Using the Tandem Fluorescent Timer as a Reporter of Dynamic Gene Regulation

Danny Salem

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the M.Sc. in Cellular and Molecular Medicine,

Specialization in Bioinformatics

Department of Cellular and Molecular Medicine

Faculty of Medicine

University of Ottawa

Ottawa, Canada

Abstract

I propose the use of the tandem fluorescent timer protein as a reporter of dynamic gene regulation. The tandem fluorescent timer is a fusion of two fluorophores with different maturation kinetics whose fluorescence ratio is a reporter of protein age. Traditional approaches to live single-cell tracking of dynamic gene expression involve the use of destabilized fluorescent reporters. The reduced stability of these reporters improve performance but also result in reduced signal and an increased signal to noise ratio. I first develop a platform to test reporter performance by designing and implementing an inducible synthetic network orthogonally in *S. cerevisiae* cells and by developing a microfluidics-enabled live cell-tracking pipeline. To test the performance of different reporters, I develop an algorithm to decode the underlying regulatory dynamic signal of a fluorescence profile. I then simulate the fluorescence output of my platform under dynamic regulatory signaling to demonstrate the potential reporter performance of a stable timer protein. Finally, I conduct live cell-tracking experiments of yeast cells expressing the timer under a periodic signal to test *in vivo* performance of the tandem fluorescent timer. I demonstrate that compared to a traditional stable fluorescent reporter, the tandem fluorescent timer is more accurate when tracking faster periodic signals and it is more robust to global fluctuations.

Acknowledgments

I would like to thank Dr. Mads Kaern for all the guidance and support I have received during my Master's degree. As a mentor you provided me with many opportunities to improve myself.

I am grateful that the Kaern lab was always populated with smart and interesting people with whom I would always have a great conversation. I want to thank my friends and labmates: Brendan Camellato, Ian Roney, Hilary Phenix, Kateland Simmons. I also want to thank my friends from the Baetz, Downey and Blais labs.

I would like to thank Mila Tepliakova for always being available to help.

I thank Dr. Downey, Dr. Perkins and Dr. Longtin for the discussion and advice that helped me complete my research.

Finally, I thank my family and friends that didn't necessarily always understand my work but supported me anyways. You made it all possible.

Contents

1 Introduction **1**

1.1 Gene Expression 1

1.1.1 Gene expression dynamics 1

1.1.2 Measurement of gene expression 2

1.2 The Development and Use of Fluorescent Proteins to Study Gene Expression 3

1.2.1 The Discovery of GFP 3

1.2.2 FP Timers 7

1.3 Hypotheses and Specific Aims 10

2 Materials **13**

2.1 Strain Assembly 13

2.2 Quantitative Fluorescent Time Lapse Microscopy (Image Acquisition) . 14

2.3 Image Analysis (Segmentation, Lineage Tracking) 15

2.4 Data Filtering 15

2.5 Model Formulation and Fitting 16

2.5.1 Full Model 16

2.5.2 Derivation of Simplified Model 18

2.5.3 Simplified Model 21

2.5.4 Fitting 22

3 Results **23**

3.1 Specific Aim 1: Implementation and validation of a system to measure FP expression in live yeast cells . . . 23

3.1.1 Development and implementation of an inducible synthetic construct to compare fluorescent protein gene expression tracking performance . . . 24

3.1.2 Development of an image acquisition and analysis pipeline to produce single-cell gene expression profiles . . . 27

3.1.3 Characterization and testing of the TFT testing platform . . . 32

3.2 Specific Aim 2: Modeling and simulating regulated FP and TFT expression 33

3.2.1 Model formulation . . . 33

3.2.2 Model characterization and parameter fitting . . . 33

3.3 Specific Aim 3: Development of an algorithm to infer regulatory signal dynamics from simulated data . . . 38

3.3.1 Algorithm development . . . 38

3.3.2 Algorithm testing using simulated data . . . 42

3.4 Specific Aim 4: Reconstructing regulatory signals from FP expression . 47

3.4.1 Signal reconstruction using single-cell FP and TFT fluorescence data . . . 47

3.4.2 Signal inferring using population-wide FP and TFT fluorescence data . . . 52

4 Discussion **56**

4.1 Comparison of the tandem fluorescent timer to traditional fluorescent proteins . . . 56

4.2 Delay in fluorescence response to dynamic regulatory signalling . . . 58

4.3 Alternative construct designs to improve response dynamics . . . 59

4.4 Complications introduced by errors in image analysis . . . 60

4.5 Conclusion . . . 62

List of Tables

2.1 Parameters present in full model. 19

2.2 Parameter values from the simplified model that were used for simulations. 21

3.1 Simulation Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method . . . 43

3.2 Shifted Simulation Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method . 43

3.3 Single-Cell Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method . . . 48

3.4 Shifted Single-Cell Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method . 48

3.5 Population-wide Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method . 54

3.6 Shifted Population-wide Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method 55

List of Figures

2.1 Visual description of the simplified maturation kinetics of the tandem fluorescent timer and the expression dynamics of the synthetic gene network. 17

3.1 Visual representation of the synthetic gene network assembled for in vivo testing of the tandem fluorescent timer 26
3.2 Data acquisition and analysis pipeline 29
3.3 Measurements of the fluorescence change when an repression signal is removed. 30
3.4 Measurements of the fluorescence change when an activating signal is removed. 31
3.5 Simulations of timer expression kinetics under single-step regimes. . . . 34
3.6 Tandem Fluorescent Network Model Parameterization. 35
3.7 Tandem Fluorescent Timer expression simulation with a periodic forcing signal using fit model. 37
3.8 Flowchart describing the algorithm designed for periodic signal reconstruction from experimental time-lapse fluorescence microscopy data. . 39
3.9 The accuracy calculation used to score the performance of the algorithm. 41
3.10 Periodic signal reconstruction using the developed algorithm on simulations of periodically regulated TFT expression 45
3.11 Periodic signal reconstruction using developed algorithm on time-shifted simulations of periodically regulated TFT expression 46

3.12 Representative plots of cell profiles tracked from periodic forcing experiments 49
3.13 Periodic signal reconstruction using the developed algorithm on fluorescence data of periodically regulated TFT expression 50
3.14 Representative plots of cell profiles obtained using the improved segmentation data 51
3.15 Periodic signal reconstruction using the developed algorithm on population-wide data of periodically regulated tandem fluorescent timer expression 53

Chapter 1

Introduction

1.1 Gene Expression

1.1.1 Gene expression dynamics

Gene expression is an inherently stochastic process, thus genetically identical cells are heterogeneous. Fluctuations in transcription and translation caused by the limited machinery present in a cell is an unavoidable source of noise that propagates throughout the genome. It has been shown that genetically identical cells can exhibit different distinct phenotypes, a phenomena known as non-genetic plasticity [1]. This phenomena is prevalent in the field of cancer therapeutics where it has also been observed that tumour cell populations become resistant to cancer-killing drugs [2, 3]. The current understanding of the origin of drug-resistance in cancer cells is that it is caused by the Darwinian selection of mutant cancer cells, whose increased fitness enable their survival and proliferation in the remaining population [4]. However, recent work has demonstrated that this resistance is obtained through non-genetic state transitions [3]. To uncover the cause of non-genetic state transitions of cancer cells to more resistant less differentiated states, population level analysis is not enough. It is necessary to track individual cells and to monitor their gene expression profiles over time.

A common topic of study in molecular biology is the discovery and characterization

of genes that are expressed dynamically [5]. This behaviour is prevalent in diverse cell types, from microbial cells to mammalian cells, and has been observed to be critical to their survival [6]. Well studied systems that involve dynamic gene expression include: stress response systems [6], cell cycle [7], circadian rhythm [8, 9], and development pathways [10]. Dynamic expression is observed in varying complexity, ranging from sustained state transitions and transient bursts of expression to periodic and stochastic pulsatile expression.

The variability of gene expression has been observed for decades but efforts to classify and quantify noise began with work done by Elowitz in 2002 [11]. Gene expression involves the random binding of molecules, which are often present in limited quantities. Intrinsic noise is described as the variation introduced by the stochasticity inherent to biochemical processes. Meanwhile, extrinsic noise is characterized by fluctuations in the quantity and state of molecules that are involved in regulating expression. By using a novel dual reporter system, Elowitz decomposed the variability in gene expression observed between identically regulated reporter proteins into intrinsic and extrinsic noise. It is important to study the temporal variation of gene expression between cells because the variation in genetically identical cells can exhibit phenotypic heterogeneity. Since population-level measurement techniques are not good candidates for investigation of this type of cellular behaviour, single-cell time-lapse measurements of protein concentration are necessary. A reporter with the ability to track dynamic gene expression is favourable for this field of study. Fluorescent proteins, paired with time-lapse microscopy, are a powerful tool for single-cell analysis.

1.1.2 Measurement of gene expression

The ability to measure gene expression accurately has been a constant topic of study. The first methods developed include Northern [12] and Western Blotting [13]. These blotting methods enabled the detection of transcripts and expressed proteins, respectively. However the consequences associated with these methods are not conducive to

single-cell and live-cell analysis. Blotting methods require the extraction of material from a cell population; RNA in the case of the Northern blot, protein in the case of the Western blot. Furthermore, while their use as an indicator of the presence of target molecules in a cell population is still relied on today, the quantification of the bands is prone to error. A common issue in this respect, is the inability to differentiate the signal intensity of very bright bands due to saturation [14].

More recently developed methods for measuring gene expression are more accurate and are much higher throughput. RT-qPCR is a very useful technique that can be used to quantify the gene expression of many mRNA transcripts using fluorescent probes [15]. While the amplification inherent to the PCR reaction enables the quantification of transcripts with very low expression, it would amplify any contaminating nucleic acids as well. RNA-Seq is another technique that leverages recent technological advances to study gene expression. Using next-generation sequencing techniques, RNA-seq can quantify all of the transcripts present in a cell [16]. There are also modern techniques to measure gene expression via protein count, namely quantitative mass spectrometry [17] and flow cytometry [18]. These techniques are very useful but there is one disadvantage inherent to all of them. They require the destruction of the cell to extract the nucleic acids and proteins for quantification. While some techniques can do this on single-cell basis and allow for more accurate analysis of variation in gene expression, there is no way of obtaining quantitative temporal single-cell measurements of gene expression.

1.2 The Development and Use of Fluorescent Proteins to Study Gene Expression

1.2.1 The Discovery of GFP

Fluorescent proteins are crucial to the research being done in a wide range of fields of biology and medicine including the fields of proteomics, immunology, synthetic biology

and systems biology. They have been instrumental in many major breakthroughs in various fields of science since the first GFP clone was successfully expressed in *C. elegans* in 1994 [19]. GFP was the first fluorescent protein found in nature to be characterized. It was first discovered in 1962 in the jellyfish *Aequorea victoria*, a bioluminescent jellyfish found along the Pacific coast of America that emits green light [20].

The active bioluminescent protein was determined to be aequorin, a luminescent protein that emitted blue light in the presence of Ca^{2+} substrate [20]. The discrepancy between the expected blue light emission of aequorin and the observed green light was a strange phenomenon that only began to be uncovered in the early 1970s. It was discovered by several groups [21, 22, 23] that the aequorin luminescence wavelength (470nm) closely matches the excitation wavelength of GFP (460nm). It was posited that there was an energy-transfer from the excited aequorin to GFP occurring, resulting in green light emission. This hypothesis was later validated experimentally in 1974 [24].

In 1992, the GFP gene from the Aequorea victoria jellyfish was cloned for the first time by Chalfie [19]. The availability of GFP as a biotechnological tool enabled the detailed monitoring and study of protein location and abundance. The work of Shimomura, Chalfie and Tsien in developing and characterizing GFP as a biological tool has been so influential on the field that in 2008, they were awarded the Nobel prize for Chemistry [25].

The cloning of GFP into another organism demonstrated that fluorescent proteins would become an indispensable tool in the biological sciences. Biological processes like subcellular protein trafficking [26, 27], gene expression [28] and protein-protein interactions [29] could now be detected and tracked in an easily quantifiable manner through the intensity of fluorescence. The first version of the cloned GFP gene was not yet, however, a potent tool which could be used in different model organisms for different types of analysis. It was a dimer, with a long maturation time and a low fluorescence yield [26, 28].

Since 1992, GFP has been modified many times such that an ever-growing fluores-

cence protein toolkit has been developed. To enable fluorescent protein multiplexing, variants of different colors were required. Therefore, variants with different emission and excitation wavelength properties were developed [30, 31, 32]. To use fluorescent protein techniques in different organisms, the coding sequence of GFP was modified to fit the codons used preferentially in different organisms [33]. This modification optimized GFP expression in different cell types. In addition to these characteristics, the maturation time and the fluorescence yield were also optimized [31].

Conventional fluorescent proteins, like GFP, are versatile reporters of gene expression which can be analysed in single live-cells using flow cytometry or fluorescence microscopy. In the case of fluorescence microscopy, it is also possible to track individual cells over time and measure the dynamics of gene expression; in essence, investigating how gene expression changes over time. The characteristics of the initial GFP clone were not detrimental to the investigation of protein location and movement, but they were disadvantageous in the the study of gene expression. The poor fluorescence quantum yield, the long maturation time, and the high stability of the wild-type GFP (wtGFP) impairs the sensitivity and response time of wtGFP [31].

The first improvement on the amino acid sequence of GFP came by way of a single point mutation, S65T [34]. Improvements to the photostability, excitation characterisics, and fluorescence quantum yield of GFP were a result of the mutation. Further improvement in 1995 by Tastrup and Falcow [35] resulted in the enhanced GFP variant (EGFP). The discovered mutation improved the poor folding kinetics observed in wtGFP. Further work was done to enable the use of fluorescent proteins in other organisms. In the model organism *C. albicans*, the wtGFP gene was cloned successfully but fluorescence could not be detected. A yeast enhanced GFP(yEGFP) variant was developed by codon optiziming the coding sequence of the gene [33]. The yeast enhanced variant was expressed successfully, and fluorescence was observed, in both *C. albicans* and *S. cerevisiae*.

Further work to develop a fluorescent reporter with faster response time to changes

in gene expression was necessary for investigating gene expression dynamics. EGFP could track changes in gene expression from low expression states to high expression states faithfully but its high stability, as well as the maturation time, would cause changes from high expression states to low expression states to be hard to detect. The cause of this discrepancy lies with the cause for the change in fluorescence being different processes, which occur at different rates. Changes from low to high fluorescence depend on gene transcription and translation. Changes from high to low fluorescence depend on the degradation rate of EGFP, which is slow due to its high stability [26]. To address this weakness, destabilized variants of GFP were developed, first for use in bacterial cells [36], then for use in yeast cells [37], making modifications to EGFP and yEGFP, respectively. The destabilized yEGFP took advantage of a PEST-rich destabilization tag taken from the *CLN2* gene to reduce the half-life of the expressed protein. They succeeded in reducing the half-life from the average 7 hours to 30 minutes. They tested the reporters by cloning the genes such that they were regulated by a *CUP1* promoter, which transiently activates in the presence of copper. While both reporters reported the activation of expression, the fluorescence signal of the stable variant remains high and the fluorescence signal of the destabilized variant faithfully tracked the event as a transient spike in expression. The shorter half-life of this new variant is an asset for tracking dynamic gene regulation, however potential sources for error are introduced as a result. Due to faster protein turnover, fewer GFP proteins will be present. In the experiments conducted by Mateus and Avery, the signal produced by the destabilized GFP variant was 90% lower than that of the stable GFP variant. This issue will also increase the signal-noise ratio of the data, which makes it difficult to differentiate true fluctuations in the signal from noise.

An alternative to FPs that has seen a resurgent focus, is the family of luminescence reporter proteins (LPs). LPs, though discovered much earlier than FPs [38], saw less use as reporters due to their reliance on co-factors to complete the luminescence reaction. LPs, however, have certain characteristics which make them an enticing alternative to

FPs, especially in the study of fast gene dynamics. The most important difference between the kinetics of the two types of reporters is that FPs undergo a lengthy folding process upon completion of translation while LPs fold co-translationally [39]. Due to there being no reliance on a maturation step, an LP reporter will have a faster response to positive changes in gene expression dynamics than a FP reporter. Recent work has also gone into developing destabilized variants of FPs [7, 40], and higher photon yield variants [40]. This work has been successful in tracking the dynamic gene expression observed in the cell division and respiratory cycle of *S. cerevisiae* more accurately than their FP counterparts [40]. In addition to the issues seen in destabilized FPs, LPs have certain unique disadvantages. Their reliance on co-factors to luminesce increases the complexity of conducting experiments, especially in long time-lapse experiments. Furthermore, the photon yield of LPs is very low and it is only in the past several years that LPs bright enough to use sub-minute exposure times were developed [40]. A stable fluorescent reporter that can accurately track fast gene dynamics would be an improvement over the current state of the art.

1.2.2 FP Timers

Protein turnover rate is a very important detail that can be used to study disease. Abnormal protein turnover dynamics have been linked with a variety of diseases including neurodegenerative disorders like Huntington's and Alzheimer's disease [41]. There are limited techniques available for the measurement of protein turnover and the associated consequences of these techniques limit their versatility. Radioactive pulse-chase metabolic labeling is often used for this purpose, however this process involves the immunoprecipitation of the protein in question, and as such can't be used in living cells [42]. Another popular method is cycloheximide chase, which can be used in living cells, but since it prevents expression of genes globally, it has severe effects on a cell's metabolism. More recently a method consisting of pulse-labeling fluorescent proteins and tracking the decay of fluorescence over time has been employed. While this method

can more comfortably be used in live cells, it requires a time-lapse analysis to extract the turnover rate. [43]

Protein age is a characteristic of an expressed protein that implies its turnover rate; a higher average protein age indicates a slower turnover rate. Conventional fluorescent proteins are unable to track the age of a protein. When a fluorescent tagged protein is expressed it is only capable of emitting a digital signal; it either emits a full fluorescence signal or no fluorescence signal. This feature is useful to track the presence, abundance, and position of expressed proteins but secondary features like protein age can't be extracted. A protein that can emit a signal that changes over time in its colour (emission wavelength) would be a reporter of protein age, and therefore, a reporter of protein turnover rate.

Fluorescent proteins have been discovered that exhibit a multi-stage maturation process, consisting of two stages which fluoresce at different wavelengths [44]. The fluorescence data from this kind of protein can be expressed as an analog signal by calculating the ratio between the fluorescence intensity produced by each state. As the average age of a protein in a cell increases, the majority of protein will move from the first fluorescent stage of maturation to the last stage of maturation. These proteins have been described as timer proteins. Timer proteins are a promising technology but have seen limited use in research due to several shortcomings that limit their performance. Timer proteins have: a tendency to form oligomers, an exponential signal shape over time, and a low fluorescence quantum yield.

The Khlemminski group developed a new fluorescent protein construct, named the tandem fluorescent timer (TFT) that seeks to remedy the weaknesses of the conventional fluorescent timer [45]. The TFT leverages the same fluorescence ratio idea used with conventional fluorescent timers but separates the source of fluorescence into two separate proteins. The ratio between the fluorescence intensity of the two fluorophores that make up the TFT is useful because of their differential maturation kinetics. In a well designed TFT, one protein will mature at slower rate than another, therefore older TFT proteins

will consist of two mature, and therefore fluorescent, fluorophores than younger proteins. This framework enables one to generate TFTs with different characteristics by choosing component fluorophores to match the effective range of the TFT to the application. The TFT combines the localization tracking ability of conventional fluorescent proteins and the protein turnover tracking ability of measurement techniques pulse-chase radioactive labeling. Furthermore, TFTs can be used safely in live cell analysis. In contrast to the pulse-labeleling method of turnover measurement which requires timecourse data, a single snapshot is enough to extract protein age.

The TFT protein with optimal maturation kinetics for use in *S. cerevisiae*, is a fusion of sfGFP and mCherry. With maturation half-lives of 5 and 45 minutes respectively [45], this fusion's optimal range for turnover analysis is of proteins whose turnover ranges between 10 minutes and 8 hours [45]. In addition to the genes for mCherry and sfGFP, the TFT construct consists of a degron, downstream of an N-terminal ubiqiuitin, which is exposed following translation. The degron is a tag which triggers degradation via the N-degron pathway [46]. The rate of this process depends on two key residues in the degron. By constructing strains with different residues in the degron, variants of the TFT with different degradation kinetics can be constructed. Investigation of the steady state fluorescence ratio of the timer given different degradation rates, using various degrons, and different expression rates using various promoters reveals the steady state fluorescence ratio is not dependent on production rate. The only parameter that affects the ratio for a given TFT is the degradation rate.

Using the TFT system, the Khlemminski group studied the differential inheritance of spindle pole bodies (SPBs) by mother cells following budding in *S. cerevisiae*. By tagging the Spc42 protein found in SPBs, they were able to track the SPB structures during the budding process and discovered that the older SPB (mother cell's SPB) is inherited by the daughter cell. Using this same method, they were able to 'color-code' sites of budding in a time-progessive manner by creating a yeast strain containing a TFT tagged *RAX2* gene. Rax2 localizes to cell division sites, and is part of a structure

that forms at, and marks, sites of previous cell division. Using a single snapshot of the fluorescence ratio between mCherry and sfGFP, distinct bud scars are observed, of increasing age (ratio).

The TFT system has been shown to be a very effective reporter to study localization and protein turnover simultaneously. We suggest that the TFT can also be used as a reporter of gene expression under dynamic gene regulation. The average protein age, which is tracked by a TFT, changes as the regulatory signal changes. During times of low expression, the average age would be expected to increase, while the opposite is true during times of high expression. Furthermore, the TFT system is reliant on differential maturation kinetics of its constituent fluorophores, rather than expression and degradation kinetics, this technique should be able to track fast gene dynamics without a degradation tag.

1.3 Hypotheses and Specific Aims

The goal of my thesis research is to contribute to the development of methods to accurately measure single-cell gene regulatory dynamics by testing two hypotheses:

Hypothesis 1

Stable FP timers are effective reporters of fast gene dynamics.

Hypothesis 2

FP timers can improve the reconstruction of dynamic gene regulatory signals compared to conventional FPs

Specific Aims

Specific Aim 1: Implementation and validation of a System to measure FP expression in live yeast cells

To test my hypotheses I developed a system in which I can control the regulating signal of the FP timer which allows me to record time-lapse single cell data over a long culturing period. This design was implemented using a synthetic system in *S. cerevisiae* in which the expression of an FP timer is under the control of small molecules.

Specific Aim 2: Modeling and simulating regulated FP and timer expression

To investigate the expected behaviour of the FP timer in my synthetic system, I aimed to create a mathematical model to simulate the cellular abundance of a dynamically regulated FP timer protein.

Specific Aim 3: Development of algorithms to infer regulatory signal dynamics from simulated data

To test the FP timer's characteristics in reporting dynamic regulatory signaling, I aimed to develop a derivative based algorithm designed to infer regulatory signal dynamics and to test it using simulated data generated by my mathematical model.

Specific Aim 4: Reconstructing regulatory signals from FP expression

To determine if and when FP timers are an improvement over conventional FPs, I will compare the signal reconstruction performance of the FP timer and conventional FPs. I will use my platform to conduct periodic forcing experiments and apply my signal

reconstruction algorithm to the data.

Chapter 2

Materials

2.1 Strain Assembly

To build my dna constructs, I used homologous recombination and overlap extension PCR. In-depth comparisons and descriptions have been reported by the Kaern Lab [47]. To summarize the procedure, DNA fragments with overhangs are extracted by PCR using long primers. These overhangs provide homology between adjacent and separately extracted parts of a designed DNA construct. Homology for the locus of insertion is also necessary for targeted cloning of the DNA construct. These fragments were designed with large overlapping sequences (>40bp). I transformed all the fragments into a BY4742 strain of *S. cerevisiae*, which is a mutant strain derived from the reference strain S288C [48]. I allowed the native DNA synthethic machinery to assemble and insert the DNA construct into the targeted locus. Transformations were conducted using the lithium acetate method [49].

Transformations were visually confirmed by agarose gel of PCR products. PCRs were performed in a BioRad C1000 Touch. The PCR protocol was designed as follows: 30 reaction cycles with a denaturing step of 15 seconds at a temperature of 98C°, an annealing step of 20 seconds at a temperature of 55-60C°and an extension step at a temperature of 72C°for $30 seconds * kb$. Reactions contained: 1 unit of Phusion

High Fidelity Polymerase (M0530L, New England Biolabs), 10μM of Primers (Value Oligos purchased from Invitrogen), 20ng of template DNA, 200μM Deoxynucleotide mix (N0446S, New England Biolabs) and 1x High Fidelity Buffer (B0518S, New England Biolabs).

2.2 Quantitative Fluorescent Time Lapse Microscopy (Image Acquisition)

Single colonies from antibiotic plates were inoculated in synthetic media (1% adenine, 2% glucose). Overnight cultures were diluted to logarithmic phase(OD_{600} of 0.1) prior to any network induction. Yeast cells prepared for microscopy were cultured in CellASIC ONIX Y03C microfluidic plates during microscopy. Cell cultures were diluted to an OD_{600} of 0.07 prior to loading of the chamber. Diluted cell culture was added to well 8 (70uL), and loaded into the chamber using the ONIX software loading protocol (8 psi for 15 seconds). The media required for the designed media regime is added to the first 6 chambers. Synthetic media with doxycycline, and synthetic media with β-estradiol were added to wells 1 and 2 respectively. Synthetic media lacking any small molecule inducers was added to well 3. Before beginning the periodic signaling of doxycycline and β-estradiol, cells are perfused with non-inducing media to prime them to the environment.

Five to ten frames of view are located for each condition containing a single centered cell or colony. Imaging is controlled by Nikon NIS-Elements software. Each field of view is imaged several times per imaging cycle: a gfp fluorescence image, an mCherry fluorescence image, and bright field images at 3 focal lengths. Three bright field images at and around the centre of cells are taken to enable cell segmentation and to counteract focal drift in the Nikon PFS auto-focus system.

2.3 Image Analysis (Segmentation, Lineage Tracking)

Timelapses were extracted as .nd2 files and separated into individual .tiff image files using Nikon NIS-Elements software. Each field of view was first segmented, then tracked using code from [50] that was adapted and modified by myself and other members of the Kaern Lab. The format of the output is as .mat files containing arrays of quantified fluorescence for each detected and tracked cell.

2.4 Data Filtering

Cell profiles were filtered by completeness and length. All cell profiles shorter than 3 hours with more than 50% missing data were filtered out. The remaining cell profiles were interpolated using the spline implementation in MATLAB 2017a. Interpolation was done using a quadratic spline to fill in missing data lost due to losing cells in image analysis. Cell profiles were then smoothed with a window size of 3, which was done using the moving average smoothing implementation in MATLAB 2017a. Mean traces were calculated using the remaining filtered and smoothed cell profiles. Average sfGFP and mCherry fluorescence at each time-point was calculated and appended to create a mean fluorescence trace.

Pixel data was calculated using the segmented image data. Using the masks generated by segmentation, the pixel intensity of every segmented cell at each time point is exported into arrays. The mean trace is calculated by calculating the mean pixel fluorescence intensity at each time-point and appending the resultant means into a mean fluorescence trace.

2.5 Model Formulation and Fitting

In Figure 2.1, the method in which fluorophore maturation is described in my model is shown. I separated the population of the expressed protein into multiple populations and used two different maturation rates, one describing the faster maturation rate of sfGFP(m_1), and one describing the slower maturation rate of mCherry(m_2). The populations include the completely immature version of the TFT which is the state of a newly translated protein as well as all possible combinations of mature and immature sfGFP and mCherry. In addition, the populations of the 2 inducible transcription factors are modeled too. The network dynamics of my synthetic system are described using classical Hill equations to model promoter occupancy of the transcription factors GEV, and rTetR. The $_m$ subscript denotes that the modeled molecule is the mRNA of a protein. The parameters used in the full model are shown in Table.2.1.

2.5.1 Full Model

Expression Dynamics

$$[\dot{GEV_m}] = p_1 - \gamma_m[GEV_m] \tag{2.1}$$

$$[\dot{GEV}] = p_2[GEV_m] - k_1[GEV][est] + k_{-1}[GEV/est] + \gamma_p[GEV] \tag{2.2}$$

$$[\dot{GEV/est}] = k_1[GEV][est] - k_{-1}[GEV/est] - \gamma_p[GEV/est] \tag{2.3}$$

$$[\dot{rTetR_m}] = p_1 - \gamma_m[rTetR_m] \tag{2.4}$$

$$[\dot{rTetR}] = p_2[rTetR_m] - k_2[rTetR][dox] + k_{-2}[rTetR/dox] - \gamma_p[rTetR] \tag{2.5}$$

$$[\dot{rTetR/dox}] = k_2[rTetR][dox] - k_{-2}[rTetR/dox] - \gamma_p[rTetR/dox] \tag{2.6}$$

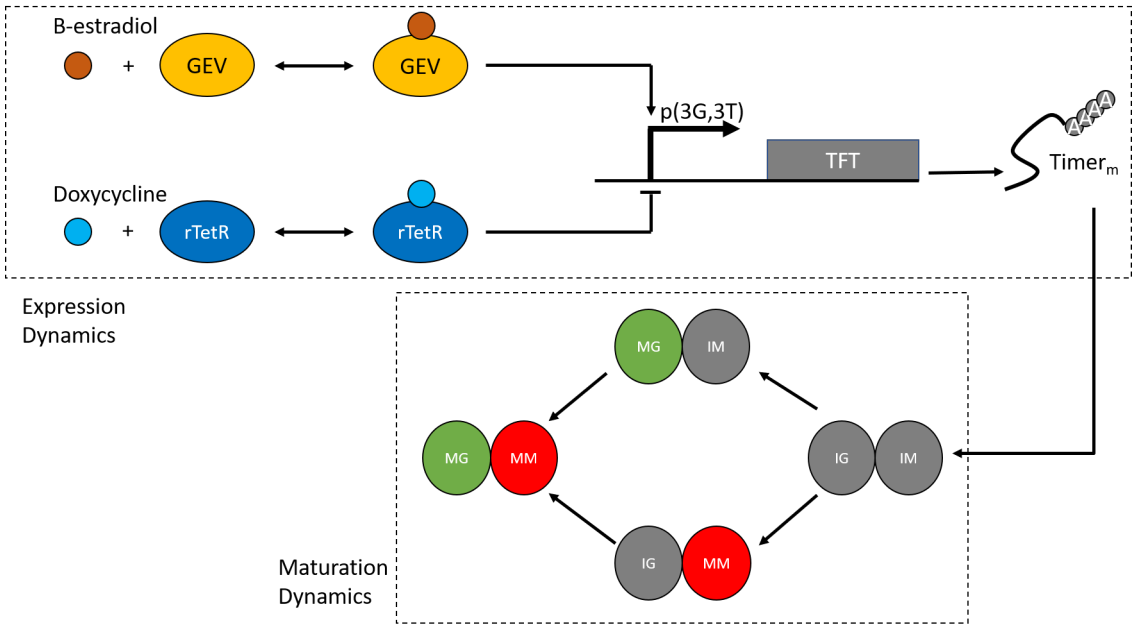

Figure 2.1: Visual description of the simplified maturation kinetics of the tandem fluorescent timer and the expression dynamics of the synthetic gene network. The tandem fluorescent timer is represented by the G(sfGFP) and M(mCherry) subunits which change colours to the colour of their emitting wavelength. The promoter, p(3GAL,3TET) contains 6 binding sites, 3 GAL binding sites that bind GEV, and 3 TET sites that bind rTetR. β-estradiol and doxycycline are exogenous molecules that are added to media as part of the experiment.

Maturation Kinetics

$$[\dot{Timer_m}] = b_0 + b_1 \left(\frac{k_d}{k_d + [rTetR/dox]^n} \right) \left(\frac{[GEV/est]^n}{k_e + [GEV/est]^n} \right) - \gamma_m [Timer_m] \tag{2.7}$$

$$[\dot{IG/IM}] = p[Timer_m] - (\gamma_2 + m_1 + m_2)[IG/IM] \tag{2.8}$$

$$[\dot{MG/IM}] = m_1[IG/IM] - (m_2 + \gamma_2)[MG/IM] \tag{2.9}$$

$$[\dot{IG/MM}] = m_2[IG/IM] - (m_1 + \gamma_2)[IG/MM] \tag{2.10}$$

$$[\dot{MG/MM}] = m_1[IG/MM] + m_2[MG/IM] - \gamma_2[MG/MM] \tag{2.11}$$

2.5.2 Derivation of Simplified Model

Derivation is identical for equations (1)-(3) and (4)-(6), therefore it is only shown for GEV and β-estradiol interactions.

Assumption 1: mRNA dynamics are on a faster timescale and thus mRNA will be at steady state.

$$[\dot{GEV_m}] = 0 \tag{2.12}$$

I then solve for $[GEV_m]^*$.

$$[GEV_m]^* = \frac{p_1}{\gamma_m}, \tag{2.13}$$

$$\tag{2.14}$$

Substituting this term into equation (2) yields:

$$[\dot{GEV}] = p_2 \frac{p_1}{\gamma_m} - k_1[GEV][est] + k_{-1}[GEV/est] - \gamma_p[GEV], \tag{2.15}$$

$$\tag{2.16}$$

Table 2.1: Parameters present in full model.

Parameter	Description	Units
p_1	transcription rate	$\frac{M}{min}$
p_2	translation rate	$\frac{M}{min}$
b_0	Basal p(3,3) Trascription rate	$\frac{M}{min}$
b_1	Maximal p(3,3) Trascription rate	min^{-1}
k_1	Beta-Estradiol and GEV Binding rate	$M^{-1}min^{-1}$
k_{-1}	Beta-Estradiol and GEV Dissociation rate	min^{-1}
k_2	Doxycycline and rTetR Binding rate	$M^{-1}min^{-1}$
k_{-2}	Doxycycline and rTetR Dissociation rate	min^{-1}
γ_m	mRNA decay rate	min^{-1}
γ_1	GEV and rTetR protein decay rate	min^{-1}
γ_2	Timer protein decay rate	min^{-1}
m_1	sfGFP maturation rate	min^{-1}
m_2	mCherry maturation rate	min^{-1}
n_1	Effective Hill Coefficient for rTetR binding	-
n_2	Effective Hill Coefficient for GFV binding	-
k_d	Activation coefficient	-
k_e	Repression coefficient	-

I combine the three parameters describing GEV expression into one rate (p_3) so equation (2) becomes:

$$[G\dot{E}V] = p_3 - k_1[GEV][est] + k_{-1}[GEV/est] - \gamma_p[GEV], \tag{2.17}$$

$$\tag{2.18}$$

Assumption 2: Since GEV and rTetR are constitutively expressed, their concentration is at steady state. Also, inducer binding occurs on a much faster timescale

than gene expression and will reach steady state very quickly upon changes in inducer concentration.

$$[\dot{GEV}] = 0 \tag{2.19}$$

$$[\dot{GEV/est}] = 0 \tag{2.20}$$

$$\tag{2.21}$$

First I solve for $[GEV/est]^*$,

$$0 = k_1[GEV]^*[est] - [GEV/est]^*(k_{-1} + \gamma_p) \tag{2.22}$$

$$[GEV/est]^* = \frac{k_1[GEV]^*[est]}{k_{-1} + \gamma_p} \tag{2.23}$$

$$\tag{2.24}$$

Every term in the equation for $[GEV/est]^*$ is a constant or a parameter. To simplify the equation I will merge the parameters into one: est, which yields the following expression:

$$[GEV/est]^* = est \tag{2.25}$$

$$\tag{2.26}$$

This means that in our model, it is assumed that the active transcription factor concentration is equal to the scaled inducer concentration which is controlled by me. I substitute this into the maturation kinetics equations, specifically the tandem fluorescent timer expression equation.

$$[\dot{Timer_m}] = b_0 + b_1\left(\frac{k_d}{k_d + dox^n}\right)\left(\frac{est^n}{k_e + est^n}\right) - \gamma_m[Timer_m] \tag{2.27}$$

2.5.3 Simplified Model

With these assumptions, the simplified model becomes:

$$[\dot{Timer_m}] = b_0 + b_1 \left(\frac{k_d}{k_d + dox^n} \right) \left(\frac{est^n}{k_e + est^n} \right) - \gamma_m [Timer_m] \tag{2.28}$$

$$[\dot{IG/IM}] = p[Timer_m] - (\gamma_2 + m_1 + m_2)[IG/IM] \tag{2.29}$$

$$[\dot{MG/IM}] = m_1[IG/IM] - (m_2 + \gamma_2)[MG/IM] \tag{2.30}$$

$$[\dot{IG/MM}] = m_2[IG/IM] - (m_1 + \gamma_2)[IG/MM] \tag{2.31}$$

$$[\dot{MG/MM}] = m_1[IG/MM] + m_2[MG/IM] - \gamma_2[MG/MM] \tag{2.32}$$

Table 2.2: Parameter values from the simplified model that were used for simulations.

Parameter	Description	Units	Value
est	Scaled β-estradiol concentration	M	
dox	Scaled doxycycline concentration	M	
p	Translation rate	min^{-1}	20
b_0	Basal Timer Transcription rate	$\frac{M}{min}$	0.002
b_1	Maximal Timer Transcription rate	$\frac{M}{min}$	0.2
γ_m	mRNA decay rate	min^{-1}	0.033
γ_p	GEV and rTetR protein decay rate	min^{-1}	0.0058
m_1	sfGFP maturation rate	min^{-1}	0.1218
m_2	mCherry maturation rate	min^{-1}	0.046
n_1	Effective Hill Coefficient for rTetR binding	-	1
n_2	Effective Hill Coefficient for GFV binding	-	1
k_d	Activation coefficient	-	0.6
k_e	Repression coefficient	-	0.6

2.5.4 Fitting

Fitting was done using the patternsearch function in the Fitting and Optimization Toolbox of MATLAB 2017a. patternsearch is a direct search method for finding local minima that requires initial conditions and an objective function to minimize. Default tolerance and maximum iteration parameters for the patternsearch method were used. The sum of the squared differences between simulated data and the experimental data was minimized using pattern search. 5 parameters were fit: $\gamma_p, m_1, m_2, n_1, n_2$. The fit parameters are shown in table 2.5.3. The initial parameters values were based on biologically relevant values obtained from the literature, and the ranges used for fitting were within biological feasibility. Simulation results were re-scaled to a range of [0,1] such that the general behavior was emphasized in fitting over abundance. The translation rate and the transcription rates were chosen such that the steady state protein count would be approximately 4000.

Chapter 3

Results

3.1 Specific Aim 1: Implementation and validation of a system to measure FP expression in live yeast cells

To enable the testing of my hypotheses that stable FP timers are effective reporters of fast gene dynamics, I developed a biological test system and a data analysis pipeline. The test system is a synthetic network implemented in *S. cerevisiae* that expresses the TFT protein under user-controlled regulation. I developed an image acquisition and analysis pipeline to record and analyse time-lapse single cell data over a long culturing period. Lastly, I developed an algorithm to reconstruct regulatory signals from fluorescence time-courses.

3.1.1 Development and implementation of an inducible synthetic construct to compare fluorescent protein gene expression tracking performance

Synthetic biology uses the building blocks of eukaryotic and prokaryotic genomes to assemble novel gene networks. In addition to genes, transcription factors (TF) like transcriptional activators and repressors are key components in synthetic biology. TFs are proteins that usually consist of two domains. The DNA binding domain (DBD) that binds specific sequences in the gene promoter region is required to target a TF's activity. The optional second domain can either be a transactivation domain (TAD) which attracts transcriptional machinery [51] or a repression domain that prevents the binding of transcriptional machinery. The different domains of transcription factors, namely the activation, the repression, and the DNA binding domains have been demonstrated to be discrete components with isolated function. This feature enables the modular creation of custom transcription factors [52].

In addition to designing transcription factors, another feature underpinning effective network design, is the placement of DNA binding sites in the promoter region. I can target transcription factors to regulate genes by placing corresponding binding sites in gene promoter regions. Furthemore, promoters can contain binding sequences for multiple transcription factors, facilitating regulation from multiple transcription factors.

An overview of the desired network behaviour is shown in Figure 3.1A. The TFT gene has binding sites for two transcription factors: an activator and a repressor. The activator (GEV) promotes expression of the TFT gene when induced by β-estradiol. Meanwhile the repressor (rTetR) blocks expression of the TFT when induced by doxycycline. The TFT protein will be expressed in an immature state and, upon successive folding of first sfGFP and then mCherry, will fluoresce at red and green wavelengths.

I first designed and assembled 3 DNA parts that form a genetic network and transformed them into *S. cerevisiae*. The order, components and location of the DNA parts

used are shown in Figure 3.1B. I needed to design a genetic network whose expression of the TFT could be induced and repressed strongly in a dynamic manner. GEV [53] and rTetR [54] were chosen as targetable, inducible transcription factors. GEV consists of a *GAL4* DNA binding domain and a vp16 transactivation domain. Meanwhile, rTetR only consists of a TET binding domain so I fused a Sin3 [55] repressor protein to turn rTetR into an enzymatic repressor.

The inducible nature of GEV and rTetR allows for manipulation of the active transcription factor concentration dynamically via the regulation of the concentration of small inducer molecules in the media. In my designed network, the GEV and rTetR transcription factors are expressed contstitutively in their inactive form, such that the total GEV and rTetR concentration is expected to be at steady state. Thus the timescale of GEV and rTetR activity is dictated by inducer binding and entry into the nucleus rather than the expression of the GEV and rTetR genes.

I used a dual input promoter composed of 3 GAL binding sequences and 3 TET binding sequences. This enables the binding of two different transcription factors to the upstream region of the TFT gene. By fusing a transactivation domain to the GEV gene and an repressor to the TeT gene, it is possible to regulate expression of the TFT gene positively and negatively with just one dual-input promoter upstream of the TFT gene.

The DNA parts assembled in this work are shown in Figure. 3.1B. The first DNA part assembled and cloned into the host organism was part 1, which contained a dual input promoter, p(3GAL,3TET), and the TFT gene. This part was cloned into the *ADE2* locus because a deletion in the *ADE2* gene of *S. cerevisiae* induces an accumulation of red pigment, causing colonies of successfully cloned populations to be red [56]. In addition to colony color, screening of transformants was be done by fluorescence in galactose enriched media. The 3 GAL sites in the promoter of the TFT gene are bound by the gal4 transcription factor from the galactose metabolism pathway of yeast, inducing expression of the TFT protein.

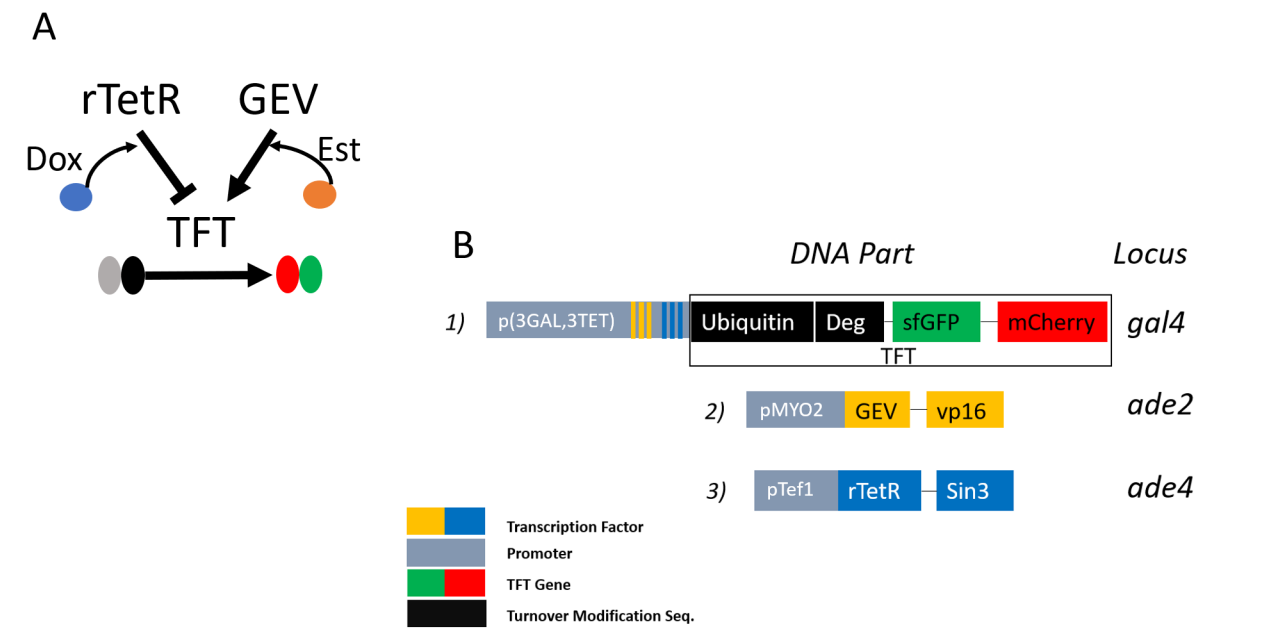

Figure 3.1: Visual representation of the synthetic gene network assembled for in vivo testing of the tandem fluorescent timer. A) The interaction between expressed proteins in the designed network is shown visually. Transcription factors rTetR and GEV repress and activate TFT expression, respectively, when activated by their respective inducer molecules. Doxycycline(Dox) and β-Estradiol(Est) are small molecule inducers that are added exogenously. B) The detailed color-coded maps for the three transformed DNA constructs as well as the loci in which they were inserted are shown. Transcription factor binding sites are shown using small color-coded stripes in the promoter region of a gene.

DNA part 2 contained the GEV-vp16 gene regulated by the strong constitutive promoter pMYO2, an endogenous promoter of the *MYO2* gene. It was cloned into successful transformants at the *GAL4* locus of *S. cerevisiae*. The *GAL4* locus was chosen to prevent *GAL4* expression. GEV and gal4 bind the same DNA sequence and would interfere with the induction activity of the newly transformed GEV-vp16 gene. Successfully cloned cells no longer expressed TFT in galactose enriched media, and were confirmed by screening for β-estradiol-induced fluorescence.

The last DNA part to be cloned was DNA part 3 and it was inserted at the *ADE4* locus. DNA part 3 consisted of the rTetR protein with a Sin3 repressor tag (rTetR-Sin3). Insertion at the *ADE4* locus was chosen due to the effect on colony colour upon successful deletion of the *ADE4* gene in a *ADE2* deletion mutant of *S. cerevisiae*. The red colony phenotype observed in the *ADE2* deletion mutant [56] reverts back to a white colony phenotype if the *ADE4* gene is deleted. In addition to the selection techniques made available by the chosen loci and the order of transformation, further confirmation of successful cloning was done by PCR.

3.1.2 Development of an image acquisition and analysis pipeline to produce single-cell gene expression profiles

To investigate the behaviour of the TFT system, I developed a data acquisition and analysis pipeline. A visual representation of the individual processes in this pipeline are shown in Figure 3.2. A vital feature of my system is the use of a microfluidic device. I used a microfluidic device, shown in Figure 3.2A,B to culture my cells during multi-hour fluorescence imaging experiments. The pressure control system enabled the automated switching of media conditions in the cell culture, such that any dynamic media conditions could be used, including a periodic repression/induction regime.

I was interested in analyzing individual cell traces over long periods of time, therefore I used a fluorescence microscope with an automated stage for image acquisition. I recorded multiple colonies for 12-24 hour experiments at an imaging resolution of 10

minutes (6 images per hour) that produced cell histories over colony growth of up to 8 doublings.

I helped develop image analysis software for cell segmentation and tracking of *S. cerevisiae* cells cultured in a microfluidic plate and imaged using a Nikon Ti-E microscope. Cell segmentation was done primarily using an adaptive thresholding algorithm followed by watershedding. Successful implementation of our algorithm would yield images like the ones shown in Figure 3.2C, which overlaid red contours over the detected cell boundaries. Detected cell boundaries are used as input to our cell tracking algorithm. The goal of cell tracking is to accurately assign imaged cells with labels over the duration of experiment. This means that the identity of cells has to be matched in every successive image. Our cell tracking algorithm is based on the Hungarian algorithm published by Ricicova et al. in 2013 [50]. We improved the implementation and adapted their code to fit the specific images acquired using my pipeline.

The next step in the pipeline is the use of the extracted cell profiles to construct and analyze single cell and mean traces as shown in Figure 3.2 D,E. Fluorescence intensity traces for the mCherry and sfGFP components of the TFT (Figure 3.3A,C) are produced by smoothing the raw cell profiles created using our cell tracking code. The mCherry/sfGFP fluorescence intensity ratio trace (Figure 3.3B,D) was produced using the smoothed mCherry and sfGFP traces. These traces would be created for each individual cell, such that for each cell there would be a fluorescence history of up to 20 hours for each successfully imaged and tracked cell. The mean trace, as shown in Figure 3.2D with the standard deviation, would be created using the individual fluorescence and fluorescence ratio traces to investigate population level dynamics. Details of the data-filtering protocol are available in the Materials section.

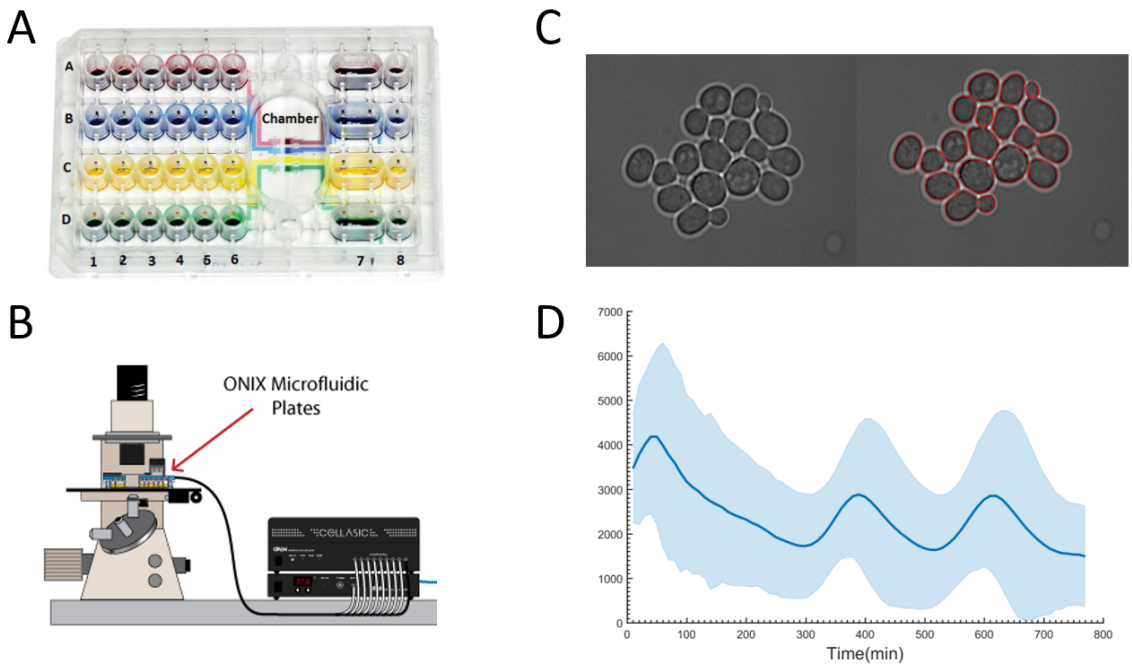

Figure 3.2: Data acquisition and analysis pipeline developed for this work. A) A microfluidic device paired with a B) microfluidic pressure controller were used to contain growing cell culture during imaging with an automated inverted microscope. C) Bright field images were taken for the purpose of segmentation and tracking using our algorithm. D)sfGFP and mCherry fluorescence images were quantified and combined with cell lineages to form single-cell fluorescence profiles. Figures A and B are adapted and edited from ONIX promotional material (http://www.emdmillipore.com/CA/en/product/CellASIC-ONIX2-Microfluidic-System,MM_NF-CAX2-S0000).

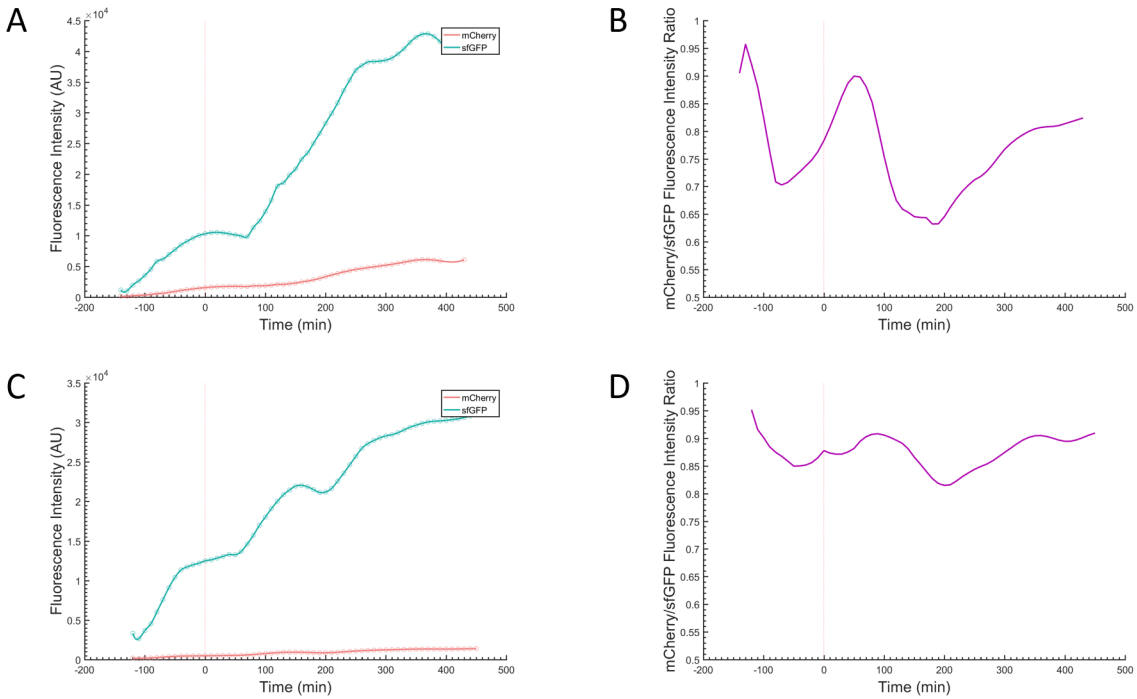

Figure 3.3: Characterization of the fluorescence change when an repression signal is removed. Network is induced by a saturating repression signal and the removal of an activation signal at time 0. Time 0 is denoted by a vertical dotted red line. Data is shown for both the stable variant (A,B) of the TFT and the unstable variant of the TFT (C,D). A,C) sfGFP and mCherry fluorescence calculated using the mean of all recorded cell profiles. B,D) Mean mcherry/sfGFP ratio trace calculated using the mean fluorescence data shown in A and C.

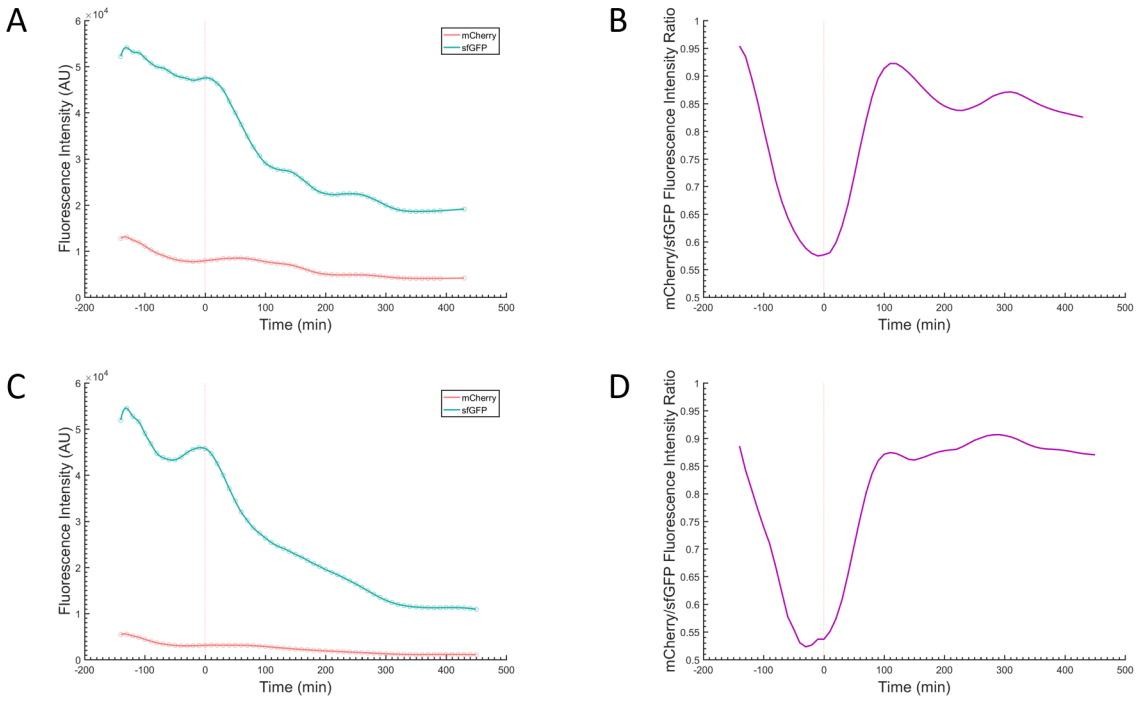

Figure 3.4: Characterization of the fluorescence change when an activating signal is removed. Network is induced by a saturating activation signal and the removal of a repressing signal at time 0. Time 0 is denoted by a vertical dotted red line. Data is shown for both the stable variant (A,B) of the TFT and the unstable variant of the TFT (C,D). A,C) sfGFP and mCherry fluorescence calculated using mean of all recorded cell profiles. B,D) Mean mcherry/sfGFP ratio trace calculated using the mean fluorescence data shown in A and C.

3.1.3 Characterization and testing of the TFT testing platform

Upon successful transformation of the three DNA parts comprising the genetic circuit, experiments were conducted to determine the response of the TFT expressed from our designed circuit. Two simple regulatory regimes were used in these experiments: high rTetR/low GEV to low rTetR/high GEV and vice-versa. More simply I named these two regimes Off to On and On to Off since they represented the two possible changes between two discrete states: repression and induction. In Figure 3.3, the initial system characterization is shown. The mean traces for the sfGFP, mCherry and the mCherry/sfGFP ratio are shown for both the stable (Panels A and B) and unstable (Panels C and D) variants of the TFT under regulation by the Off to On regime.

The expected response to a regime of this kind is an increase in sfGFP and mCherry signal, and a temporary dip in the mCherry/sfGFP ratio followed by a return to the steady state. This response is observed in the data I obtained in my experiment as seen in both the mean fluorescence traces and the fluorescence ratio trace. There is a delay of approximately 80 minutes observed between the moment the signal changes to an On signal, marked by the dot red line at time 0, and the increase in fluorescence. This is reflected by an identical delay in the dip seen in the ratio trace. The stable sfGFP signal appears to begin to oscillate in Figure 3.3 even though the regulatory signal is a constant activating signal. Since the oscillation begins to occur as the sfGFP fluorescence profile rapidly approaches a new steady state, the likely cause is the settling of the TFT concentration to the new steady state equilibrium.

The opposite regime is characterized in Figure 3.4. The expected response is an increase in the mCherry/sfGFP ratio and a decreasing mCherry and sfGFP fluorescence. This response is observed in both the stable and unstable variants of the TFT. Furthermore, the onset of the response is much faster than in the Off to On regime. In Figures 3.3 and 3.4, there is variation in the fluoresence traces prior to the induction of a new regulatory signal. The fluctuation in fluorescence response is worrying because

it demonstrates a potential for variation and noise to confuse true changes in the regulatory signal. I consider that this may be an artifact from the cell loading and priming process of the microfluidics platform. Loading of yeast cells into the microfluidics device stresses them and may lead to increased variation. The response to the changing regulatory signal is much greater than the variation observed prior and therefore I do not think it will have a great effect on my ability to compare the effectiveness of the TFT protein and sfGFP.

I developed and characterized a testing platform to enable the testing of my two hypotheses.

3.2 Specific Aim 2: Modeling and simulating regulated FP and TFT expression

To study and make predictions regarding the behaviour of the test system under dynamic regulatory regimes, I developed a mathematical model of my system. My model incorporates the regulatory dynamics of the network as well as the dynamics of fluorophore maturation. It was necessary to include the different maturation dynamics of the two fluorophores mCherry and sfGFP because it is the reason for the response in the mCherry/sfGFP ratio.

3.2.1 Model formulation

3.2.2 Model characterization and parameter fitting

To characterize my constructed synthetic system, I ran simulations using my model to compare the simulated dynamic behaviour to previous characterization experiments, shown in Figures 3.3 and 3.4. I simulated a switch from On to Off (Figure 3.5D,E) and Off to On(Figure 3.5B,C), reflecting the experiment shown in Figure 3.3. The response observed in the model simulations closely mirrors the response in the experimental

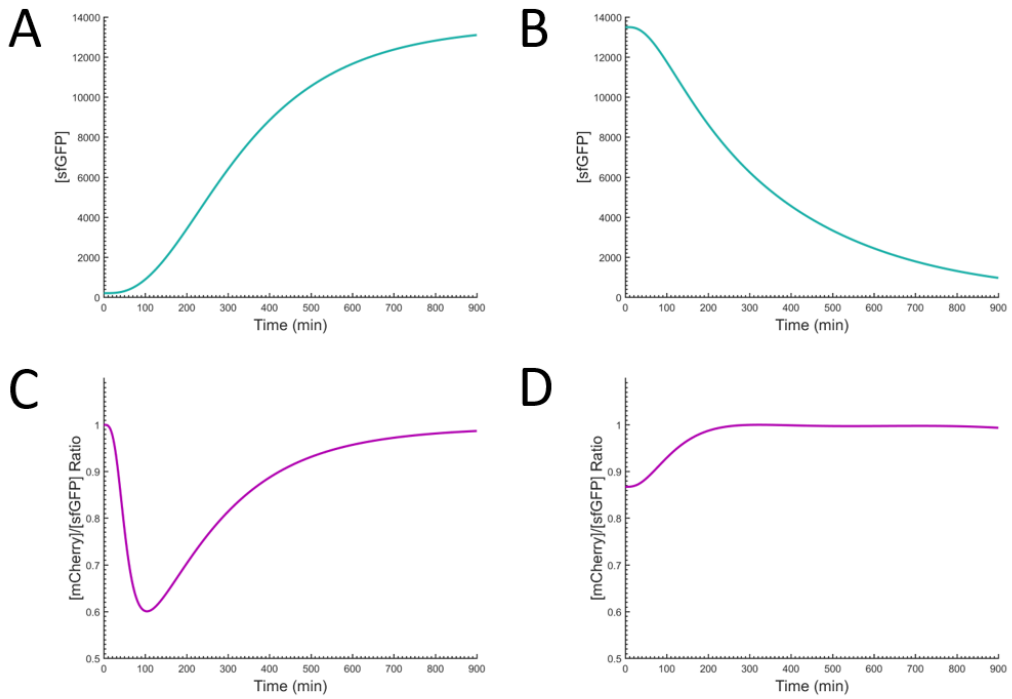

Figure 3.5: Simulations of timer expression kinetics under single-step regimes demonstrate step change in the sfGFP fluorescence response and the mCherry/sfGFP ratio response. Data for the following conditions are shown: A) sfGFP On—Off, B) mCherry/sfGFP ratio On—Off, C) sfGFP Off—On, D) mCherry/sfGFP ratio Off—On. Regulatory signal step-change occurs at time zero.

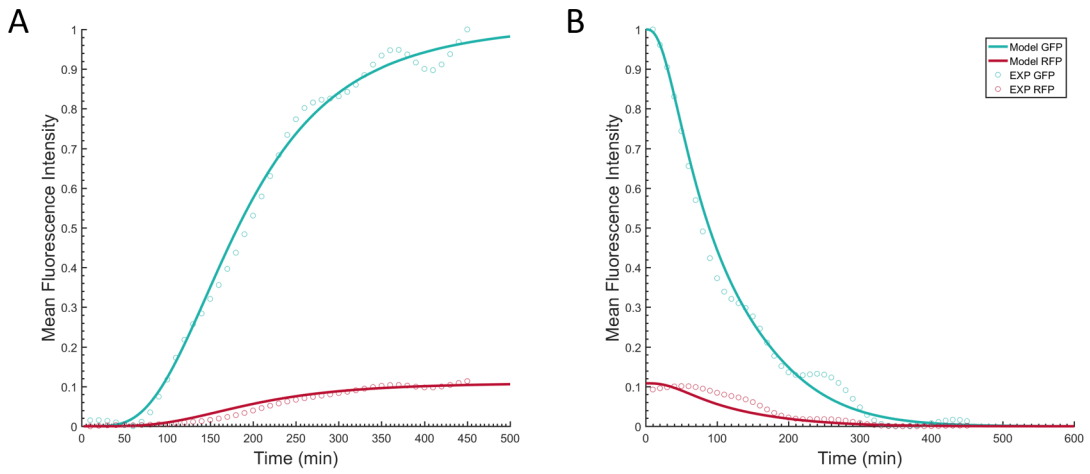

Figure 3.6: Tandem Fluorescent Network Model Parameterization. A) Fitting of RFP and GFP data from Off-On experiment. B) Fitting of RFP and GFP data from On-Off experiment. Fitting was conducted using the patternsearch function in MATLAB 2015a. Default function parameters for tolerance and iterations were used. The sum of the squared differences was used as the the objective function for minimization. N=1804 cells (A), 533 cells (B)

data traces. Upon removal of an activating signal, the ratio increases, and remains at a higher steady state than the initial condition. Meanwhile, when an activating signal replaces a repression signal, the ratio dips and returns to the previous steady state. The response observed in the model simulations occurs much quicker following the change in signal than in the experimental data.

To more closely resemble experimental data, and allow for my informed predictions, I fit the model parameters. Using the numerical fitting method of pattern search, I fit 6 model parameters to the experimental data shown in Figure 3.3. The parameters chosen for fitting were the Hill coefficient and the EC50 for both transcription factors as well as their nuclear exit rates. The model simulations were fit to both the On-Off and the Off-On experimental data; the final fit is shown in Figure 3.6. The initial conditions used in the simulations for fitting mirrored those used in the single change experiment: saturating levels of GEV for the On-Off simulation and saturating levels of rTetR for the Off-On simulation.

The fit parameters, shown in the Materials section, are used in all the following simulations done using our model. Using this fit model, I simulated the network under a periodic signaling regime of 3 different frequencies: 1, 0.5 and 0.25 cycles per hour. The simulation output is shown in Figure 3.7. It is clear that both the sfGFP and the ratio trace track the periodic regulatory signal, however certain features are more pronounced in the ratio trace. The amplitude appears to be noticeably higher in the ratio trace. It is also not exhibiting a downward trend, like observed in the sfGFP fluorescence trace. The simulations done using my fit model suggest there are potential avenues by which the mCherry/sfGFP trace could outperform the sfGFP trace in tracking a dynamic regulatory signal.

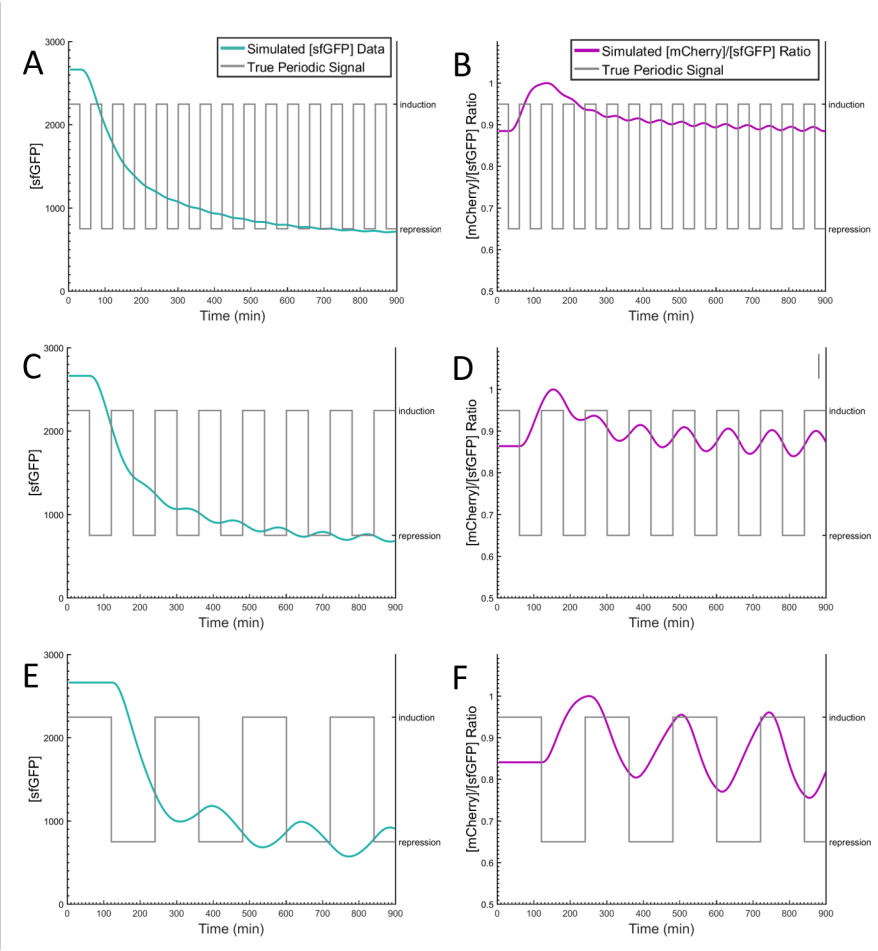

Figure 3.7: Periodic response is observed in both the sfGFP fluorescence and the mCherry/sfGFP ratio at all simulated periods. Model simulations were done with a periodic forcing signal using fit model parameters. Regulatory signal change occurs at time zero. The regulatory signal is plotted as a grey box wave. All simulations are done assuming a stable tandem fluorescent timer protein. The following conditions are shown: A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period, E) sfGFP – 240min period, F) mCherry/sfGFP ratio – 240min period.

3.3 Specific Aim 3: Development of an algorithm to infer regulatory signal dynamics from simulated data

To enable comparison of the performance of different reporters, I developed an algorithm to infer regulatory signals from time-lapse reporter data. Using the inferred regulatory signal from different reporters, I am able to quantify and compare their accuracy. Using a derivative based algorithm, my code is designed to detect discrete changes between two expression modes: On and Off (induction/repression). This reflects the two possible signal states possible using the microfluidic system employed in this work.

3.3.1 Algorithm development

A flowchart describing the algorithm I developed is shown in Figure 3.8. It was necessary to choose a feature of the reporter data that reflected changes in expression dynamics. There is an obvious correlation between fluorescence intensity and the regulatory signal; changes between states of increasing and decreasing fluorescence are markers of changes in regulatory signal. I chose to use the first derivative's zeros or the second derivative's local minima and maxima as markers of signal change.

The algorithm is comprised of two major operations: the detection of features denoting switches between state, and the sorting and assembly of detected events into a discrete signal. I denote a switch from an increasing fluorescence signal to a decreasing signal as an off-signal(off-time) and the reverse as an on-signal(on-time). These names are given because it is assumed in the formulation of the algorithm that these events are correlated with the changes in the regulatory signal.

The algorithm is separated into two steps: the detection step and the sorting step. The design is such that the sorting step is only dependent on the output of the detection step. I exploit the isolated nature of the detection step by employing two different

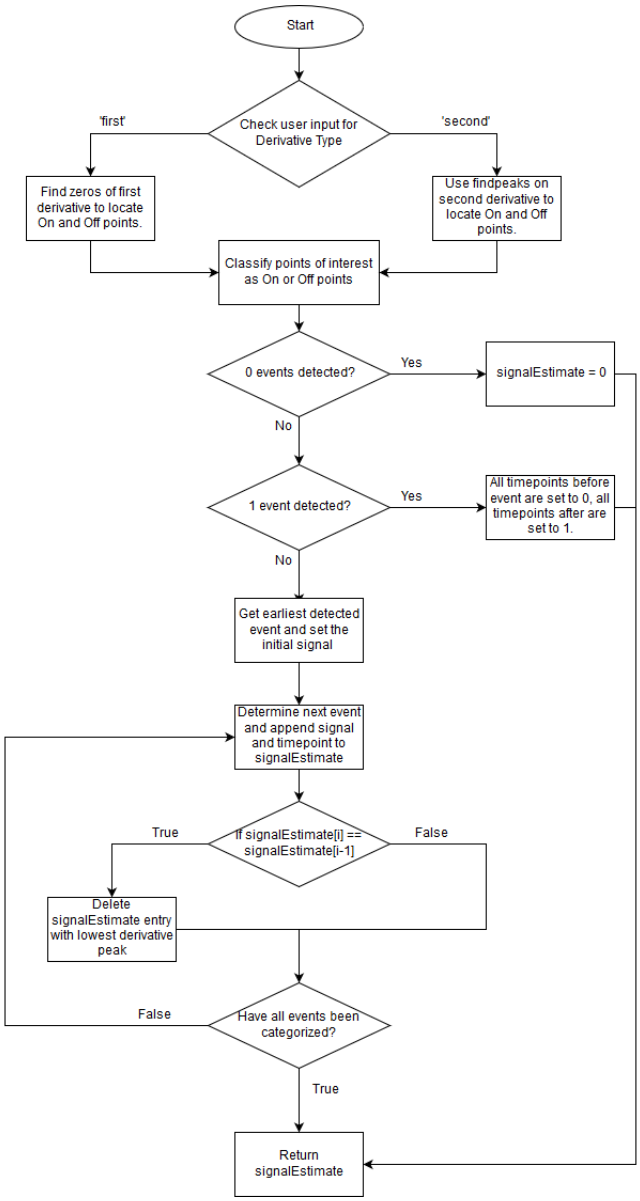

Figure 3.8: Flowchart describing the algorithm designed for periodic signal reconstruction from experimental time-lapse fluorescence microscopy data. Diamond boxes indicate situations in which a decision has to be made. Rectangular boxes indicate operations that are to be carried out. Ovals indicate the start of the algorithm logic. The variable signalEstimate is an array containing zeroes and ones which when paired with a signalTime array, describes the estimated signal. The time when switches in the signal occur are stored in signalTime, the type of signal at each time is stored in signalEstimate.

methods for feature detection: the aforementioned first derivative zeros and the second derivative local minima and maxima. The output of the detection step, regardless of the method employed is an array of detected off-times, an array of detected on-times, and an array of the derivative values detected at each on and off-time. Ultimately, I observed that while the second derivative peaks should be more accurate of a feature, it was also clear the derivation of a discrete signal introduces noise that outweighed any improvement in accuracy that it enabled. As a result, all the data analysed using my signal inferring algorithm first underwent feature detection using the first derivative. Upon successful feature detection, edge cases which did not require sorting are filtered out. Otherwise the sorting and assembly portion of my algorithm is used to infer a regulatory signal from the detected on and off-times.

The sorting algorithm loops through the arrays of on-times and off-times once until it reaches the last detected event and constructs a signal by appending subsequent opposing events in an array. This means that only an off-time may follow an on-time and vice-versa. A best-case scenario would entail all detected events to alternate between on and off times, and the sorting algorithm would append each event until it reaches the last event. However, in the case of adjacent identical events (e.g. two on-times without an off-time in between), heuristic logic is required to choose the right event to store.

Adjacent identical events can only be found when using the second derivative feature discovery method. This is the case because the zeros of the first derivative will always alternate between a change from positive to negative and a change from negative to positive. As such, sorting of the events is only necessary when events were detected in the second derivative. Adjacent event conflicts are resolved by comparing the height of the peak detected in the second derivative for each of the events, and removing the lower peak. This is done because I assumed that fluctuations in the fluorescence trace that resulted in peaks in the second derivative were likely to be of lesser amplitude than those caused by external regulatory signaling.

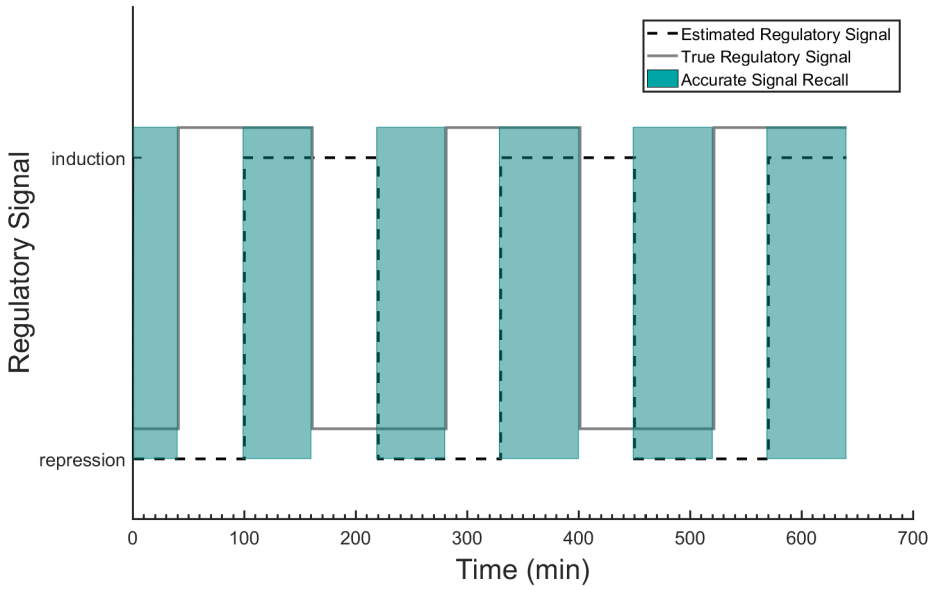

Figure 3.9: The accuracy calculation used to score the performance of the algorithm with different data and experimental conditions is demonstrated visually. Green rectangles indicate the sections of the data that align between the detected signal and the true signal. The accuracy is calculated as the fraction of time that the predicted signal is correct.

3.3.2 Algorithm testing using simulated data

I tested my signal recreation algorithm using simulated data. To evaluate the performance of my algorithm, I also described a method to quantify the accuracy of my recreated signal. The method, shown in Figure 3.9, compares the true regulatory signal to the reconstructed signal using the fraction of time the signals are identical. Using this metric as a measure of accuracy, there are several key score values. A score of 1 is perfect, a score of 0.5 can be achieved by guessing a constant signal, and a score of 0 indicates a perfectly recreated signal, albeit out of phase by half a wavelength.

In Figures 3.10 and 3.11, the use of the signal reconstruction algorithm is shown on a set of simulated data of a stable timer protein under a forced periodic signal. Three different periods are used: 60 minutes(A and B), 120 minutes(C and D), and 120 minutes(E and F). The simulation begins with a 900 minute priming period (not shown) of a saturating activating signal prior to the beginning of the periodic signaling. This is done so that the periodic forcing begins with a system at steady state. Time 0 in the plots shown in Figures 3.10 and 3.11 is 200 minutes after the periodic signal begins. There is a downward trend in the sfGFP trace, most notably in the 60 minute period plot in Figure 3.10A. This is caused by the system's steady state being disturbed and a new steady state, which approximately half the intensity of the initial steady state, is approached. The downward trend, paired with the low amplitude of the 60 minute period sfGFP simulation data, leads to no detected on/off-times in over half of the duration of the simulation data. This effect is observed considerably less in the ratio trace. Furthermore, the amplitude of the periodic fluorescence ratio response is greater than the sfGFP periodic response. This indicates why more signal events are detected in the ratio profiles than the sfGFP profiles. The disparity between Figures 3.10A and 3.10B demonstrates this clearly.

The accuracy of the detected signals is shown in Table 3.1. While the shape of the true signal is recreated almost perfectly, there is a severe phase-shift that leads to very low accuracy. An example of this is the sfGFP fluorescence profile for the 120 minute

Table 3.1: Simulation Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method

	sfGFP		Ratio	
Period	Unstable	Stable	Unstable	Stable
60min	0.39	0.40	0.27	0.29
120min	0.19	0.19	0.35	0.33
240min	0.50	0.48	0.79	0.76

period experiment (Figure 3.9C). The phase-shift is present in both the sfGFP trace and the ratio trace, but it is more severe in the sfGFP trace. This delay is mostly caused by the timescales of activation and repression in the synthetic test platform. In addition to translation and transcription, a switch in expression requires the diffusion of inducers into cells, the binding of inducers to transcription factors, transport, and binding of transcription factors in the nucleus.

Table 3.2: Shifted Simulation Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method

	sfGFP		Ratio	
Period	Unstable	Stable	Unstable	Stable
60min	0.67	0.65	0.91	0.89
120min	0.78	0.75	0.96	0.97
240min	0.84	0.81	0.85	0.88

I correct for the phase-shift in the reconstructed signal by implementing a delay. Using a 50 minutes delay as a benchmark, I shifted the reconstructed signal back by 50 minutes. The accuracy of the signal increases substantially, as shown in Table 3.2. Comparing the plots in Figures 3.10 and 3.11 shows that that the signal shape and frequency is recreated almost perfectly in the ratio trace, and also in the sfGFP trace once the amplitude overcomes the downward trend. The accuracy of the signal

recreated using the fluorescence ratio of 120 minute period data increases nearly 3-fold from 0.33 to an almost perfect 0.97. Similar results are observed for all the other data as well. The same analysis was done using simulations of an unstable TFT protein and the accuracy is reported in Tables 3.1 and 3.2. Nearly identical accuracy results are observed for the simulations of the unstable TFT variant. This is not unexpected since the deterministic nature of the simulations are not subject to the expected deficiencies related to noise. The results of this testing suggests that if it is possible to precisely determine the delay in the response of the timer protein, accurate regulatory signals can be detected, with temporal accuracy.

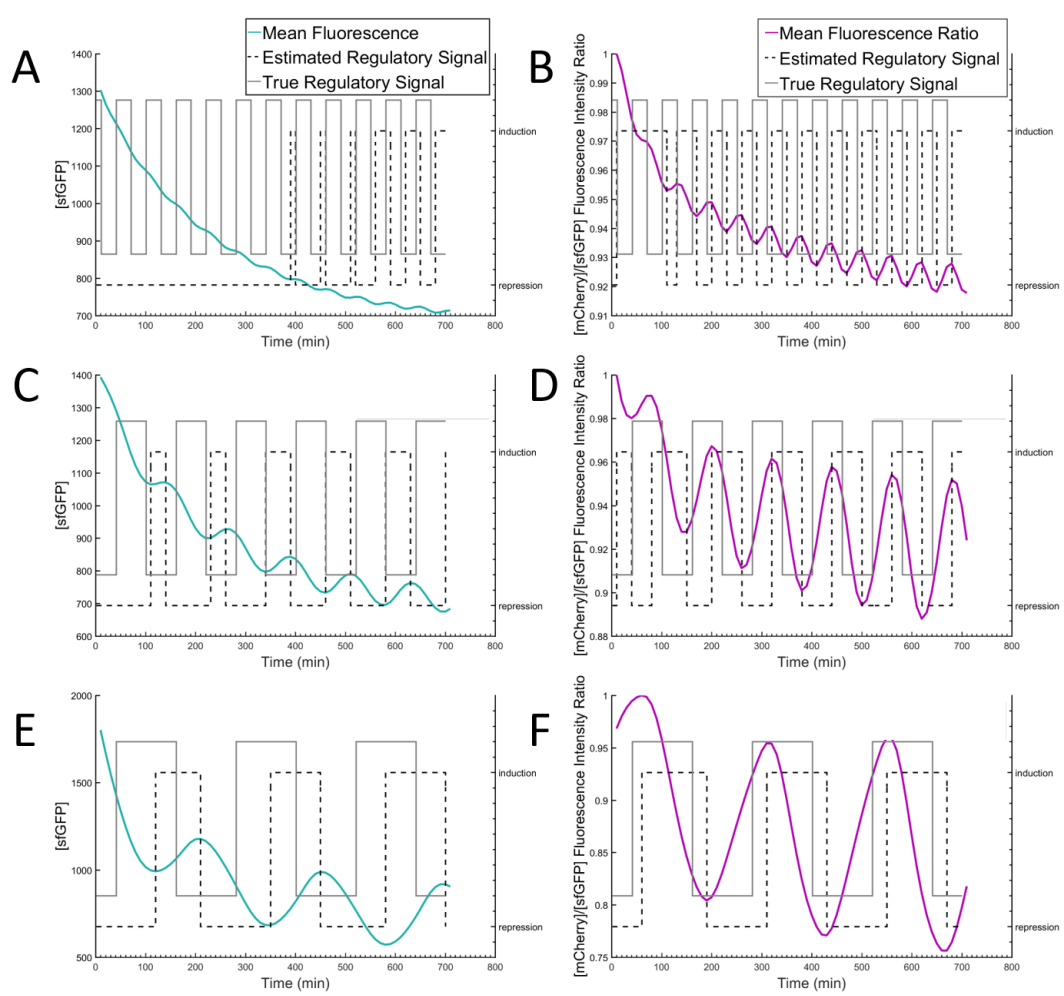

Figure 3.10: The mCherry/sfGFP ratio has a wider range of effectiveness as a reporter of dynamic gene expression. Periodic signal reconstruction using the developed algorithm on simulations of periodically regulated stable TFT expression is shown. The true signal is plotted using a solid grey box wave. The reconstructed signal is shown using a dotted black box wave. The following conditions are shown: A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period, E) sfGFP – 240min period, F) mCherry/sfGFP ratio – 240min period.

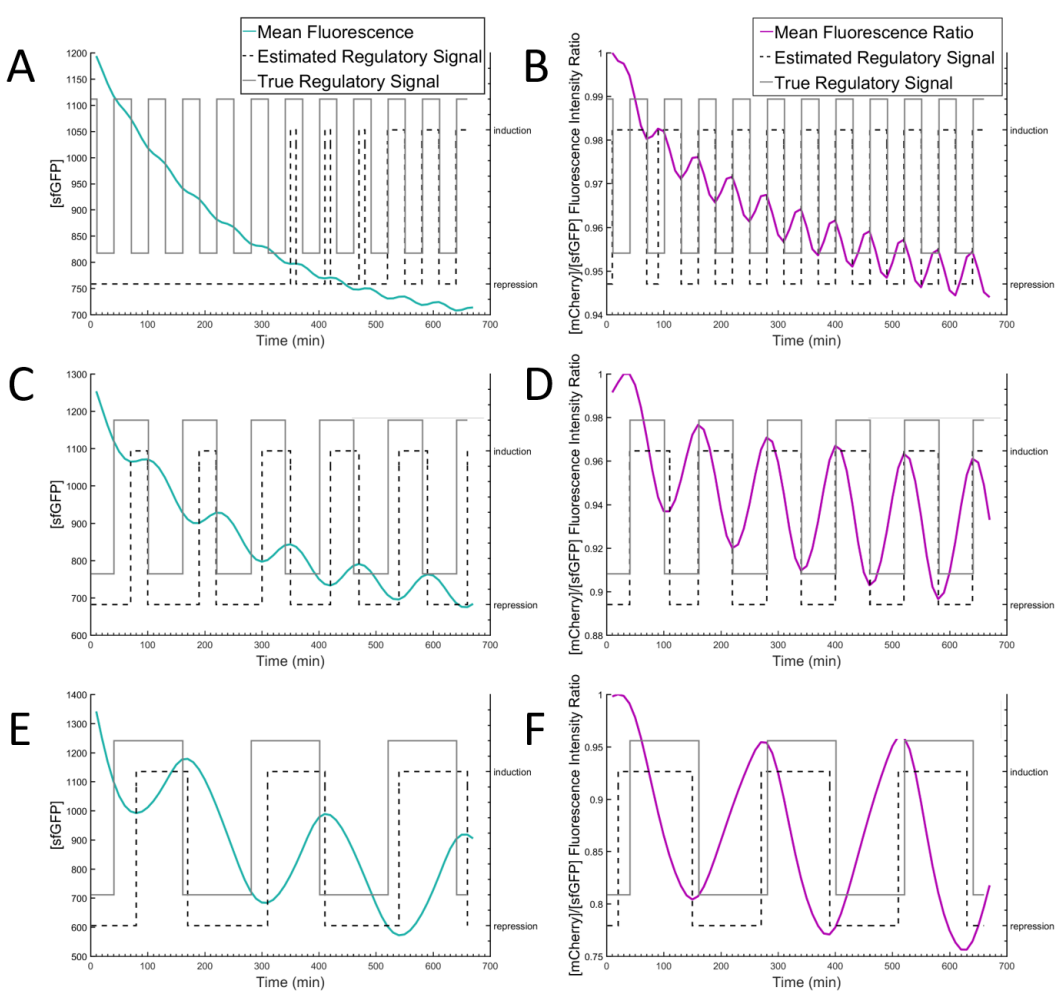

Figure 3.11: Shifting the fluorescence traces rectifies the reduced accuracy caused by the delay in the response to a dynamic signal. Periodic signal reconstruction using developed algorithm on time-shifted simulations of periodically regulated stable TFT expression is shown. The true signal is plotted using a solid grey box wave. The reconstructed signal is shown using a dotted black box wave. The following conditions are shown: A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period, E) sfGFP – 240min period, F) mCherry/sfGFP ratio – 240min period.

3.4 Specific Aim 4: Reconstructing regulatory signals from FP expression

To test my second hypotheses, I conducted periodic forcing experiments and employed my algorithm on experimental data. The same regulation pattern that was used to generate the simulated data was used to in my experiments. Using the platform described in Aim 1 (Figures 3.1, 3.2), I conducted time-lapse quantitative fluorescence microscopy experiments to track yeast cells periodically expressing the tandem fluorescent protein. I first examined the individual cell traces obtained from the set of experiments conducted.

3.4.1 Signal reconstruction using single-cell FP and TFT fluorescence data

The sfGFP fluorescence traces and ratio traces for a 60 minute and a 120 minute period signal is shown in Figure 3.13. In Figures 3.13A,B, the cell profiles for the stable timer variant under a periodic signal of 60 minutes is shown, whereas in Figures 3.13CD, the cell profiles for the stable variant under a 120 minute periodic signal is shown. The periodic regulatory signal is reflected in the 120 minute period traces traces in both the mean and the individual cell profiles, however the mean in the 60 minute variant traces do not appear to exhibit any periodicity. The individual cell profiles of the ratio oscillate between high and low intensity ratio. The oscillation in the ratio cell profiles of the 60 minute periodic signaling experiment appears to be out of phase, which may indicate why the mean trace is effectively a straight line. However, the oscillation in the 120 minute period ratio trace is much closer to being in phase, leading to a clear trend in the mean trace.

The model simulations predict there is going to be a delay between the fluorescence signal and the underlying pulsing regulatory signal. Variation in this delay is a potential cause for the variation in the phase shift seen in individual cell profiles. Since the

delay, previously found to be approximately 50 minutes, is very close to the signal pulsing period, the same amount of variation will have a greater effect on the shorter period signaling trace than longer period traces. A strategy to detect the underlying periodic signal using the mean trace would likely not work for the 60 minute period data. Alternatively, applying the algorithm to each cell profile would yield many potential signals, each with a different phase shift.

Table 3.3: Single-Cell Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method

	sfGFP		Ratio	
Period	Unstable	Stable	Unstable	Stable
60min	0.49	0.42	0.42	0.45
120min	0.40	0.41	0.31	0.46
240min	0.65	0.47	0.62	0.67

Table 3.4: Shifted Single-Cell Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method

	sfGFP		Ratio	
Period	Unstable	Stable	Unstable	Stable
60min	0.50	0.54	0.55	0.48
120min	0.62	0.58	0.73	0.60
240min	0.75	0.63	0.73	0.67

Even with this issue identified for the 60 minute period data, I moved forward with analysing the mean fluorescence trace data using my signal reconstruction algorithm. In Figure 3.13 the mean GFP and ratio traces are shown for the 3 different period lengths (60, 120 and 240 minutes), as well as the signal detected using my algorithm. The signal detected for the 60 minute period traces are unrecognizable; as seen in Figure 3.12 the mean traces are not an accurate way to track the regulatory signal. The rest of the

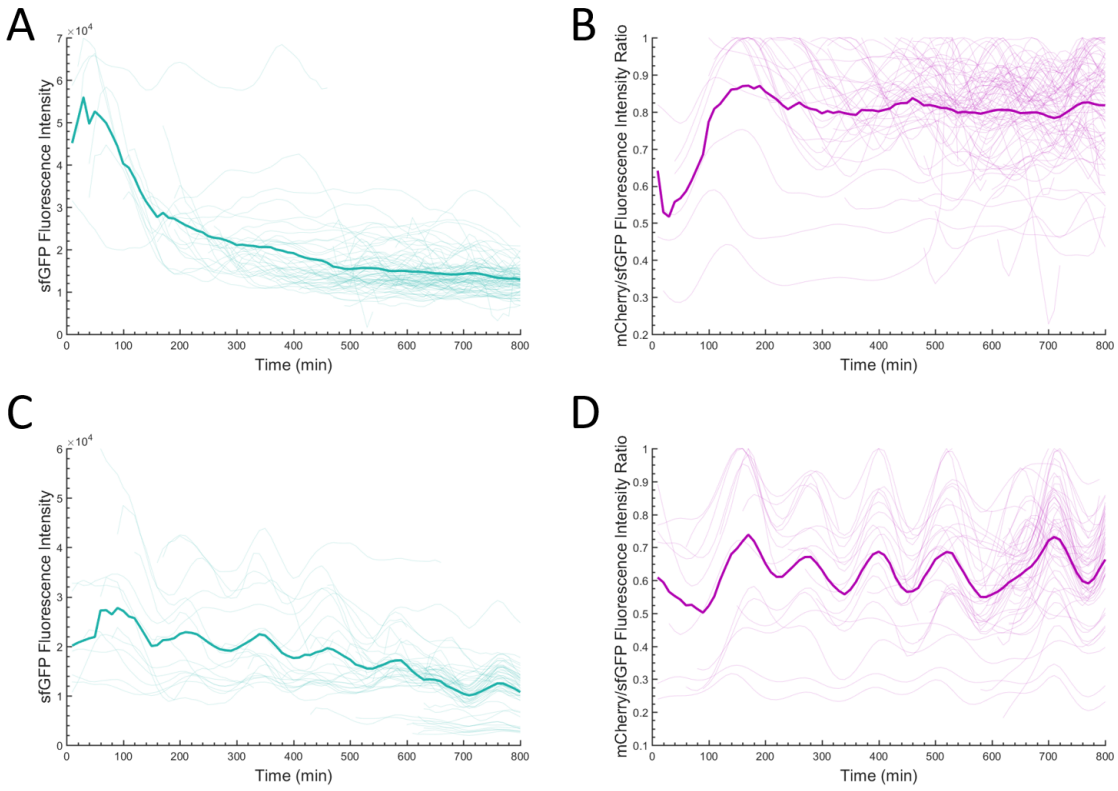

Figure 3.12: Variation in individual cell profiles has a negative effect on mean fluorescence traces at shorter periods of dynamic signalling. Representative plots of cell profiles tracked from periodic forcing experiments are shown. All plots are derived from data obtained for the stable variant of the tandem fluorescent timer. The thicker and darker line denotes the mean trace calculated using the individual cell profiles. Regulatory signal change occurs at time zero. A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period.

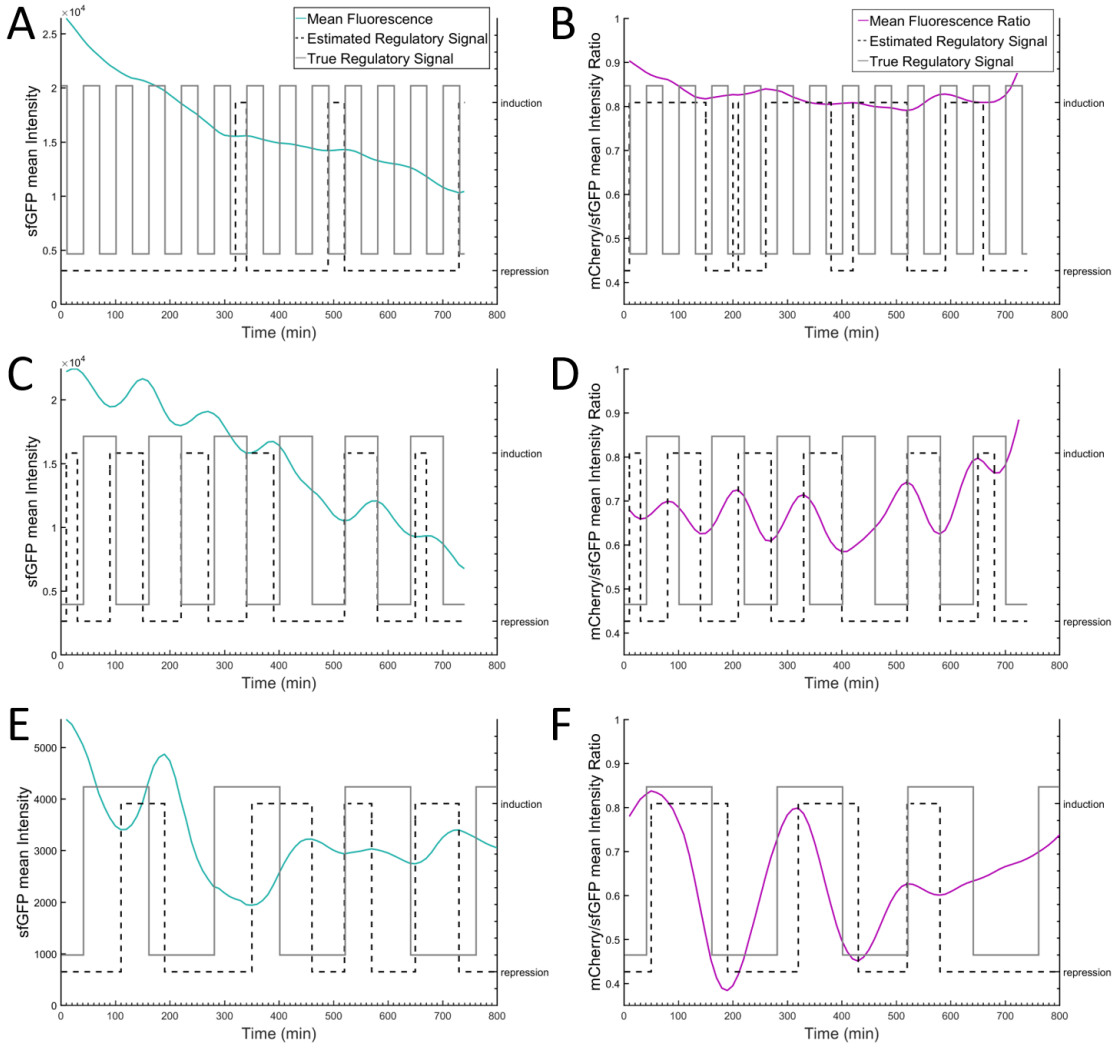

Figure 3.13: Failure to track the signal at low periods using experimental data demonstrates the impact of noise and variation on accurate dynamic signal tracking. Periodic signal reconstruction using the developed algorithm on fluorescence data of periodically regulated stable TFT expression is shown. Mean trace is calculated using cell profiles tracked using our image analysis package. The true signal is plotted using a solid grey box wave. The reconstructed signal is shown using a dotted black box wave. The following conditions are shown: A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period, E) sfGFP – 240min period, F) mCherry/sfGFP ratio – 240min period. N=90 cells (AB), 57 cells (CD), 232 cells (EF)

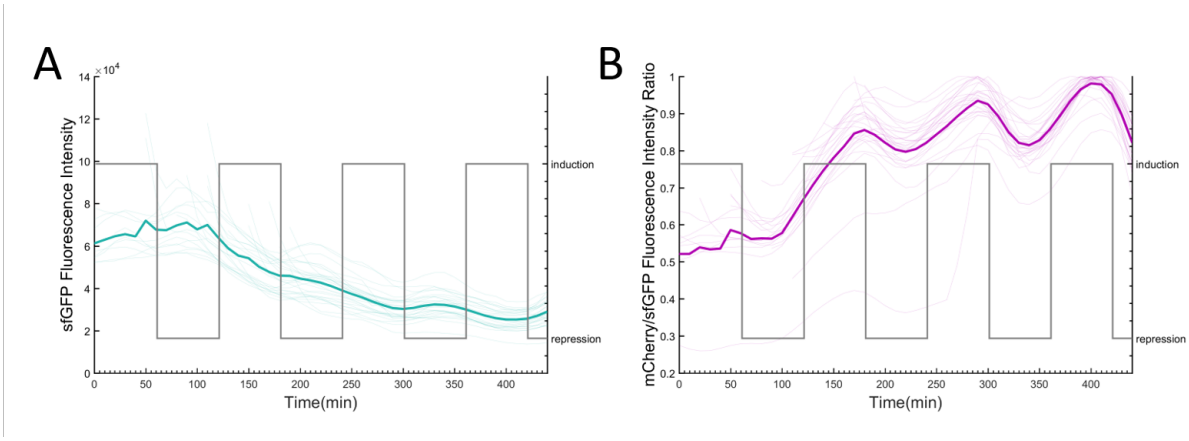

Figure 3.14: Improved segmentation and tracking leads to less variation and more complete cell profiles. Representative plots of cell profiles obtained using the improved segmentation data are shown. All plots are derived from data obtained for the stable variant of the tandem fluorescent timer. The thicker line denotes the mean trace calculated using the shown individual cell profiles. Regulatory signal change occurs at time zero. A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period. N=90 cells (AB), 57 cells (CD)

attempts at signal reconstruction were more successful, especially with recreating the shape of the signal. The detected signals are shifted, however, leading to low accuracy measures as shown in Table 3.3. All recreated signals for the 60 and 120 minute periods are less accurate than random guessing. The cause of this in the 120 minute period data is that the delay in the fluorescence response to regulatory signaling changes leads to a shift in the detected signal. Correcting for this shift using delay times determined using simulations improves the signal in certain conditions by almost 100%. This data is shown in Table 3.3.

The representative plots in Figure 3.12 show the tracked cells that survived filtering. The filtering consisted of cell profiles that contained data for at least 3 hours and was missing data from less than half the time-points. I did this to exclude very short and incomplete cell profiles, of which there were many. In addition due to the incompleteness in the filtered cell profiles, I interpolated the missing data to produce complete fluorescence traces. In certain conditions, up to 90% of detected cell profiles were filtered out using this filter, leaving < 50 tracked cell profiles from populations of over 500 cells. The reasons for this failure to segment and track cells accurately are discussed in detail in the Discussion.

3.4.2 Signal inferring using population-wide FP and TFT fluorescence data

I investigated if there was a way to ameliorate the data extracted from the images. By employing a semi-manual approach to segmentation, I was able to eliminate the problem with missing data. The cell profiles produced by tracking the manually segmented images, as shown in Figure 3.14, demonstrate that much more cell profiles are detected, and the variation is much lower. While leading to near perfect cell tracking, this method is very time-consuming and not feasible in an experiment that requires the segmentation of thousands of images.

I grouped the single-cell data into population wide data. This reduced the sparseness

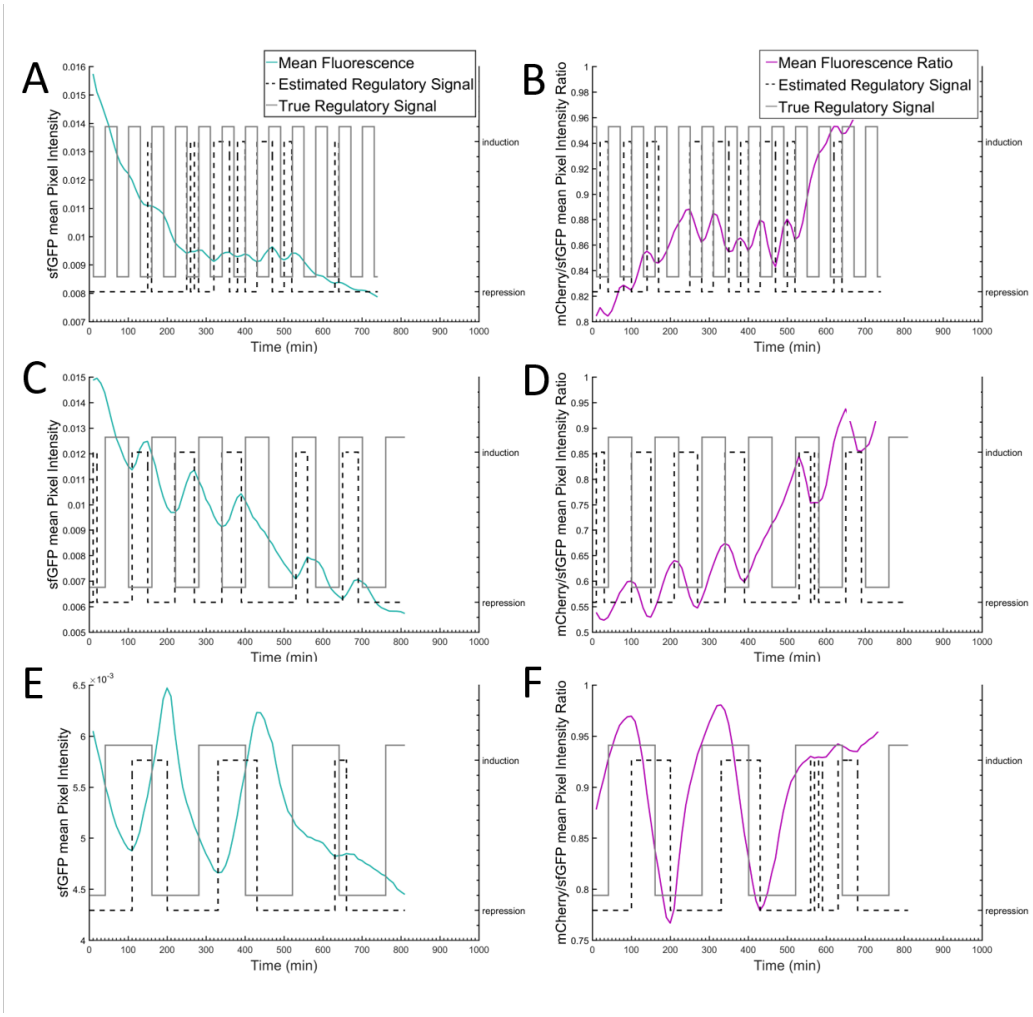

Figure 3.15: By excluding less data with population-wide fluorescence data, a wider range of TFT effectiveness is observed. Periodic signal reconstruction using the developed algorithm on population-wide data of periodically regulated stable TFT expression is shown. The mean trace is calculated using pixel intensity distributions collected from all segmented cells in our images. The true signal is plotted using a solid grey box wave. The reconstructed signal is shown using a dotted black box wave. The following conditions are shown: A) sfGFP – 60min period, B) mCherry/sfGFP ratio – 60min period, C) sfGFP – 120min period, D) mCherry/sfGFP ratio – 120min period, E) sfGFP – 240min period, F) mCherry/sfGFP ratio – 240min period. N=4142 cells (AB), 5217 cells (CD), 6119 cells (EF)

of the data but masked the cell-to-cell variation in the data. I did this by removing the cell tracking step in my image analysis pipeline, and only using the segmented images for fluorescence data extraction. By using distributions of individual pixel fluorescence intensity for all segmented cells in my images, I was able to obtain more accurate mean fluorescence traces (Figure 3.15). Most importantly, by using this new image analysis technique, a periodic fluorescence signal is now detected in the 60 minute period experiments (Figure 3.15A,B). Due to a delay in the signal, the signal accuracy is very low (< 0.5), however, with the delay corrected, the accuracy of the signal is improved (Tables 3.6, 3.5). Otherwise, there is no clear trend with regards to the accuracy of signal detection in the remaining conditions. The signal accuracy of the 240 minute period is improved using the population level data, compared to the single cell data. Otherwise, the 120 minute period population level data yielded equal or lower accuracy signal detection using my algorithm.

Table 3.5: Population-wide Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method

	sfGFP		Ratio	
Period	Unstable	Stable	Unstable	Stable
60min	0.41	0.36	0.40	0.26
120min	0.39	0.40	0.37	0.40
240min	0.62	0.55	0.64	0.56

Ultimately, it is shown that the fluorescence ratio of a stable tandem fluorescent timer is an effective alternative to conventional fluorescent proteins in the tracking of dynamic gene signaling. It is not clear, however, whether the fluorescence ratio measure of a TFT is an improvement over destabilized fluorescent proteins.

Table 3.6: Shifted Population-wide Data Signal Accuracy of the Stable and Unstable TFT at Different Periods calculated using first-derivative inference method

	sfGFP		Ratio	
Period	Unstable	Stable	Unstable	Stable
60min	0.62	0.65	0.62	0.72
120min	0.59	0.56	0.60	0.60
240min	0.88	0.81	0.87	0.87

Chapter 4

Discussion

In this thesis, I have shown that Tandem Fluorescent Timer proteins are effective reporters of dynamic gene regulation. Using a model, I showed that the mCherry/sfGFP fluorescence ratio of the TFT protein responds to changes in gene regulation and can in principal be used to detect an underlying signal. After fitting the model to experimental expression dynamics data, simulations demonstrated fluorescence behaviour that mirrored the regulatory signal closely. For the purpose of decoding the underlying signal, I developed an algorithm to detect a dynamic gene regulatory signal from a fluorescence trace or a mCherry/sfGFP fluorescence ratio trace. This algorithm enabled me to decode the underlying regulatory signal using an mCherry/sfGFP fluorescence ratio trace obtained from simulated TFT protein data, demonstrating the effectiveness of TFT ratio trace as a reporter of dynamic signalling.

4.1 Comparison of the tandem fluorescent timer to traditional fluorescent proteins

In analyzing my model, I was unable to conclude that the TFT ratio trace was a more accurate reporter due to my deterministic model not having noise. I had the options to add noise to my deterministic model or to formulate an alternative stochastic

model. The alternative I chose however, was to move on to experimental work. I was not interested in modeling the process extensively if the TFT did not work *in vivo*. Moreover, I had the means to investigate the *in vivo* effectiveness of the TFT.

I conducted time-lapse experiments using microfluidics-enabled periodic forcing to generate time-lapse fluorescence traces in several conditions. In addition to 3 different periods, I also conducted experiments using a destabilized variant of the TFT protein. This data enabled me to investigate the TFT in different conditions, and to compare *in vivo* effectiveness of stable TFT to destabilized TFT, as well as stable and destabilized sfGFP. Ultimately, this comparison is the crucial one needed to address my second hypothesis. I analyzed my data by applying my algorithm to decode the underlying periodic signal. I demonstrated the successful use of the TFT as a reporter of dynamic signalling, supporting my hypothesis that the TFT is an alternative to traditional fluorescent proteins for tracking dynamic gene expression. In addition, I present evidence that the TFT fluorescence ratio trace can detect periodic signalling of a shorter period than sfGFP, as well as more effectively detect fast periodic signalling hidden by longer time-scale fluctuations.

Given that destabilized variants of fluorescent reporters are favoured in studies where tracking of fast gene dynamics is required, my comparison of the stable and destabilized TFT is important since it also allows me to compare the stable TFT to destabilized sfGFP. Surprisingly, we observe that destabilizing the TFT protein does not confer an improved ability to track fast gene dynamics. When using the fluorescence ratio of the TFT protein to detect changes in signalling, the differential maturation kinetics of the two fluorophore subunits, mCherry and sfGFP are the critical factors. Alternatively, when using a conventional fluorescent reporter, its expression and decay kinetics are the important parameters. This difference between the TFT protein and conventional fluorescent reporters is a likely reason why changes in the stability of the TFT protein do not improve tracking effectiveness, unlike conventional fluorescent reporters. This characteristic of the TFT protein confers an advantage over conventional fluorescent

reporters since it will not suffer the pitfalls of destabilized proteins, namely the increased noise. Less noise in the fluorescence response will result in more accurate signal decoding.

4.2 Delay in fluorescence response to dynamic regulatory signalling

The algorithm I developed in this work to decode gene regulatory signalling was successful when used with ideal simulation data. When analysing simulation data from a fit model, the more realistic timescale of response to changes in expression regulation led to a reduced accuracy of the decoded signal, specifically in the phase of the periodic signal. A more significant delay between the changes in the underlying signal and the corresponding responses in the TFT fluorescence/ratio traces was observed with the fit model simulations. This delay was then observed in the periodic forcing experimental data as well, with a comparable delay time (40min). The fitting of the model was done using data from non-periodic signalling fluorescence responses, yet a similar response delay is observed in the periodic forcing experiment fluorescence traces. Given these observations, it is likely that the delay is a characteristic of the system rather than the consequence of experimental conditions. Implementing a correction for this delay improved the the accuracy of the decoded signal.

The delay observed between the regulatory signal and the fluorescence response is very long, even though the form of the signal is reported fairly accurately. Determining the factors that contribute to this delay is an important first step to minimizing it. There are obvious factors that are responsible for part of the delay like the timescale of transcription, translation and fluorophore maturation. The response time of any fluorescent reporter protein will be constrained by these three processes, however, there may be other factors that affect the response time that are characteristic of the construct used for *in vivo* testing. The GEV transcription factor, which in our construct

is bound to an transcription activation domain, enters the nucleus via active transport when it is induced by β-estradiol. There is however no active mechanism to remove the induced GEV from the nucleus, as such induced GEV can stockpile in the nucleus with periodic β-estradiol signalling due to the 40 minute washout rate [53]. Since nuclear entry and exit kinetics of the rTetR protein are faster [57], the aforementioned problem with GEV is not present in the rTetR system of our construct. Another factor is the implementation of the microfluidics plate used for periodic forcing. Before any expression of the reporter can take place, inducer-enriched media has to replace the current growth media. Furthermore, media enters the growth chamber via unidirectional flow, leading to colonies closer to the inlet, and cells closer to the outside of a colony, being preferentially exposed to inducer-enriched media.

4.3 Alternative construct designs to improve response dynamics

The nuclear exit kinetics of the GEV transcription factor are specific to the synthetic construct we developed for testing and therefore it is also a possible target for improving the response time. The options are to change the transcription factor used to one with different nuclear entry/exit kinetics, or to alter the GEV construct to increase its nuclear exit rate. Switching to a different transcription factor would also require changes to the binding sites of the dual-input promoter regulating TFT expression. Not only is this a lengthy process to implement, it would also take extra time to characterize the kinetics of the newly assembled promoter. A less labor-intensive solution is to assemble a GEV gene with a nuclear export tag [58].

Given the additional factors inherent to the construct implemented in our work to compare fluorphores, the effectiveness of the TFT may be higher when used to study endogenous processes. For example, to study stochastic switching between high and low expressing states of a protein, the TFT gene would be fused to the gene of interest

and expression dynamics would be regulated by endogenous processes. The factors that affected the delay in our work, such as nuclear GEV stockpiling and inducer diffusion in the growth chamber of the microfludics plate would be eliminated. Furthemore, this system for studying gene expression dynamics can be used to study the rate of switching, the frequency of switches, or the fraction of time spent in either state. The only measure of accuracy that would be useful is the the accuracy of the reconstructed signal shape, whereas the accuracy of the reconstructed signal phase would be mostly irrelevant.

In many applications, the phase of the decoded signal is less important than its shape. For use in these cases, a measure of accuracy that focuses on describing the decoded signal's shape would be more applicable to real-world applications. One potential solution is a method that incorporates a cross-correlation measure comparing true and decoded signals. Another solution could include the use of the Fourier transform to determine the frequency, irrespective of the phase of the signal. The disadvantage of this method is that it would only be useful when studying explicitly periodic phenomena. In the case of stochastic or unknown behaviour, the Fourier transform of the fluorescence response would only be useful in demonstrating the absence of periodicity. Furthemore, the data required for Fourier analysis is generally high resolution in time, which is difficult to obtain in this environment due to overcrowding(length of experiment) and photobleaching(imaging frequency). The first solution, using a similarity measure like cross-correlation or convolution is useful for describing accuracy in more applications: stochastic and periodic phenomena.

4.4 Complications introduced by errors in image analysis

Noise introduces a lot uncertainty into the experimental data, in contrast to the deterministic simulations of my model. While noise is expected, the poor image analysis

negatively affects the quality of the extracted data, magnifying the effect. Upon investigation, it was clear that the issue was poor cell segmentation, leading to short and incomplete cell profiles. To determine if the poor image analysis affected the tracking performance, I developed a way to compile the pixel intensity data into population wide data. While, this reduced the resolution of the data, the periodic range that I could detect was extended to include a 60 minute period. This made it obvious that the limitation of my image analysis inhibits my ability to conclude if the TFT fluorescence ratio is a more effective reporter of dynamic gene regulation.

A major issue with automated microscopy is drift in the focus over time. Autofocus systems are used by most automated microscopes to enable the fast imaging of multiple fields of view but focus drift remains a problem. Currently the only ways to combat this problem is to consistently correct the focus, which can introduce other errors, and to image at multiple z-levels as a fail-safe. Focus drift still affected the quality of my microscopy data even with my employment of these methods. Changes in focus affect cell segmentation because algorithms are designed for to detect specific bright-field intensity-based features like the bright cell outlines found in out-of-focus images. Focus drift affects the appearance of these features, making them difficult to detect. Ultimately, an improved image analysis tool for cell segmentation would be one whose performance is robust to changes in focus.

A deep learning approach to solve the problem of cell segmentation would likely lead to a robust tool. Deep learning is a machine learning method that can be used to learn features from annotated training data. By feeding training data of perfect segmentation masks associated with bright-field images at many levels of focus. A deep learning model will learn features associated with cells at different levels of focus even if they are vastly different, effectively learning to segment bright field images, regardless of focus. Deep learning has already been used to successfully develop very accurate cell segmentation tools for other microscopy images and organisms [59, 60, 61], as such it is a great approach to solving the same problem in *S. cerevisiae* bright-field microscopy

images.

4.5 Conclusion

In conclusion, during the completion of my thesis research, I have made the following contributions to the advancement of science. Using a mathematical model first, then with a synthetic construct, I demonstrated the effective use of a tandem fluorescent timer's fluorescence ratio as a reporter of dynamic gene regulation. I simulated and conducted periodic forcing experiments to compare the accuracy and range of uses between conventional fluorescent reporter sfGFP and the tandem fluorescent timer. I developed an algorithm to decode the underlying gene regulatory signal from fluorescence traces and implemented it to analyze the results of my periodic forcing experiments. I observed that the tandem fluorescent timer has a larger range in effectiveness due to a higher accuracy at lower periodic signals as well as robustness to fluctuations of much slower timescales. These contributions lay the foundation for new approaches to the measurement of dynamic gene expression and regulation. Specifically, my work demonstrates that the tandem fluorescent timer is a stable and effective reporter of dynamic gene regulation.

Bibliography

[1] Sui Huang. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23):3853–3862, 2009.

[2] Genevieve Housman, Shannon Byler, Sarah Heerboth, Karolina Lapinska, Mckenna Longacre, Nicole Snyder, and Sibaji Sarkar. Drug resistance in cancer: an overview. *Cancers*, 6(3):1769–1792, 2014.

[3] Angela Oliveira Pisco and Sui Huang. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse:'what does not kill me strengthens me'. *British journal of cancer*, 112(11):1725, 2015.

[4] Peter C Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1976.

[5] Tineke L Lenstra, Joseph Rodriguez, Huimin Chen, and Daniel R Larson. Transcription dynamics in living cells. *Annual review of biophysics*, 45:25–47, 2016.

[6] Dietmar Kültz. Molecular and evolutionary basis of the cellular stress response. *Annu. Rev. Physiol.*, 67:225–257, 2005.

[7] J Brian Robertson, Chris C Stowers, Erik Boczko, and Carl Hirschie Johnson. Real-time luminescence monitoring of cell-cycle and respiratory oscillations in yeast. *Proceedings of the National Academy of Sciences*, 105(46):17988–17993, 2008.

[8] Siddharth Sukumaran, Richard R Almon, Debra C DuBois, and William J Jusko. Circadian rhythms in gene expression: relationship to physiology, disease, drug

disposition and drug action. *Advanced drug delivery reviews*, 62(9-10):904–917, 2010.

[9] Adam Claridge-Chang, Herman Wijnen, Felix Naef, Catharine Boothroyd, Nikolaus Rajewsky, and Michael W Young. Circadian regulation of gene expression systems in the drosophila head. *Neuron*, 32(4):657–671, 2001.

[10] Katharina F Sonnen, Volker M Lauschke, Julia Uraji, Henning J Falk, Yvonne Petersen, Maja C Funk, Mathias Beaupeux, Paul François, Christoph A Merten, and Alexander Aulehla. Modulation of phase shift between wnt and notch signaling oscillations controls mesoderm segmentation. *Cell*, 172(5):1079–1090, 2018.

[11] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[12] Christopher G Kevil, Loren Walsh, F Stephen Laroux, Theodore Kalogeris, Matthew B Grisham, and JS Alexander. An improved, rapid northern protocol. *Biochemical and biophysical research communications*, 238(2):277–279, 1997.

[13] James C Alwine, David J Kemp, and George R Stark. Method for detection of specific rnas in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with dna probes. *Proceedings of the National Academy of Sciences*, 74(12):5350–5354, 1977.

[14] Graham Bell. Quantifying western blots: none more black. *BMC biology*, 14(1):116, 2016.

[15] Marcia J Holden and Lili Wang. Quantitative real-time pcr: fluorescent probe options and issues. In *Standardization and quality assurance in fluorescence measurements II*, pages 489–508. Springer, 2008.

[16] Zhong Wang, Mark Gerstein, and Michael Snyder. Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10(1):57, 2009.

[17] Pawel L. Urban. Quantitative mass spectrometry: an overview. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2079), 2016.

[18] Julien Picot, Coralie L Guerin, Caroline Le Van Kim, and Chantal M Boulanger. Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology*, 64(2):109–130, 2012.

[19] Martin Chalfie, Yuan Tu, Ghia Euskirchen, William W Ward, and Douglas C Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, 1994.

[20] Osamu Shimomura, Frank H Johnson, and Yo Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. *Journal of cellular and comparative physiology*, 59(3):223–239, 1962.

[21] Y Kohama, O Shimomura, and FH Johnson. Molecular weight of the photoprotein aequorin. *Biochemistry*, 10(22):4149–4152, 1971.

[22] Osamu Shimomura and Frank Harris Johnson. Properties of the bioluminescent protein aequorin. *Biochemistry*, 8(10):3991–3997, 1969.

[23] Osamu Shimomura and Frank H Johnson. Calcium binding, quantum yield, and emitting molecule in aequorin bioluminescence. *Nature*, 227(5265):1356, 1970.

[24] Hiroshi Morise, Osamu Shimomura, Frank H Johnson, and John Winant. Intermolecular energy transfer in the bioluminescent system of aequorea. *Biochemistry*, 13(12):2656–2662, 1974.

[25] Mans Ehrenberg. The green fluorescent protein: discovery, expression and development. *Information Department, The Royal Swedish Academy of Sciences*, 2008.

[26] Roger Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67(1):509–544, 1998.

[27] Bradley J Feilmeier, Ginger Iseminger, Diane Schroeder, Hannali Webber, and Gregory J Phillips. Green fluorescent protein functions as a reporter for protein localization in escherichia coli. *Journal of bacteriology*, 182(14):4068–4076, 2000.

[28] Marc Zimmer. Green fluorescent protein (gfp): applications, structure, and related photophysical behavior. *Chemical reviews*, 102(3):759–782, 2002.

[29] Atsushi Miyawaki and Roger Y Tsien. Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. In *Methods in enzymology*, volume 327, pages 472–500. Elsevier, 2000.

[30] Geoffrey S Baird, David A Zacharias, and Roger Y Tsien. Biochemistry, mutagenesis, and oligomerization of dsred, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences*, 97(22):11984–11989, 2000.

[31] Roger Heim and Roger Y Tsien. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current biology*, 6(2):178–182, 1996.

[32] Richard N Day and Michael W Davidson. The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews*, 38(10):2887–2921, 2009.

[33] Brendan P Cormack, Gwyneth Bertram, Mark Egerton, Neil AR Gow, Stanley Falkow, and Alistair JP Brown. Yeast-enhanced green fluorescent protein (yegfp): a reporter of gene expression in candida albicans. *Microbiology*, 143(2):303–311, 1997.

[34] Roger Heim. Improved green fluorescence. *Nature*, 373:663–664, 1995.

[35] Brendan P Cormack, Raphael H Valdivia, and Stanley Falkow. Facs-optimized mutants of the green fluorescent protein (gfp). *Gene*, 173(1):33–38, 1996.

[36] Xianqiang Li, Xiaoning Zhao, Yu Fang, Xin Jiang, Tommy Duong, Connie Fan, Chiao-Chain Huang, and Steven R Kain. Generation of destabilized green fluorescent protein as a transcription reporter. *Journal of Biological Chemistry*, 273(52):34970–34975, 1998.

[37] Carolina Mateus and Simon V Avery. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast*, 16(14):1313–1323, 2000.

[38] Arda A Green and William D McElroy. Crystalline firefly luciferase. *Biochimica et biophysica acta*, 20:170–176, 1956.

[39] Dmitriy M Chudakov, Mikhail V Matz, Sergey Lukyanov, and Konstantin A Lukyanov. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological reviews*, 90(3):1103–1163, 2010.

[40] Anyimilehidi Mazo-Vargas, Heungwon Park, Mert Aydin, and Nicolas E Buchler. Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy. *Molecular biology of the cell*, 25(22):3699–3708, 2014.

[41] Richard I Morimoto. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes & development*, 22(11):1427–1438, 2008.

[42] Julie M Pratt, June Petty, Isabel Riba-Garcia, Duncan HL Robertson, Simon J Gaskell, Stephen G Oliver, and Robert J Beynon. Dynamics of protein turnover, a missing dimension in proteomics. *Molecular & Cellular Proteomics*, 1(8):579–591, 2002.

[43] Lijuan Zhang, Nadya G Gurskaya, Ekaterina M Merzlyak, Dmitry B Staroverov, Nikolay N Mudrik, Olga N Samarkina, Leonid M Vinokurov, Sergey Lukyanov, and Konstantin A Lukyanov. Method for real-time monitoring of protein degradation at the single cell level. *Biotechniques*, 42(4):446–450, 2007.

[44] Alexey Terskikh, Arkady Fradkov, Galina Ermakova, Andrey Zaraisky, Patrick Tan, Andrey V Kajava, Xiaoning Zhao, Sergey Lukyanov, Mikhail Matz, Stuart Kim, et al. " fluorescent timer": protein that changes color with time. *Science*, 290(5496):1585–1588, 2000.

[45] Anton Khmelinskii, Philipp J Keller, Anna Bartosik, Matthias Meurer, Joseph D Barry, Balca R Mardin, Andreas Kaufmann, Susanne Trautmann, Malte Wachsmuth, Gislene Pereira, et al. Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nature biotechnology*, 30(7):708, 2012.

[46] Alexander Varshavsky. The n-end rule pathway of protein degradation. *Genes to cells*, 2(1):13–28, 1997.

[47] Afnan Azizi, Wilson Lam, Hilary Phenix, Lioudmila Tepliakova, Ian J Roney, Daniel Jedrysiak, Alex Power, Vaibhav Gupta, Nada Elnour, Martin Hanzel, et al. No training required: experimental tests support homology-based dna assembly as a best practice in synthetic biology. *Journal of biological engineering*, 9(1):8, 2015.

[48] Robert K Mortimer and John R Johnston. Genealogy of principal strains of the yeast genetic stock center. *Genetics*, 113(1):35–43, 1986.

[49] R Daniel Gietz and Robert H Schiestl. High-efficiency yeast transformation using the liac/ss carrier dna/peg method. *Nature protocols*, 2(1):31, 2007.

[50] Marketa Ricicova, Mani Hamidi, Adam Quiring, Antti Niemistö, Eldon Emberly, and Carl L Hansen. Dissecting genealogy and cell cycle as sources of cell-to-cell variability in mapk signaling using high-throughput lineage tracking. *Proceedings of the National Academy of Sciences*, 110(28):11403–11408, 2013.

[51] Manfred Gossen and Hermann Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, 89(12):5547–5551, 1992.

[52] Ivan Sadowski, Jun Ma, Steve Triezenberg, and Mark Ptashne. Gal4-vp16 is an unusually potent transcriptional activator. *Nature*, 335(6190):563, 1988.

[53] R Scott McIsaac, Sanford J Silverman, Megan N McClean, Patrick A Gibney, Joanna Macinskas, Mark J Hickman, Allegra A Petti, and David Botstein. Fast-acting and nearly gratuitous induction of gene expression and protein depletion in saccharomyces cerevisiae. *Molecular biology of the cell*, 22(22):4447–4459, 2011.

[54] Manfred Gossen, Sabine Freundlieb, Gabriele Bender, Gerhard Muller, Wolfgang Hillen, and Hermann Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218):1766–1769, 1995.

[55] David Kadosh and Kevin Struhl. Targeted recruitment of the sin3-rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Molecular and cellular biology*, 18(9):5121–5127, 1998.

[56] Simone Ugolini and Carlo V Bruschi. The red/white colony color assay in the yeast saccharomyces cerevisiae: epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells. *Current genetics*, 30(6):485–492, 1996.

[57] Manfred Argast and Christoph F Beck. Tetracycline diffusion through phospholipid bilayers and binding to phospholipids. *Antimicrobial agents and chemotherapy*, 26(2):263–265, 1984.

[58] Shunichi Kosugi, Masako Hasebe, Masaru Tomita, and Hiroshi Yanagawa. Nuclear export signal consensus sequences defined using a localization-based yeast selection system. *Traffic*, 9(12):2053–2062, 2008.

[59] David A. Van Valen, Takamasa Kudo, Keara M. Lane, Derek N. Macklin, Nicolas T. Quach, Mialy M. DeFelice, Inbal Maayan, Yu Tanouchi, Euan A. Ashley, and Markus W. Covert. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Computational Biology*, 12(11):1–24, 11 2016.

[60] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[61] Mo Zhang, Xiang Li, Mengjia Xu, and Quanzheng Li. Image segmentation and classification for sickle cell disease using deformable U-Net. *arXiv preprint arXiv:1710.08149*, 2017.