctaacgttact
72	atatggccacagcctgcaggatgga
73	cgaggccgcaaagagac
74	gtggcggaggtggtggcggtatgga
75	tttgtcacagctaaca
76	tagatctcccgatccgtctggcgta
77	aattgtaagcgttaatat
78	gttagctgtgacaaatccataccgccaccacctccgccaccaccagccagtcgccgttgcgaat
79	agtcgcggccgctcgagttagtaggcagcttcaggctcaa
80	gctggggctctaggggtatcttaagatacattgatgagtttgga
81	cgcgcgatttctttttccaggtaagcgttttcattttcgatgaacgcagcgtcattctccagagtggtaacttcgttcatggtggcgaccggtg
82	aaacgcttacctggaaaaagaaatcgcgcgtctgcgtaaagaaaaagcggcgctgcgcaaccgtctggcgcacaaaaaaggtggtagcggaaccagcg
83	gttacggttcttcagctcttcaactttgtttttcagctgttcgttacgattcagtttaaccgcaacacggtttttgagttccgcaactttctgcatcctgcaggctgtggccatattatc
84	accgtaacgcttacctcaagaacgaactggcgaccctggagaacgaggttgcgcgtctggaaaacgacgttgcagaaggtggtggcggaggtggtggcggtgacgaggccgcaaagagac
85	gattatgatctagagtcgcggccgctcgagttacagccagtcgccgttgc

Supplementary Table 6.6 continued

86	tttttgtgcaggttctttttcttcagggtttcgttctcattttccagagacgcaacttcgttttcgagctgcgcaaccaggttcatggtggcgaccg gtg
87	gaaaccctgaagaaaaagaacctgcacaaaaaagacctgatcgcgtacctggagaaagaaatcgcgaatctgcgtaagaaaatc gaagaaggtggtagcggaaccagc
88	gatgatttttttccaggttctgttcatcacgttccagctgcaggttgtctttttttcagacgtgcgattttcttacgcagatacgcgttacgcgccatc ctgcaggctgtggccatattatc
89	atgaacagaacctggaaaaaatcatcgcgaacctgcgtgacgaaatcgcgcgtctcgaaaacgaagttgcgtctcacgaacagggt ggtggcggaggtggtgacgaggccgcaaagagac
90	gctggcgacgtggaggagaaccctggacctggcatggcgcgtaacgcg
91	aggtccagggttctcctccacgtcgccagcctgcttcagcaggctgaagttagtagctccgcttccgtccagctttttcttctttgc
92	ctcttgttgcctggccgttaacgctttcatcgagccactgcctcctttatacagttcatccataccatggg
93	atgggatgtattaaatcaaaaagg
94	gagggatccatcgatctag
95	gaaaaagctggactagatcgatggatccctcgggtttattacagggac
96	gtctttccttttgatttaatacatcccatggtggcgaccggtacccg
97	tttcttacgcagatacgcgttacgcgccatggtggcgaccggtacccg
98	gggtttattacagggacagc
99	aaaaagctggactagatcgatggatccctcgtgattaacctcaggtgc
100	ctggatctctgctgtccctgtaataaacccgccgcatgctttcagcaaaaaacc
101	tagatcgatggatccctc
102	aggtccagggttctcctccacgtcgccagcctgcttcagcaggctgaagttagtagctccgcttccgtccagctttttcttctttgctgcggct gcggc
103	gctggcgacgtggaggagaaccctggacctatgttagccttgaaattagcaggtcttgatatcggcgtgagcaagggcgaggagc
104	gaggttaatcacgagggatccatcgatctattttggggcctgggcatcg
105	ttagggggggggagggatccatcgatcta
106	caccggccttattccaagcg
107	cccggtgaacagctcctcgcccttgctcaccatggcgaccggtggatccgcacggcccagcggcac
108	agatctggagccgacacg
109	ggttaattaacccgtgtcggctccagatctactcgaggttaacgaattc
110	tttcttacgcagatacgcgttacgcgccatggtggcgggatgcaggtc
111	gattatgatctagagtcgcggccgctcgagctattttggggcctgggc
112	gctggcgacgtggaggagaaccctggacctatggtgggccgcaacagc
113	gctggcgacgtggaggagaaccctggacctatgggctgcatcaagagcaagcgcaaggacaaggacgagttcggaggcggggg cagcgtgagcaagggcgaggag
114	gctggcgacgtggaggagaaccctggacctgtgagcaagggcgaggag
115	tatgatctagagtcgcggccgctcgagctacataattacacactttgtctttgacttctttttcttctttttaccttatattttggggcctgggcatc
116	tcgtgaactttctgaccatcattccggtccgggatgaaaccggagagtacagatactccatgggattccagtgcgaaaccgaaggtggt ggcggaggtggt
117	aatgtcataaccgccgggagcgtagagcgtatgcatggtggcgaccggtggatcccgggc
118	acgctctacgctcccggcggttatgacattatgggctatctgcggcagattaggaaccgcccaaaccctcaggtcgaactggggcctgt ggacacgtcatgtgctctgattctgtgcgac
119	cccgccgctttcgctggttccgctaccaccttccgtttcgcactggaaac
120	gacgtgtccacaggccccagttcgacctgagggtttgggcggttcccaatctgatccagatagcccataatgtcataaccgccgggag cgtagagcgtatgcatcctgcaggctgtggcc
121	tcaggtcgaactggggcctgtggacacgtcatgtgctctgattctgtgcgac
122	ggaccggaatgatggtcagaaagttcacgaaccgctggccgttcttcttaaagttgaccacctcaacctgcacctcggcgttcctatcaa tcgctttcctgatcgtattgatcgtgttgg
123	gaggttaatcacgagggatccatcgatctacataattacacactttgtctttgac
124	aacctgcacaaaaaagacctgatcgcgtacctggagaaagaaatcgcgaatctgcgtaagaaaatcgaagaaggtggtggcgga cccaagaagaagcgcaaggtggccaccatgaaagcg
125	gatttttccaggttctgttcatcacgttccagctgcaggttgtctttttcagacgtgcgattttcttacgcagatacgcgttacgcgccatggtg gcgaccggtggatcccgggcc

Supplementary Table 6.6 continued

126	gaacgtgatgaacagaacctggaaaaaatcatcgcgaacctgcgtgacgaaatcgcgcgtctcgaaaacgaagttgcgtctcac gaacagggtggtagcggaaccagc
127	gatcaggtctttttgtgcaggttctttttcttcagggtttcgttctcattttccagagacgcaacttcgttttcgagctgcgcaaccaggttcat ggtggcgaccggtacccggggatc
128	aaaagaaatcgcgcgtctgcgtaaagaaaaagcggcgctgcgcaaccgtctggcgcacaaaaaaaggtggtggcggacccaa gaagaagcgcaaggtggccaccatgaaagcgttaacgg
129	ttacggttcttcagctcttcaactttgtttttcagctgttcgttacgattcagtttaaccgcaacacggtttttgagttccgcaactttctgcatggt ggcgaccggtggatcccgggcc
130	caaagttgaagagctgaagaaccgtaacgcttacctcaagaacgaactggcgaccctggagaacgaggttgcgcgtctggaaaa cgacgttgcagaaggtggtagcggaaccagc
131	ttttctttacgcagacgcgcgatttctttttccaggtaagcgttttcattttcgatgaacgcagcgtcattctccagagtggtaacttcgttcat ggtggcgaccggtacccggggatc
132	gtctgctaacatgcggtgacgtcgaggagaatcctggcccaatggcgcgtaacgcgtatc
133	gtcaccgcatgttagcagacttcctctgccctctccactgccgtccagctttttcttctttgctgcggc
134	aatatggccacagcctgcaggatggcgcgtaacgcg
135	agtggcggaagcggaggcagcggcggaagcgtgagcaagggcgaggagg
136	ccacgtcgccagcctgcttcagcaggctgaagttagtagctccgcttcccagctccttggcggcctcgtcaatattttctgcagttttctta atcagc
137	tcagcctgctgaagcaggctggcgacgtggaggagaaccctggacctatgaacagcaacgagctggccctgaagctggccggc ctggacatcaacaagggcgtgagcaagggcgaggagc
138	ggatccacttccagaaccttgtacagctcgtccatgccgagagtg
139	ggttctggaagtggatccggcagtggagagggcagaggaagtctgc
140	ggatccacttccagaacc
141	cacgatgataatatggccacagcctgcaggatgaaagcgttaacggccaggcaacaagag
142	atcaagacctgctaatttcaaggctaacatggtggcgaccggtggatcccgggc
143	gagtctagtctgtacatataacc
144	gtctggcgcacaaaaaaggtggtggcggacccaagaagaagcgcaaggtg
145	acctttttgtgcgccagacggttgcgcagcgccgctttttctttacgcag
146	accggtcgccaccatggaaagcggcgggccggccgacg
147	tttccatggtggcgaccggtggatcccgggcccgcggtacc
148	tggcgacgtggaggagaaccctggacctatgggcccaaaaaagaaaagaaaagttggctacccctacgacgtgcccgactacg ccatcgaaggccgccatatgctagccgttaaaatggc
149	cgatgataatatggccacagcctgcaggatgggcccaaaaaagaaaagaaaagttggctacccctacgacgtgcccgactacgc catcgaaggccgccatatgctagccgttaaaatggc
150	aaaagaaatcgcgcgtctgcgtaaagaaaaagcggcgctgcgcaaccgtctgg
151	gagccaggaaacctttagcgatctgtggaaactgctgccgactagtggcggaagcggagg
152	ctaaaggtttcctggctcagcggcggttccacgctgctggttccgctaccacc
153	atattagctttccgggcctggatgaagatccgctgggcccggataacattaactggagccagtttattccggaactgcagactagtgg cggaagcggagg
154	caggcccggaaagctaatatccagcagaattttgctcaggctatcgttcatggtatccagcaccaggccgctggttccgctaccacc
155	tagcagacttcctctgccctctccactgcctttttggggcctgggcatcg
156	cgatgcccttgacgactttgaccttgatatgctgcccgggagtggcggaagcggaggcag
157	caaagtcgtcaagggcatcggctggcccgccgctttcgctggttccgctaccaccc
158	agcggcgggagtggcggttccggcggcagccccagcacccggatccagc
159	atggccacagcctgcaggatgaacgaagttaccac
160	tggccacagcctgcaggatgaacctggttgcgcag
161	ggagaatcctggcccaatgaacctggttgcgcagc
162	aggttcattgggccaggattctcctcgacgtcacc
163	tcgttcattgggccaggattctcctcgacgtcacc
164	gaatagaccgagatagggttgagtgttg
165	acccccctagagccccagctg
166	cggaaagaaccagctggggctctaggggggtttttgatttataagggattttgccgatttc
167	gaacaacactcaaccctatctcggtctattcatagtaatcaattacggggtcattagttc

6.3 Supplementary Sequences

Gblocks:

3xFox1Mc-TAD1

GGCGACTGGCTGggtggtagcggaaccagcGGCTTGGTCCTTGATACAATGAATGATAGCC
TGAGCAAAATTCTGCTGGATATATCCTTTCCCGGTCTGGATGAAGATCCCCTCGGTC
CTGACAATATTAACTGGTCCCAATTCATCCCAGAACTCCAGGGCCTCGTCCTTGACA
CAATGAATGACTCATTGAGTAAAATCCTTCTGGACATCTCCTTTCCTGGGCTGGATG
AGGACCCGCTCGGACCAGATAACATAAACTGGAGCCAATTTATACCGGAACTTCAA
GGCCTTGTTCTGGATACCATGAACGACTCCTTGAGTAAGATCTTGTTGGATATTAGT
TTCCCAGGCCTCGACGAAGACCCATTGGGACCTGACAATATTAATTGGTCACAATTC
ATACCTGAATTGCAGAGCGGCGGGAGTGGCGGTTCCGGCGGCAGC

2x RandSeq_2xCTCF

AACCAGTGGTTATATGTACAGACTAGACTCCCGCGAGGTGGCAGGGTGTCGCTCCT
TCATCTGACAATATGCAGCCGCTACCACCATCGATTAATACAACGAACGGTGATGTT
GTCATAGATTCGGCACATTTCCCTTGTAGGTGTGAAATCACTTAGCTTCGCGCCGAA
GTCTTATACAACTATGTGGATGTGACGAGCTTCTTTTATATGCTTCGCCCGCCGGAC
CGGCCTCGCGATGGCGTAGCTGCGCATAAGCAAATGACAATTAACCACTGTGTACT
CGTTATAACATCTGGCAGTTAAAGTCGCCGCGTGGAGGCAGGAGTCCGAGCGGAG
ACTCTAGAGGGTATATAATGGATCCCCGGGTACCGAGCTCGAATTCCAGCTTGGCA
TTCCGGTACTGTTGGTAAATAGGCGCCACCATGGCGCGTAAGGTCGATCTCACCTC
CTGC

Promoter p1:

CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
ctcgagtccgagcggagactctagacgcTATAAAagggatccccgggtaccgagctcgaattccagcttggcattccggtac
tgttggtaaataggcgccaccATGGCGCGTAAGGTCGATCTCACCTCCTGCGATCGCGAGCCGA
TCCACATCCCCGGCAGCATTCAGCCGTGCGGCTGCCTGCTAGCCTGCGACGCGCAGG
CGGTGCGGATCACGCGCATTACGGAAAATGCCGGCGCGTTCTTTGGACGCGAAACTCC
GCGGGTCGGTGAGCTACTCGCCGATTACTTCGGCGAGACCGAAGCCCATGCGCTGCG
CAACGCACTGGCGCAGTCCTCCGATCCAAAGCGACCGGCGCTGATCTTCGGTTGGCG
CGACGGCCTGACCGGCCGCACCTTCGACATCTCACTGCATCGCCATGACGGTACATCG
ATCATCGAGTTCGAGCCTGCGGCGGCCGAACAGGCCGACAATCCGCTGCGGCTGACG
CGGCAGATCATCGCGCGCACCAAAGAACTGAAGTCGCTCGAAGAGATGGCCGCACGG
GTGCCGCGCTATCTGCAGGCGATGCTCGGCTATCACCGCGTGATGTTGTACCGCTTCG
CGGACGACGGCTCCGGGATGGTGATCGGCGAGGCGAAGCGCAGCGACCTCGAGAGC
TTTCTCGGTCAGCACTTTCCGGCGTCGCTGGTCCCGCAGCAGGCGCGGCTACTGTACT
TGAAGAACGCGATCCGCGTGGTCTCGGATTCGCGCGGCATCAGCAGCCGGATCGTGC
CCGAGCACGACGCCTCCGGCGCCGCGCTCGATCTGTCGTTCGCGCACCTGCGCAGCA
TCTCGCCCTGCCATCTCGAATTTCTGCGGAACATGGGCGTCAGCGCCTCGATGTCGCT
GTCGATCATCATTGACGGCACGCTATGGGGATTGATCATCTGTCATCATTACGAGCCGC
GTGCCGTGCCGATGGCGCAGCGCGTCGCGGCCGAAATGTTCGCCGACTTCTTATCGC
TGCACTTCACCGCCGCCCACCACCAACGCTATAAGGGTAAAAAGAAGAAAAAGAAGTC
AAAGACAAAGTGTGTAATTATGtag

Promoter p2:

CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
ctcgagtccgagcggagactctagagggTATATAatggatccccgggtaccgagctcgaattccagcttggcattccggtact
gttggtaaataggcgccaccATGGCGCGTAAGGTCGATCTCACCTCCTGCGAT...

Promoter p3:

CTGTAAAAAAAAACAGtggttatatgtacagactagaCTGTAAAAAAAAACAGtgatcagacatgtatattgga
CTGTAAAAAAAAACAGtggttatatgtacagactagaCTGTAAAAAAAAACAGtggttatatgtacagactaga
ctcgagtccgagcggagactctagacgcTATAAAagggatccccgggtaccgagctcgaattccagcttggcattccggtac
tgttggtaaataggcgccaccATGGCGCGTAAGGTCGATCTCACCTCCTGCGAT...

Promoter p4:

CTGTAAAAAAAAACAGtggttatatgtacagactagaCTGTAAAAAAAAACAGtgatcagacatgtatattgga
CTGTAAAAAAAAACAGtggttatatgtacagactagaCTGTAAAAAAAAACAGtggttatatgtacagactaga
ctcgagtccgagcggagactctagagggTATATAatggatccccgggtaccgagctcgaattccagcttggcattccggtact
gttggtaaataggcgccaccATGGCGCGTAAGGTCGATCTCACCTCCTGCGAT...

Promoter p7:

CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
ctcTGGAGAATCCCGGTCTGCAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCT
TAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGT
CTCCAGGCACGTGTCAGATATATACATCCTGATgagtccgagcggagactctagagggTATATAatg
gatccccgggtaccgagctcgaattccagcttggcattccggtactgttggtaaataggcgccaccATGGCGCGTAAGG
TCGATCTCACCTCCTGCGAT...

Promoter p8:

CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
ctcTGGACTCCTCCCCGGGTGTCGCTCCTTCATCTGACAATATGCAGCCGCTACCACCAT
CGATTAATACAACGAACGGTGATGTTGTCATAGATTCGGCACATTTCCCTTGTAGGTGT
GAAATCACTTAGCTTCGCGCCGAAGTCTTATgagtccgagcggagactctagagggTATATAatggat
ccccgggtaccgagctcgaattccagcttggcattccggtactgttggtaaataggcgccaccATGGCGCGTAAGGTC
GATCTCACCTCCTGCGAT...

Promoter p9:

CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
ctcTGGACTCCTCCCCGGGTGTCGCTCCTTCATCTGACAATATGCAGCCGCTACCACCAT
CGATTAATACAACGAACGGTGATGTTGTCATAGATTCGGCACATTTCCCTTGTAGGTGT
GAAATCACTTAGCTTCGCGCCGAAGTCTTATgagtccgagcggagactctagacgcTATAAAagggat
ccccgggtaccgagctcgaattccagcttggcattccggtactgttggtaaataggcgccaccATGGCGCGTAAGGTC
GATCTCACCTCCTGCGAT...

Promoter p22:

CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
CTGTATATAAAACCAGtggttatatgtacagactagaCTGTATATAAAACCAGtggttatatgtacagactaga
ctcgagtccgagcggagactcTTACTCCCTAGTCTtggatccccgggtaccgagctcgaattccagcttggcattccggt
actgttggtaaataggcgccaccATGGCGCGTAAGGTCGATCTCACCTCCTGCGAT...

Purple – Responsive element

Blue – TATA-box

Brown – iRFP670

Cyan – CAAX motif

Red – 601 nucleosome positioning sequence

Pink – Randomly generated sequence

Orange, Ascent 2 – 2x Initiator sequence

SynTF:

ATGAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTTGATCTCATCCGTGATCACATCAG
CCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCGCAGCGTTTGGGGTTCCGTTC
CCCAAACGCGGCTGAAGAACATCTGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATT
GTTTCCGGCGCATCACGCGGGATTCGTCTGTTGCAGGAAGAGGAAGAAGGGTTGCCG
CTGGTAGGTCGTGTGGCTGCCGGTGAACCACTTCTGGCGCAACAGCATATTGAAGGTC
ATTATCAGGTCGATCCTTCCTTATTCAAGCCGAATGCTGATTTCCTGCTGCGCGTCAGC
GGGATGTCGATGAAAGATATCGGCATTATGGATGGTGACTTGCTGGCAGTGCATAAAA
CTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGTATTGATGACGAAGTTACCGTT
AAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTGCCAGAAAATAGCGAGTTTAA
ACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATTGAAGGGCTGGCGGTTGGG
GTTATTCGCAACGGCGACTGGCTGggtggtagcggaaccagcttaattaacGAAAGCGGCGGGCC
GGCCGACGCCCTTGACGATTTTGACTTAGACATGCTCCCAGCCGATGCCCTTGACGAC
TTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGACCTTGACATGCTCCC
CGGGactagtggcggaagcggaggcagcggcggaagcCCCAGCACCCGGATCCAGCAGCAGCTGG
GCCAGCTGACCCTGGAGAACCTGCAGGTGAGCAAGGGCGAGGAGGATAACATGGCCA
TCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGA
GTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCA
AGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTC
AGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTT
GAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGG
CGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAA
GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGAC
CATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGG
GCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCA
AGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACAT
CAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGC
GCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACaagggtggatctggaggttca
ggtggaagtTTGGCTACTACACTTGAACGTATTGAGAAGAACTTTGTCATTACTGACCCAAG
ATTGCCAGATAATCCCATTATATTCGCGTCCGATAGTTTCTTGCAGTTGACAGAATATAG
CCGTGAAGAAATTTTGGGAAGAAACTGCAGGTTTCTACAAGGTCCTGAAACTGATCGCG
CGACAGTGAGAAAAATTAGAGATGCCATAGATAACCAAACAGAGGTCACTGTTCAGCTG
ATTAATTATACAAAGAGTGGTAAAAAGTTCTGGAACCTCTTTCACTTGCAGCCTATGCGA
GATCAGAAGGGAGATGTCCAGTACTTTATTGGGGTTCAGTTGGATGGAACTGAGCATGT
CCGAGATGCTGCCGAGAGAGAGGGAGTCATGCTGATTAAGAAAACTGCAGAAAATATT
GACGAGGCCGCAAAGAGACTGCCCGACGCCAACCTGGCCGCAGCAGCCGCAGCCGC
AGCAAAGAAGAAAAAGCTGGACtag

Blue-Grey – LexA Dark-yellow – VP48 Orange - IKBα

Red – mCherry Light blue – AsLOV2 Gold – biNLS10

6.4 List of Figures

FIGURE 1.1 | EUKARYOTIC CORE PROMOTER ELEMENTS. 5
FIGURE 1.2 | DEPICTION OF TRANSCRIPTION PRE-INITIATION COMPLEX STABILIZED BY VP16. 7
FIGURE 1. 3 | ILLUSTRATION OF TRANSCRIPTION FACTOR DYNAMICS. 13
FIGURE 1.4 | LIGHT-INDUCIBLE NUCLEAR LOCALIZATION SIGNAL (LINUS). 18
FIGURE 1.5 | STRUCTURE OF YES-ASSOCIATED PROTEIN. 19
FIGURE 1.6 | STRUCTURAL DETAILS OF TEA DOMAIN PROTEIN 1. 21
FIGURE 2.1.1 | DESIGN AND CHARACTERIZATION OF THE SYNTHETIC TF. 26
FIGURE 2.1.2 | SEQUESTERING TF1 FROM THE NUCLEUS USING LOVTRAP. 27
FIGURE 2.1.3 | CHARACTERIZATION OF THE SYNTHETIC TRANSCRIPTION FACTOR (SYNTF). 28
FIGURE 2.1.4 | DESIGN AND CONSTRUCTION OF REPORTER LIBRARY AND QUANTIFICATION OF REPORTER EXPRESSION. 29
FIGURE 2.1.5 | REPORTER EXPRESSION UNDER DIFFERENT PLASMID ARRANGEMENTS. 31
FIGURE 2.1.6 | EXPRESSION OF REPORTER ON SYNPLASMID1 SHOWS A WEAK DEPENDENCE ON SYNTF LEVELS. 32
FIGURE 2.1.7 | REPORTER EXPRESSION IS DIFFERENTIALLY INDUCED BY DIFFERENT TRANSACTIVATION DOMAINS. 33
FIGURE 2.1.8 | EXPERIMENTAL SETUP FOR STUDYING TRANSCRIPTION FACTOR DYNAMICS. 35
FIGURE 2.1.9 | A PROMOTER WITH HIGH AFFINITY RES AND TATA-BOX CANNOT DISTINGUISH BETWEEN SUSTAINED AND PULSATILE TF DYNAMICS. 37
FIGURE 2.1.10 | PROMOTER WITH LOW AFFINITY TATA-BOX AND HIGH AFFINITY RES DOES NOT DISTINGUISH DIFFERENT SYNTF DYNAMICS. 39
FIGURE 2.1.11 | PROMOTER WITH HIGH AFFINITY TATA-BOX AND LOW AFFINITY RESPONSIVE ELEMENT EFFICIENTLY INTEGRATES NUCLEAR SYNTF INTO REPORTER GENE EXPRESSION. 40
FIGURE 2.1.12 | PROMOTER WITH A LOW AFFINITY RES AND TATA-BOX DISPLAYS A DELAYED REFRACTORY RESPONSE UNDER SUSTAINED SYNTF DYNAMICS. 42
FIGURE 2.1.13 | PROMOTER WITH A LOW AFFINITY RES AND TATA-BOX INDUCES LOWER REPORTER EXPRESSION UNDER PULSATILE SYNTF DYNAMICS. 43
FIGURE 2.1.14 | LOW DNA LOOPING EFFICIENCY HELPS TO DISTINGUISH DIFFERENT SYNTF DYNAMICS. 47
FIGURE 2.1.15 | PROMOTERS WITH LOWER PROBABILITY OF INTERACTION BETWEEN TF-BOUND RE AND TBP-BOUND TATA-BOX SHOW DIFFERENTIAL EXPRESSION IN RESPONSE TO DIFFERENT SYNTF DYNAMICS. 48
FIGURE 2.1.16 | REPORTER WITH LOWER AFFINITY TATA-BOX AND TRANSLATION RATE EFFECTIVELY DISTINGUISHES DIFFERENT SYNTF DYNAMICS. 50
FIGURE 2.2.1 | CONSTRUCTION OF SYNTEAD/SYNYAP TF AND THE CHARACTERIZATION OF SYNYAP NUCLEAR ACCUMULATION. 52
FIGURE 2.2.2 | NUCLEAR ACCUMULATION OF THE SYNYAP IS ENHANCED BY STRONGER INTERACTION WITH SYNTEAD AND HIGHER NUCLEAR SYNTEAD LEVELS. 54

FIGURE 2.2.3 | HIGHER INTERACTION AFFINITY AND LOWER SYNTEAD LEVELS YIELD HIGHER REPORTER PROTEIN EXPRESSION. 56
FIGURE 2.2.4 | REPORTER EXPRESSION INDUCED BY SYNTF IS SIGNIFICANTLY HIGHER THAN INDUCTION BY SYNTEAD/SYNYAP TFS. 57
FIGURE 2.2.5 | NUCLEOCYTOPLASMIC TRANSLOCATION OF SYNYAP BASED ON INTERACTION WITH PROTEINS LOCALIZED TO DIFFERENT SUBCELLULAR COMPARTMENTS. 59
FIGURE 2.2.6 | HIGHER ZIPA AND ZIPB AFFINITY FACILITATES NUCLEAR ACCUMULATION OF SYNYAP IN THE ABSENCE OF ANY NLS. 61
FIGURE 2.2.7 | REPORTER EXPRESSION INDUCED BY SYNTEAD/SYNYAP WAS SIGNIFICANTLY LOWER WITHOUT NUCLEAR LOCALIZATION SIGNAL. 62
FIGURE 3.1 | THEORETICAL PREDICTION OF A PROMOTER THAT INDUCES HIGHER REPORTER EXPRESSION UNDER PULSATILE TF DYNAMICS. 70
FIGURE 6.1 | PROMOTER WITH LOW AFFINITY TATA-BOX AND HIGH AFFINITY RE INDUCES HIGH REPORTER EXPRESSION. 99
FIGURE 6.2 | PROMOTER WITH LOW AFFINITY RE AND HIGH AFFINITY TATA-BOX DISPLAYS NO REFRACTORY RESPONSE. 100
FIGURE 6.3 | REPORTER EXPRESSION INDUCED BY LOW AMPLITUDE SUSTAINED AND HIGH AMPLITUDE PULSATILE SYNTF DYNAMICS. 101
FIGURE 6.4 | REPORTER PROTEIN EXPRESSION INDUCED BY DIFFERENT PROMOTER DESIGNS UNDER SUSTAINED SYNTF DYNAMICS. 102
FIGURE 6.5 | REPORTER EXPRESSION UNDER DIFFERENT TF SETUPS AND DIFFERENT INTERACTION AFFINITIES. 103
FIGURE 6.6 | REPORTER EXPRESSION UNDER DIFFERENT LEVELS OF SYNTEAD.103

6.5 List of Tables

TABLE 1.1: SEQUENCES OF LEXA RESPONSIVE ELEMENTS. 3
TABLE 5.1 | DESCRIPTION OF MODEL PARAMETERS AND HOW THEY WERE VARIED. 95
TABLE 6.1 | TRANSCRIPTION FACTOR CONSTRUCTS. 104
TABLE 6.2 | LIST AND COMPOSITION OF SYNTHETIC PROMOTERS. 104
TABLE 6.3 | TRANSCRIPTION FACTOR DYNAMICS PLASMID CONSTRUCTS. 106
TABLE 6.4 | HETERODIMER PAIRS. 108
TABLE 6.5 | SYNTEAD/SYNYAP PLASMIDS. 108
TABLE 6.6 | LIST OF OLIGONUCLEOTIDES. 110

6.6 Abbreviations

3’ UTR	Three prime untranslated region
5’ UTR	Five prime untranslated region
40S	Eukaryotic small ribosomal subunit
60S	Eukaryotic large ribosomal subunit
80S	Eukaryotic ribosome
λ Repressor	Lambda repressor
AP-1	Activator protein 1
AP-2	Activator protein 2
ASPP	Apoptosis stimulating protein of p53
BGH	Bovine growth hormone
bHLH	Basic helix-loop-helix
BREd	Downstream TFIIB recognition element
BREu	Upstream TFIIB recognition element
BRN3A	Brain-specific Homeobox/POU domain protein 3A
BRN3B	Brain-specific Homeobox/POU domain protein 3B
C-terminal	Carboxyl-terminal
CTD	Carboxyl-terminal hepta-peptide repeat domain
DBD	DNA-binding domain
DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
DSIF	DRB-sensitivity-inducing factor
EGF	Epidermal growth factor
eIF	Eukaryotic initiator factor
ERK	Extracellular signal-regulated kinase
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
HAT	Histone acetyltransferase
HTH	Helix-turn-helix

Hzf	Hematopoietic zinc finger protein
iASPP	Inhibitor of ASPP
IFN-β	Interferon-β
IkBα	NF-κB inhibitor alpha
INR	Initiator sequence
IRES	Internal ribosome entry site
JNK	c-Jun N-terminal kinase
KZ	Kozak sequence
LATS1/2	Large tumour suppressor kinase
LPS	Lipopolysaccharide
MCP	MS2 coat protein
MBS	MS2 binding site
Mdm2	Mouse double minute 2 homolog
MOB1	MOB kinase activator 1
mRNA	Messenger RNA
MST1/2	Mammalian hippo pathway core kinases STE20-like protein kinase 1/2
N-terminal	Amino-terminal
NELF	Negative elongation factor
NF-κB	Nuclear factor-κB
NFAT	Nuclear factor of activated T cells
NGF	Nerve growth factor
P-TEFb	Positive transcription factor b
PKA	Protein kinase A
PBS	PP7 RNA hairpin or binding site
PCP	PP7 coat protein
PIC	Pre-initiation complex
PGK	Phosphoglycerate kinase
PKIt	truncated cAMP-dependent protein kinase inhibitor alpha
POI	Protein of interest
RE	Responsive elements

RG4	G-quadruplex structures
RNA	Ribonucleic acid
RNAP II	RNA polymerase II
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAV1	Salvador family WW domain containing protein 1
SynPlasmid	synthetic plasmid containing both TF and reporter
SynPromoter	Synthetic promoter
SynTF	Synthetic transcription factor
Sp1	Sp1 transcription factor
TAD	Transactivation domain
TAF	TATA binding protein associated factors
TAZ	Transcriptional coactivator with PDZ-binding domain
TB	TATA-box
TBP	TATA binding protein
TCF4	Transcription factor 4
TEAD	TEA domain transcription factor
TF	Transcription factor
TFIIA	Transcription factor II A
TFIIB	Transcription factor II B
TFIID	Transcription factor II D
TFIIE	Transcription factor II E
TFIIF	Transcription factor II F
TFIIH	Transcription factor II H
TLR4	Toll-like receptor 4
TNFα	Tumour necrosis factor-a
TSS	Transcription start site
uORF	Upstream open reading frame
UBC	Ubiquitin C
VGLL	Vestigial-like protein
WCC	White Collar Complex
YAP	Yes-associated protein