

TUNING GENE EXPRESSION IN MAMMALIAN CELLS BY TRANSCRIPTIONAL AND POST- TRANSCRIPTIONAL CONTROL

Annalise Zouein

Submitted for the Degree of Doctor of Philosophy

Imperial College London, Department of Bioengineering

Supervised by Professor Tom Ellis & Dr Francesca Ceroni

JANUARY 16, 2024

ABSTRACT

Engineering biology also known as synthetic biology is rapidly advancing, offering numerous genetic tools for the engineering and reprogramming of cells. These tools enhance cell performance, enabling novel functions, with a wide range of potential applications. The future of cell engineering would benefit from strategies that facilitate precise and tuneable modulation of gene expression without permanent genetic alteration or a large resource footprint. In this direction, we explore the use of microRNAs (miRNAs) and Transcription Factor Binding Sites (TFBS) to control gene expression at transcriptional and post-transcriptional levels. Inspired by favourable native characteristics that ensure tight regulation within cells, we develop a highly tuneable, modular and programmable platform for the control of gene expression in mammalian cells.

The developed miRNAs and TF decoy tools show promise in modulating gene expression in mammalian cells. The use of TF decoys, empowered by our in-house DNA assembly method, led to the sequestration of intracellular TFs, subsequently causing changes in gene expression and cellular responses. On the other hand, sRNAs such as miRNAs were used to alter transcriptional and post-transcriptional outputs of genes showing their versatility. Further development may be required to fully harness their potential for gene expression modulation in mammalian cells.

This work contributes to the growing field of biology-inspired cell engineering, providing insights into the precise tuning of gene expression levels in mammalian cells. The development of these tools opens possibilities for studying diseases, cellular reprogramming for cell therapies and the production of relevant recombinant products.

DECLARATION

I declare that the following work results from my own independent and original research.

Any work from other sources has been acknowledged.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following individuals and organizations who have supported me throughout my doctoral research at the Department of Bioengineering and Chemical Engineering, Imperial College London. I would also like to thank the Centre for Synthetic Biology at Imperial College London for providing the necessary resources and infrastructure that facilitated a significant portion of the laboratory work essential to this research. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the EPSRC Centre for Doctoral Training in BioDesign Engineering (EP/S022856/1), which provided funding and learning opportunities.

I am sincerely thankful to my supervisors Professor Tom Ellis and Dr Francesca Ceroni for their guidance, support, and insights throughout the course of my research. I am grateful to have had the opportunity to work and learn from you. Moreover, I would like to thank Dr Mark Wigglesworth (formerly AstraZeneca, now Alderley Lighthouse Labs) for early input and consumables funding.

I am grateful to my lab colleagues for their collaborative spirit, camaraderie, and support. I would especially like to thank Roberto and Carolina who made every day in the lab enjoyable. Additionally, I am thankful to Elisenda and Gabriela, who were always willing to lend an ear and provide support when needed.

On a personal note, I am indebted to my parents and sister, whose endless encouragement and sacrifices have been the cornerstone of my journey. Last but certainly not least, I also extend my heartfelt thanks to my partner, Jonathan, for his understanding, patience, and unwavering support during the challenging phases of my doctoral studies. I could not have done it without you.

CONTENTS

1	Introduction	12
1.1	Summary.....	12
1.2	Background	13
1.2.1	Engineering Biology	13
1.2.2	Engineering Along The Central Dogma	17
1.2.3	Resource Aware Engineering.....	36
1.2.4	Ethical Considerations of Engineering Biology.....	38
1.2.5	Next-Generation Cell Engineering Tools.....	39
1.3	Aims & Objectives.....	43
1.4	Note to the reader	45
2	Assembly of Hard to Clone Repetitive Sequences.....	46
2.1	Introduction.....	46
2.2	Results and Discussion	49
2.2.1	Library-based approaches to repeat assembly.	50
2.2.2	Iterative approach to repeat assembly: Development of CTRL	60
2.3	Conclusion	67
3	TF Decoys AS Competitive Inhibitors Of Gene Expression.....	69
3.1	Introduction.....	69
3.2	Results and Discussion	74
3.2.1	Design of TF Decoys for Mammalian Cells.....	74
3.2.2	Sequestering Orthogonal TFs in Mammalian Cells.....	75
3.2.3	Sequestering Native TFs in Mammalian Cells	88
3.3	Conclusion	95
4	Modifying Gene Expression Using miRNAs	98
4.1	Introduction.....	98
4.2	Results and Discussion	101
4.2.1	Reducing expression of target genes.....	101
4.2.2	Increasing gene expression of target genes. miRNA activation (miRa).....	115
4.3	Conclusion	119
5	Discussion	121

5.1	Summary	121
5.2	Gene Expression Engineering.....	122
5.2.1	TF Decoy Plasmids for Transcriptional Control.....	123
5.2.2	miRNAs for Transcriptional and Post-Transcriptional Control:	125
5.2.3	Comparison Between Gene Expression Control Methods.....	128
5.3	Current Limitations.....	132
5.4	Future Directions.....	135
5.4.1	Improving Current Methodologies and Tools	135
5.4.2	Applications in Gene Expression Engineering	137
5.4.3	Applications Beyond Gene Expression Engineering	138
5.5	Conclusions	140
6	Materials and methods.....	141
6.1	Cell Culture.....	141
6.1.1	Mammalian Cell Culture	141
6.1.2	E. coli DH5 α	142
6.1.3	S. cerevisiae BY4741 and VL6-48N	143
6.2	Molecular Biology (Cloning).....	143
6.2.1	DNA Extraction.....	143
6.2.2	Restriction Digestion.....	145
6.2.3	PCR.....	146
6.2.4	RCA.....	147
6.2.5	Golden Gate Reactions	148
6.2.6	Gibson Assembly	148
6.3	DNA Transfections.....	148
6.4	Data collection and analysis	149
6.4.1	Fluorescent Microscopy.....	149
6.4.2	Flow Cytometry	149
6.4.3	Data Analysis	151
6.5	List of Materials Used.....	153
6.5.1	Table of primers:	153
6.5.2	Table of g-blocks:	156

6.5.3	Table of plasmids:.....	157
7	Supplementary Information	160
7.1	Supplementary Data	160
7.2	Statistical Test Summary	163
7.2.1	Figure 3.3	163
7.2.2	FIGURE 3.4.....	163
7.2.3	FIGURE 3.5.....	165
7.2.4	Figure 3.8	167
8	Bibliography.....	170

LIST OF TABLES

TABLE 2.1: <i>DIFFERENT DNA REPEAT TYPES</i>	48
TABLE 2.2: <i>SUMMARY OF NON-TEMPLATE PCR PRIMERS AND CONDITIONS USED.</i>	58
TABLE 2.3: <i>SUMMARY OF 4 BASE PAIR OVERHANGS REVEALED DURING CTRL.</i>	62
TABLE 2.4: <i>CTRL CLONING RESULTS.</i>	65
TABLE 3.1: <i>NEUTRAL SPACER DNA SEQUENCES.</i>	75
TABLE 3.2: <i>EFFECT OF TFBS ADDED IN SERIES.</i>	82
TABLE 3.3: <i>TESTING CUMULATIVE EFFECT OF USING MORE PLASMID IN TRANSFECTIONS.</i>	83
TABLE 6.1: <i>PCR CONDITIONS SET UP USING NEB PHUSION POLYMERASE.</i>	146
TABLE 6.2: <i>PCR CONDITIONS FOR COLONY PCR USING TAQ POLYMERASE.</i>	146
TABLE 6.3: <i>TEMPERATURE CYCLER CONDITIONS FOR PCR</i>	147
TABLE 6.4: <i>ATTUNE FLOW CYTOMETRY SETTINGS.</i>	150
TABLE 7.1: <i>MANN WHITNEY U TEST COMPONENTS OF FIGURE 3.3.B</i>	163
TABLE 7.2: <i>MANN WHITNEY U TEST OF FIGURE 3.4</i>	163
TABLE 7.3: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.4.A GRAPH TITLE IS 1X TET DECOY</i>	164
TABLE 7.4: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.4.A GRAPH TITLE IS 2X TET DECOY</i>	164
TABLE 7.5: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.4.A. GRAPH TITLE IS 4X TET DECOY</i>	165
TABLE 7.6: <i>MANN WHITNEY U TEST OF FIGURE 3.5</i>	165
TABLE 7.7: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.5.A. GRAPH TITLE IS 1X TET DECOY</i>	166
TABLE 7.8: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.5.A. GRAPH TITLE IS 2X TET DECOY.</i>	166
TABLE 7.9: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.5.A. GRAPH TITLE IS 4X TET DECOY</i>	167
TABLE 7.10: <i>MANN WHITNEY U TEST OF FIGURE 3.8.C.D. E</i>	167
TABLE 7.11: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.8.C HEK.</i>	168
TABLE 7.12: <i>DESCRIPTIVE STATISTICS FOR FIGURE 3.8.C CHO.</i>	169

LIST OF FIGURES

FIGURE 1.1: <i>CELLS CAN BE ENGINEERED FROM THE BOTTOM UP OR THE TOP-DOWN.</i>	15
FIGURE 1.2: <i>ENGINEERED REGULATION CAN OCCUR ACROSS THE CENTRAL DOGMA</i>	18
FIGURE 1.3: <i>CHARACTERISTICS THAT DETERMINE THE PERFORMANCE OF A CELLULAR CIRCUIT.</i>	20
FIGURE 1.4: <i>METHODS OF PERMANENT OR TEMPORARY DNA REGULATION.</i>	21
FIGURE 1.5: <i>MAMMALIAN TRANSCRIPTIONAL REGULATION USING TET-BASED SYSTEMS.</i>	25
FIGURE 1.6: <i>METHODS OF RNA REGULATION.</i>	31
FIGURE 1.7: <i>SINGLE-LAYER VS MULTI-LAYER CIRCUITS.</i>	35
FIGURE 1.8: <i>DESIGN OF MAMMALIAN RESOURCE COMPETITION SENSOR</i>	38
FIGURE 1.9: <i>TFs AND MIRNAS HAVE SIMILAR FUNCTIONS IN GENE REGULATION.</i>	41
FIGURE 2.1: <i>COMMON MOLECULAR BIOLOGY WORKFLOW FOR THE ASSEMBLY OF DESIRED DNA SEQUENCES FOR ENGINEERING BIOLOGY.</i>	49
FIGURE 2.2: <i>DEPICTION OF KEY STEPS OF REPEAT GENERATION USING ROLLING CIRCLE AMPLIFICATION (RCA) AND TRANSFORMATION ASSISTED RECOMBINATION (TAR) IN YEAST.</i>	51
FIGURE 2.3: <i>RCA TAR CLONING KEY STAGES AND ANALYSIS.</i>	53
FIGURE 2.4: <i>DEPICTION OF STEPS AND EXPECTED RESULTS OF REPEAT PLASMID ASSEMBLY USING REDUNDANT OLIGONUCLEOTIDES.</i>	55
FIGURE 2.5: <i>AGAROSE GEL IMAGE SHOWING RESULTING SIZES OF LIGATED OLIGOS LIGATED FOR DIFFERENT DURATIONS AND CONCENTRATIONS TO INCREASE THE PROPORTION OF LINEAR CONCATEMER PRODUCTS OVER CIRCULARISED CONCATEMER PRODUCTS. EACH SET CONTAINS LIGATION PRODUCTS OF REACTIONS CONTAINING 0.2μG, 0.4μG OR 1μG OF LIGATED OLIGOS IN SUCCESSION.</i>	56
FIGURE 2.6: <i>LIBRARY-BASED APPROACHES TO THE ASSEMBLY OF REPETITIVE SEQUENCES USING NON-TEMPLATE PCR.</i>	58
FIGURE 2.7: <i>CTRL ALLOWS ASSEMBLY OF REPEATS USING GOLDEN GATE AND BIOBRICK ITERATIVE TECHNIQUES.</i>	61
FIGURE 2.8: <i>DOMESTICATION OF REPEATS INTO CTRL.</i>	62

FIGURE 2.9: CTRL GOLDEN GATE FUNCTIONALITY.	63
FIGURE 2.10: USE OF THE BIOBRICK FUNCTION OF CTRL	64
FIGURE 2.11: RESTRICTION DIGEST CONFIRMING THE EXPANSION OF A REPEAT FROM 16 SITES TO 256 SITES.	66
FIGURE 2.12: TYPICAL IMPLEMENTATION OF CTRL WORKFLOW.	68
FIGURE 3.1: PRINCIPLES BEHIND FUNCTIONS OF TANDEM REPEATS.	70
FIGURE 3.2: USING TF DECOYS IN MAMMALIAN CELLS TO ASSESS CAPACITY TO COMPETITIVELY INHIBIT GENE EXPRESSION FROM CIRCUITS	73
FIGURE 3.3: TESTING PRINCIPLE OF INTRACELLULAR TF COMPETITION.	77
FIGURE 3.4: COMPILED ANALYSIS OF SAMPLES FROM AT LEAST THREE BIOLOGICAL REPLICATES ACQUIRED BY FLOW CYTOMETRY OF CELLS CO-TRANSFECTED WITH pTet-EGFP AND TF DECOY PLASMIDS CONTAINING 0,7,21 OR 320 TET O SITES IN CHO-K1.	79
FIGURE 3.5: COMPILED ANALYSIS OF SAMPLES FROM AT LEAST THREE BIOLOGICAL REPLICATES ACQUIRED BY FLOW CYTOMETRY OF CELLS CO-TRANSFECTED WITH pTet-EGFP AND TF DECOY PLASMIDS CONTAINING 0,7,21 OR 320 TET O SITES IN HEK-293T.	80
FIGURE 3.6: FREE TF EFFECT ON AUTOREGULATORY CIRCUITS	85
FIGURE 3.7: EFFECT OF TF DECOY PLASMIDS AND NFL (LINEARISER) CO-TRANSFECTION ON CELLS	87
FIGURE 3.8: REGULATING NATIVE GENE REGULATION OF NRF2 USING DECOY PLASMIDS.	90
FIGURE 4.1: EXPRESSION FROM EUKARYOTIC TRANSCRIPTIONAL UNIT.	99
FIGURE 4.2: DESIGN OF THE PROPOSED miRNA SENSOR.	101
FIGURE 4.3: IDENTIFYING miRNA EFFECTS	104
FIGURE 4.4: EXPERIMENTAL WORKFLOW FOLLOWED TO EXPLORE miRNA SENSOR SENSITIVITY	106
FIGURE 4.5: RESULTS OF A SENSOR WITH MREs IN 3' UTR.	109
FIGURE 4.6: HEATMAP OF FOLD CHANGE BETWEEN miRNA EXPRESSION AND CONTROL CONDITIONS OF TWO miRNA SENSORS WITH TWO EXPRESSION LEVELS.	110

FIGURE 4.7: <i>FLOWJO-GENERATED GRAPHS ILLUSTRATING THE TRANSFECTION LANDSCAPE. ...</i>	112
FIGURE 4.8: <i>ASSESSING MIRNA SENSOR SENSITIVITY TO CHANGES IN MIRNA OF INTEREST USING POLY-TRANSFECTION.</i>	114
FIGURE 4.9: <i>ENGINEERING GENE ACTIVATION USING MIRNA INDUCERS.</i>	117
FIGURE 5.1: <i>THERE IS AN INCREASING INTEREST IN TOOLS FOR ENGINEERING GENE REGULATION</i>	122
FIGURE 5.2: <i>ORIGINS, SIZE AND SCHEMATIC REPRESENTATION OF TRANSCRIPTIONAL REGULATORS DESCRIBED OR TESTED IN THIS THESIS (RIGHT).</i>	129
FIGURE 5.3: <i>DEPICTION AND EXPECTED RESPONSE CURVE OF A POTENTIAL METHOD OF INCREASING SIGNAL FROM MIRNA ACTIVATION CIRCUITS USING PFLS.</i>	134
FIGURE 6.1: <i>FLOW ANALYSIS USING FLOWJO.</i>	152

1 INTRODUCTION

1.1 SUMMARY

Engineering biology is a rapidly growing field with significant impact and broad applications. This discipline involves the integration of engineered genetic components to create engineered cells with novel functions. The engineered components can regulate DNA, RNA, or protein, each offering unique behaviours and functions such as speed of response, inheritability, and varying degrees of resource load on cells. The choice of intervention level depends on the end goal and can use single to multiple regulatory tools, integrated into increasingly complex circuits. However, the continuous addition of complexity presents challenges such as the need for tools to be 'resource-aware'. This means they should be designed to minimize their impact on the system's resources, preventing negative interference. Additionally, the delivery of large, complex systems remains a difficult and inefficient process in many mammalian cell types. As such, next-generation tools need to improve on existing ones by addressing these challenges. A promising area of focus in this field is the engineering of miRNAs (micro RNAs) and TFs (Transcription factors). These elements offer many benefits making them strategic points of intervention. This includes the ability to mediate several cellular functions, their ability to be engineered, and their multiplexing capacity. We will explore these aspects in detail in this chapter, shedding light on how they can be used to navigate the challenges of conventional gene engineering.

1.2 BACKGROUND

1.2.1 ENGINEERING BIOLOGY

Engineering biology is a rapidly evolving field with significant impacts across various sectors. Given its potential, it is not surprising that numerous governments, including the UK, are actively expanding their infrastructure and investments to translate groundbreaking research in engineering biology into tangible benefits for society (The Council for Science and Technology, 2023).

The promise of engineering biology stems from its dedication to applying engineering principles to biological systems. These principles include iteration, efficiency, simplicity, reliability and predictability. The field's notable achievements often stem from the application of these principles through various strategies and methodologies.

A key focus in engineering biology is enhancing the predictability and reliability of developed technologies. This ensures that different labs can use these tools as modules to answer their research questions (Arkin, 2013). But this is more challenging in practice as engineered components interact with, or modify, the systems of host cells they are introduced into. They may also rely on the host cell's machinery for the production of some of their own components. These interactions and roles within the complex biological framework might not be straightforward to predict (Cardinale, Joachimiak & Arkin, 2013). To address this complexity, 'omic' approaches have been employed. These approaches aim to characterize and quantify groups of biomolecules (for example: DNA, RNA, protein) within cells, providing a more comprehensive understanding of the biological settings in the presence or absence of engineered components (Schwanhäusser et al., 2011; Amer & Baidoo, 2021). For instance, the use of mammalian cells as factories for the production of biopharmaceutical proteins has only been improved with the deeper understanding offered by omics approaches, identifying bottlenecks that inform the direction in which engineered mammalian cell factories can be made with enhanced production of proteins of interest (Torres et al., 2021).

Conversely, some researchers have adopted the opposite approach, attempting to isolate engineered components from host systems to improve the predictability of responses (Liu et al., 2018; Stanton et al., 2014).

Another principle often utilized in engineering biology is iteration, which increases the throughput of approaches and thus accelerates discovery (Wu et al., 2019; Gan et al., 2022). In line with this, the design of tools and equipment for automation across several engineering biology processes promises efficient production of iterations (Appleton et al., 2017).

Lastly, another application of engineering principles, much like traditional engineering approaches, models biological systems. For these purposes, a minimal set of components and parameters are used to simulate and capture the complexity of a system, as well as understand and predict system functionality. This principle where the simplest model is often the best is called Occam's razor (Gross, 2019). An example of this is described by Professor McFadden, where overcomplication of the components of metabolic pathways may attribute the noise to inactive pathways, resulting in inaccurate representation in models (McFadden, 2023).

Engineered tools and methods are continuously being developed to achieve ever more sophisticated ways of handling cells and their components. Their use spans applications, from therapeutics to recombinant protein manufacturing or for the study of biological processes (Khalil & Collins, 2010). Tools designed by scientists are applied to create engineered cells from the top-down or bottom up.

1.2.1.1 Engineering from the Top-down vs Bottom-up

Engineered cells can be obtained by two main approaches: top-down or bottom-up. Both approaches use engineered biological components to design and construct genetic circuits for programming cells with novel functions (**Figure 1.1.a.**)

Top-down engineering redesigns biological systems to give them additional capabilities (Fritz et al., 2010). The resulting gene and protein content of the derived engineered cell maintains similarity to the parental one. Therefore, the use of this approach is particularly beneficial when host cell functions and contents are required to achieve objectives. A common application of top-down methods engineers along signal transduction pathways (Krawczyk et al., 2020; Baeumler, Ahmed & Fulga, 2017; Dagar et al., 2023; Manhas et al., 2022). Signalling pathways transmit messages within cells in response to environmental cues, to enhance the cells' ability to adapt and respond to their surroundings. Generalised Engineered Activation Regulators (GEARs) have been engineered to rewire native cell signalling, transmitting cellular messages to genomic targets of choice. This allows the activation of gene expression to different cues than

normal, changing the set of cellular responses usually observed to a signal (Krawczyk et al., 2020).

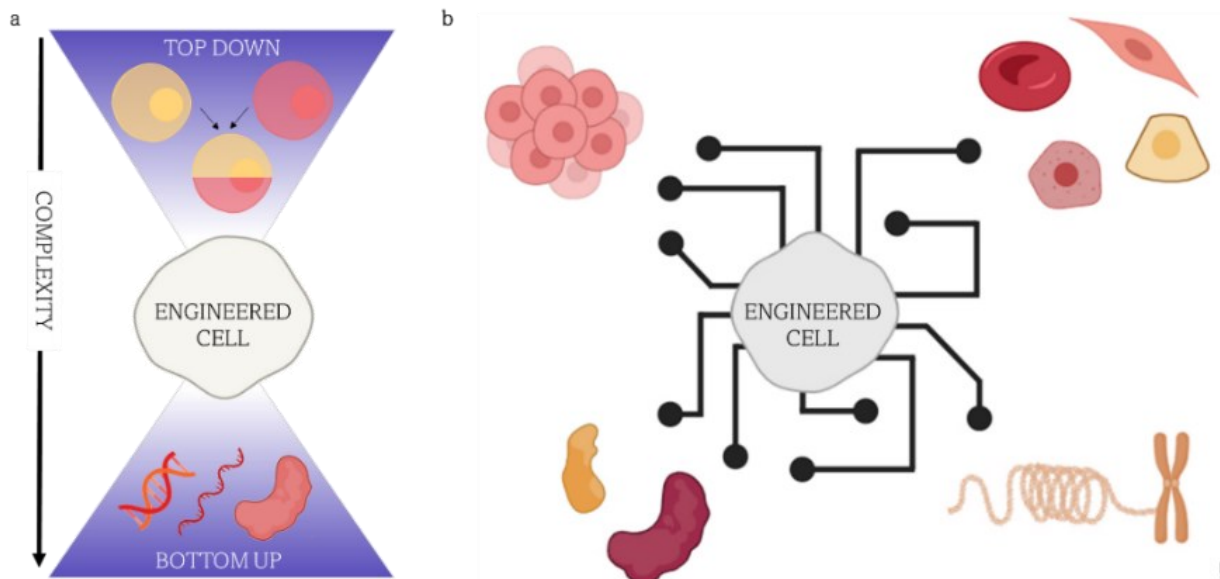


Figure 1.1: Cells can be engineered from the bottom up or the top-down. **a.** Cells can be engineered from the bottom up from non-living purified biomolecules or the top-down by introducing new instructions or functions into living cells. **b.** Mammalian engineered cells can be manipulated to make use of their unique intrinsic features like multicellularity (top-left), differentiation states (top-right), complex protein expression (bottom-left) and genome organisation (bottom-right). Some images used from BioRender.com.

Another approach that uses the host cell's natural functions is lab-based evolution which uses cellular capacity for adaptation. In these strategies, cells are encouraged to adapt to new conditions at a faster rate by evolving components, pathways or responses. The most efficient of these methods, called continuous directed evolution occurs continuously with limited researcher intervention to achieve optimal variants (Morrison, Podracky & Liu, 2020). While many methods use unicellular chassis such as bacteria and yeast, the need for methods with higher-order chassis such as mammalian cells has become apparent for the development of biomolecules for some applications, especially those for specific functions in mammalian cells (Hendel & Shoulders, 2021). This shows how top-down biology considers different biological contexts choosing the most suitable to achieve goals most efficiently.

On the other hand, bottom-up synthetic biology often assembles non-living components of cells (from natural or synthetic origins) into artificial cells (Elani, 2021; Xu, Hu & Chen, 2016) (**Figure 1.1.a**). This involves assembling the minimal components required for limited cell-like functionalities usually encapsulated by a phospholipid bilayer. The resulting engineered artificial

cells are non-living, only containing the components required for their proposed specialized function. For example, bottom-up engineers have encapsulated genetic instructions along with ingredients required for transcription and translation for the expression of therapeutic proteins. These have been functionalized to respond to several types of stimuli and have promising applications for therapeutics (Krinsky et al., 2018). Others have used the functions of actin, a component of the cell's cytoskeleton required for motility and contraction, to induce artificial contraction in artificial cells (Jahnke et al., 2020). Therefore, artificial cells are engineered to mimic many cellular functions, but achieving multipurpose cells containing more than one function remains a major challenge (Elani, 2021).

The future of bottom-up engineering looks promising with some teams looking at integrating top-down and bottom-up engineered cells. One study harnesses the bacterium *E. coli*'s (*Escherichia coli*) superior capacity to express and accurately insert proteins into membranes and synthetic cells' capacity for delivery of biomolecules. This was such that proton gradients established in *E. coli* were used to induce pH-driven shape change in synthetic cells which can be used for release of cargo from the synthetic compartments (Jahnke et al., 2021). This displays how the strengths of each strategy can be used mutually to advance the complexity with which synthetic cells can be used.

As mentioned previously, the predictability of systems is often sought after. The use of bottom-up engineered cells with reduced complexity combining a few well-characterised components and interactions, encourages and prioritises predictability in the synthetic compartment, an important requirement for many studies and applications (Buddingh' & van Hest, 2017). However, for applications where one wants to maintain and complement processes in cells like replication, growth, development, homeostasis, energy production and adaptability, bottom-up strategies may not be suitable. Top-down approaches benefit from maintaining intricate relationships between cells, their components, and their environment. The user must decide on the best host for their tasks.

The selection of a cell chassis or host can be influenced by the specific requirements of the intended applications (Morse & Alper, 2016). Mammalian cells, in particular, offer unique advantages that make them an ideal choice for understanding cell function, studying disease and the effects of therapeutics on cellular regulatory mechanisms and responses (Mathur, Xiang & Smolke, 2017; Lienert et al., 2014; Kojima & Fussenegger, 2019; Beitz, Oakes & Galloway, 2022). For example, genetically modified cell therapeutics, where cells are sourced from patients, engineered, and reintroduced as therapies, have proven a promising method of treating diseases such as cancer. A successful example of these is engineered chimeric antigen

receptors (CARs) on T-cells, which are engineered to respond to specific cues, also referred to as antigens, on cancer cells and feed into native pathways to elicit T-cell immunogenic responses (Dagar et al., 2023). These cell therapeutics are now in use in blood cancers with many looking to adapt this strategy for use in solid tumours (Martinez & Moon, 2019; Cappell & Kochenderfer, 2023).

Another condition in which using mammalian chassis is superior is in the engineering and production of mammalian proteins. This is because the bioproduction of protein-based products often requires the correct folding and addition of post-translational modifications (PTMs), a feature readily available in mammalian cells. However, mammalian cell growth is slower, and experiments are more costly to perform, causing groups to try to use other cells and organisms to circumvent this, often at the expense of protein complexity (Zang et al., 2021). Furthermore, when an application calls for the exploitation of features inherent to multicellular organisms, such as cell type differentiation, tissue engineering, or cell-to-cell communication, mammalian cells are often the preferred choice (Makri Pistikou et al., 2023; Wang, Beitz & Galloway, 2020; Glykofrydis & Elfick, 2022). This versatility underscores the importance of mammalian cells as a chassis in synthetic biology and their continuous use even when other hosts might offer simpler or more cost-effective alternatives (**Figure 1.1.b**). The following section will explore the wide range of strategies employed to redesign mammalian cells to fulfil objectives.

1.2.2 ENGINEERING ALONG THE CENTRAL DOGMA

The interest in engineering mammalian cells for applications has given rise to a multitude of technologies that redesign cells for adjusted performance. The molecular tools can impose functions at the different levels of biological organisation (**Figure 1.2**). These methodologies can intervene at the DNA level to regulate transcription, manipulate RNAs post-transcription, or adjust the RNA quantity required for protein translation. Furthermore, proteins can be regulated after translation by altering the amino acids they are made of or engineering their interactions. This demonstrates the comprehensive reach of regulatory tools, where regulation of each type of biomolecule (DNA, RNA or protein) has different pros and cons to its implementation. The abundance of tools that can be used, either individually or together, enables control over gene expression and improves the fine-tuning of cellular systems (He, 2019).

Instructions for engineering such changes are typically written into DNA plasmids and introduced into mammalian cells by transfection. Plasmids can be designed so that gene expression occurs in the nucleus over a transient period of days, be maintained by cells as

episomes for longer periods, or DNA can integrate into the host genome by various mechanisms for permanent expression (Fus-Kujawa et al., 2021).

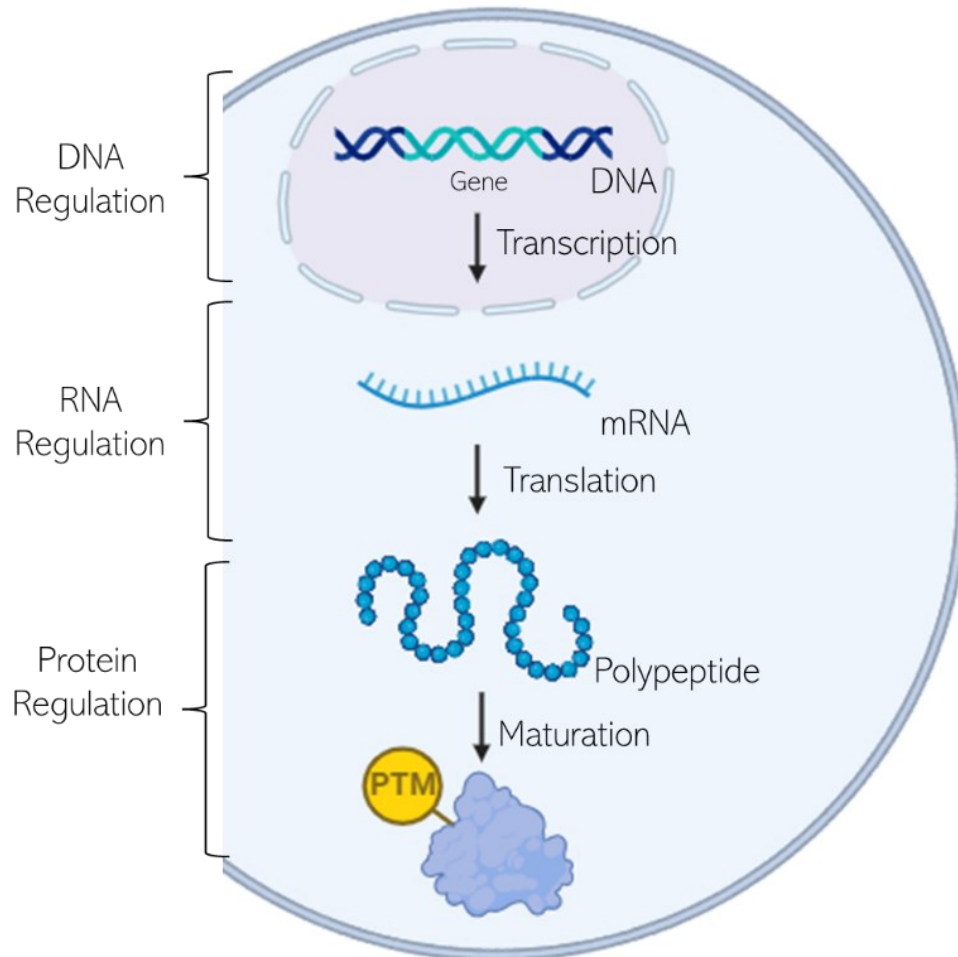


Figure 1.2: Engineered regulation can occur across the central dogma. Tools that act on DNA intervene to regulate how genes are transcribed and the rate of transcription of these genes. Common manipulations include overexpression, knock down/out or tuning transcriptional outputs. RNA interventions engineer RNA molecules to perform reactions post-transcriptionally or control the degradation of mRNAs impacting translation. Proteins can also be regulated after their translation by control of function or lifetime for example by the addition of post-translational modifications (PTMs). Image created with BioRender.com.

Most mammalian cell engineering projects begin with exposing the cells to new DNA or RNA instructions, which can direct them to perform different behaviours. The development of a wide array of DNA assembly methods, predominantly based on DNA manipulation using molecular biology enzymes, has been key to facilitating this process (Casini et al., 2015). Just as engineers follow established standards to ensure the functionality of their circuits, molecular

biologists employ DNA assembly toolkits which provide a predefined assembly framework and are rich with standardised parts adapted for use in the system. New parts can also be introduced into the system by a process often referred to as domestication. This aids in the correct arrangement of DNA segments for the regulation and production of RNA and protein, known as transcriptional units (TUs). TUs are often components of designed circuits which perform computation in cells for a wide range of applications. The use of modular genetic toolkits improves the assembling efficiency of new genetic designs and reduces the time to reach production and testing of constructs of interest (Gibson et al., 2009; Weber et al., 2011).

Many toolkits are designed with the specific goal of simplifying the process of assembling genetic circuits from characterised parts organised in a way that is compatible with expression in mammalian cells. The assembly of transcriptional units and characterisation of parts for mammalian synthetic biology has been addressed and summarised by us and others (Gómez-Schiavon et al., 2020; Zouein et al., 2021; Young et al., 2021).

For example, the mammalian toolkit (MTK) describes sequential cut and paste reactions, a workflow named Golden Gate, for the directional assembly of DNA parts into transcriptional units or multi-transcriptional unit plasmids. The kit provides access to more than 300 domesticated and characterised parts including regulatory parts, peptide linkers and reporters (Fonseca et al., 2019, 2020).

The future of DNA assembly lies in methods that promise precise, efficient, and fast synthesis of DNA. Nucleic acid synthesis technologies, which are advancing at an extraordinary pace, could potentially address these needs, surpassing molecular biology-based assembly methods. Companies are working to make this a reality by developing DNA synthesisers that accurately “print” longer DNA for lower prices (Hoose et al., 2023).

Engineered constructs regardless of the level in which they act have differing characteristics such as timelines, inheritability, range and resource requirements. We will explore these regulatory levels in the following sections.

1.2.2.1 Circuit Features & their Importance

The use and predicted response of designed circuits are often considered before their implementation and the capacity to achieve goals assessed. This allows the engineering of particular response features that enhance the performance of the engineered cells. For the purpose of this thesis, we are particularly interested in considering dynamic range, detection

range and response time. These are depicted below as changes in detected output upon increases in input concentrations (**Figure 1.3**).

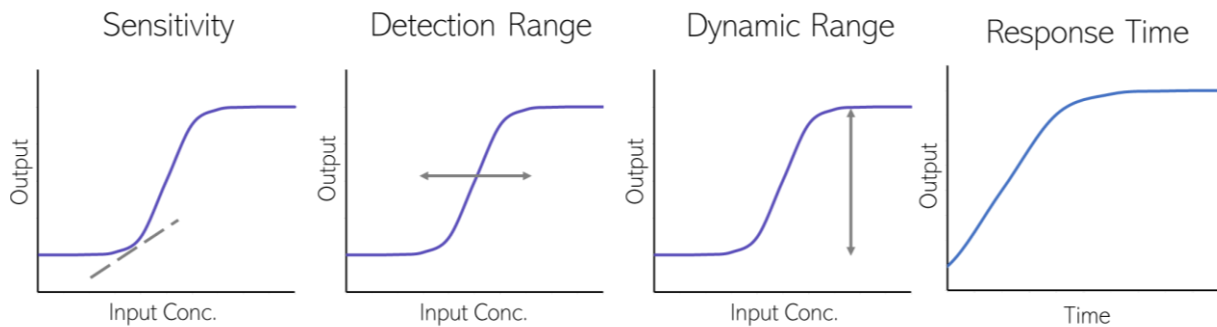


Figure 1.3: Characteristics that determine the performance of a cellular circuit. When an engineered system is introduced into the target cells, engineers evaluate the characteristics of the resulting circuit to understand its functionality. This includes sensitivity which defines how narrow the window of input concentrations is between on and off states. Sensitivity is also related to detection range, which is the range of input concentrations where a different amount of output is detected. Other features also include dynamic range which describes how different the initial and final output are and response time which describes the duration of time it takes for the system to reach an equilibrium or steady state.

The dynamic range represents the difference between the maximum output and the basal expression level, providing a measure of the range of a circuit's achievable outputs. Detection range, or sensitivity, describes the ability of the circuit to detect and respond to varying levels of input signals. A circuit with a high detection range can respond to a wide range of input signal concentrations with detectably different output signals. In comparison, ultrasensitive circuits switch almost digitally from their minimal output to maximal output within a narrow input window. Finally, response time refers to the time taken by the circuit to reach its maximum output, a critical factor in applications where rapid response is required. Depending on the specific requirements of an application, engineers can select module setups that exhibit these desired properties.

In the subsequent sections, we will delve deeper into methods of engineering transcription, translation and post-translation, discussing how these processes can be engineered to achieve diverse features.

1.2.2.2 DNA Control: Transcriptional Regulation

DNA-level interventions change the transcription of genes, which in turn alters the produced RNA. Methods can be subdivided into permanent editing of DNA, spearheaded by the use of DNA nucleases and recombinases, or transient editing, based on expression regulators which control the expression without permanent editing (**Figure 1.4**). Examples include epigenetic switches or TF modulation which directly control access to DNA by the transcriptional machinery. Regulatory interventions at this level take the longest to observe. This is because it is the initial step in the expression process, which includes subsequent stages of processing before the outcome can be observed (**Figure 1.2**).

Permanent genetic editing through nucleases such as CRISPR and recombinases allows guided genetic sequence modifications. CRISPR nuclease, which requires the co-expression of a small guide RNA (gRNA) for targeted cutting at complementary DNA, can further be coupled with endogenous cellular repair mechanisms to achieve altered DNA sequences (Scully et al., 2019; Nambiar et al., 2022). Targeted double-strand breaks which encourage repair machinery can lead to the incorporation of new DNA (when added to cells) by homology directed repair or error-prone repairs which can lead to random insertions or deletions (indels) (**Figure 1.4.a**). Error-prone repairs often lead to a change in the reading frame stopping correct expression from genes.

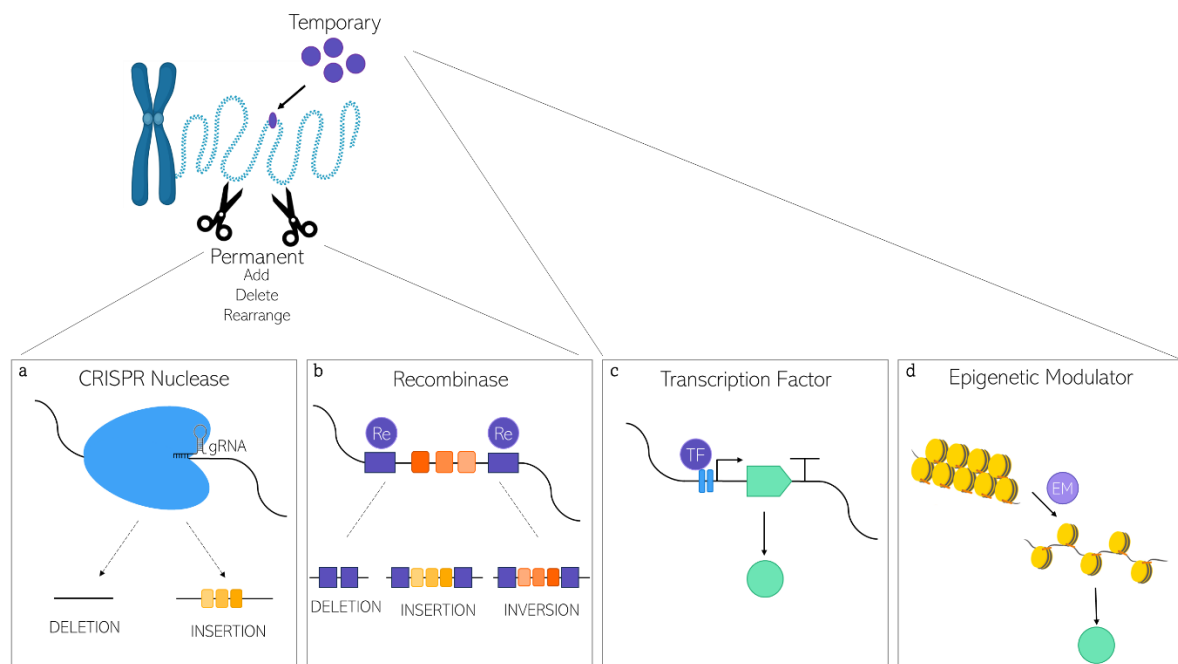


Figure 1.4: Methods of permanent or temporary DNA regulation. **a.** Nucleases such as CRISPR nuclease cause double-strand breaks in DNA that trigger native repair mechanisms which can lead to DNA sequence alterations such as insertions or

deletions. **b.** *Recombinases are a method of permanent genetic editing that recognises a specific DNA sequence (Re) and rearranges the sequence within. Recombinases can lead to insertions deletions or inversions of DNA sequence within.* **c.** *Temporary genetic edits can trigger differential expression patterns for the duration a TF is expressed for example.* **d.** *Temporary genetic edits may also lead to the unwinding of DNA elements that were previously silenced through the action of epigenetic modulators.*

The other way in which permanent changes to DNA sequence can be achieved is through the use of recombinases, which are proteins that recognise specific sequences in DNA of interest and catalyse cutting reactions that lead to editing or reorganisation of the sequence within (**Figure 1.4.b**) (Nagy, 2000). Unlike nucleases, this process does not rely on endogenous DNA repair to achieve insertions, deletions or inversions. Traditional recombinases used cannot be programmed to any target DNA sequence, therefore novel tools have been developed to incorporate CRISPR's programmability with recombinase specificity (Durrant et al., 2023; Yarnall et al., 2023).

Permanent addition, reorganisation or deletion of the genome components can be inherited by cell replication, and this type of engineering can take a few weeks to complete and verify. Verifying edits includes searching for off-target integrations or deletions and can be a significant shortcoming of this approach as unspecific editing of unintended parts of the genome can cause disruptions or introduce unpredictability into the system (Giannoukos et al., 2018). Moreover, the inheritability of the engineered genome raises safety debates and control issues, currently limiting the translation of research into the marketplace (Bruetschy, 2019; Health and Safety Executive, 2019). Despite these complications, recently the first gene-edited therapy for blood diseases has gotten regulatory approval in the UK (Wong, 2023). So, permanent DNA editing technologies are essential when cells require specific modifications, for example when a cell is unable to express a functional gene. However, cellular interactions are often complex and expression is often graded. Therefore, permanent editing for the overexpression or deletion of a gene or genes may not fully capture the intricacies involved in processes and could overlook crucial information hidden by the cells' capacity to adapt and manage changes (Michaels et al., 2019; Bowman et al., 2020). Instead, cells naturally regulate transcription, balancing expression levels without constant changing of their DNA sequence (Casamassimi & Ciccodicola, 2019).

Therefore, where possible, avenues that temporarily instruct the expression profile of cells (**Figure 1.4.c,d**) as an alternative to permanent changes would be used more readily and allow greater control (Kallunki et al., 2019). One major group of proteins which relay messages

from the cell and environment to the genome is the group of proteins collectively known as TFs (Chong et al., 2022).

1.2.2.2.1 Transcription Factors: Native TFs

Signal transduction pathways activated by extracellular and intracellular signals eventually lead to the activation of TFs influencing gene transcription (Kabir et al., 2018). TFs dictate the response to cues from the environment by allocating transcription resources to gene targets at regulatory elements, such as promoters and enhancers, to produce the correct amounts of specific RNAs at the correct times (Ali & Brewster, 2022). Regulation is flexible and reversible allowing constant adaptation to new signals. TFs are pleiotropic in that they can bind to multiple DNA sites for broad regulation, and at each site, TFs can bind in a cooperative way causing synergistic activation or repression at the Transcription Factor Binding Sites (TFBSs) (Li & Altman, 2018). Gene regulation by TFs is partially governed by these binding characteristics to targets, a process which involves a balance between strength of affinity and cooperativity, i.e. where the binding of one TF to its sites enhances the chance for another one to bind (Bragdon et al., 2023). However, the presence of binding sites is not sufficient to predict regulation by TFs as all cells of an organism contain the same DNA and therefore binding sites. Instead, TF concentration and compaction of the DNA around nucleosomes (epigenetic state) also determines the degree of impact of regulation of most TFs (Janssen et al., 2000; Pop et al., 2023).

In conditions where DNA is densely packed, TFs cannot access or bind sites and therefore these regions are not transcriptionally active. While most TFs cannot bind compacted DNA, pioneer factors, insulator TFs and bookmarking TFs have been found to bind dense DNA and displace nucleosomes (Pop et al., 2023). In the case of pioneer factors, the binding allows the remodelling of the epigenome increasing access and potential transcription for functions such as cellular differentiation into specific cell types (Cirillo et al., 2002). TFs that idle after nucleosome displacement, such as in the case of members of bookmarking and insulator TFs, ensure specific purposes. Bookmarking TFs bind loci throughout cell division and facilitates reactivation of these genes, whereas insulating TFs act as boundaries preventing the spread of dense chromatin, therefore reducing gene silencing of important genes (Wang et al., 2014). This outlines the broad utility of TFs and how they exert diverse effects on cells.

Intracellular concentration of TFs within a cell is also a major determinant of their regulatory impact. Cells employ control features, such as regulatory loops, to modulate the concentration of TFs, thereby influencing gene regulation (Calkhoven & AB, 1996; Gyorgy, 2023). As an example, key TFs for the differentiation of myoblasts into fully formed muscle cells (myogenesis) are regulated by regulatory loops. This process is governed by autoregulatory feedback loops, where myogenic regulatory factors regulate their expression ensuring the maturation of myoblasts into differentiated muscle cells (Vicente-García, Hernández-Camacho & Carvajal, 2022).

Another way that intracellular TF concentrations are regulated is through long tracts of DNA repeats called Short Tandem Repeats (STRs) that competitively bind with TFs, altering free TF concentrations and further tuning the expression of their targets (Horton et al., 2023). Due to their control of free TF levels, STRs have been studied extensively, showing functions of epigenetic tuning at STRs (Janssen et al., 2000; Liu et al., 2007) or the impact on expression in the presence of synthetic repeats (Lee & Maheshri, 2012; Wan et al., 2020). Due to their pivotal roles in regulating cells, the expansion of these regions can cause certain diseases, generally referred to as repeat expansion diseases (Paulson, 2018).

In summary, TF features, such as concentration and binding site specificity, and DNA features, such as accessibility and number of TFBSs, determine the regulation of transcription in cells. Their important impact on cells makes them the target in experimental methods such as those for the up or downregulation of genes to direct differentiation of stem cells towards specific lineages (Joung et al., 2023). Consequently, dysregulation in the regulatory networks governed by TFs can result in a wide range of diseases and disorders, spanning from developmental diseases to autoimmune diseases and even cancer (Lambert et al., 2018).

1.2.2.2.2 Transcription Factors: Synthetic TFs

Due to TFs' significant impacts on cells, the development of synthetic TFs has been investigated to develop novel tools for programmable control of gene expression in mammalian cells (Lienert et al., 2014). Synthetic TFs consist of a DNA binding domain, for target DNA sequence recognition, fused to effector domain(s) for activation or repression of transcription. The use of synthetic TFs is attractive as they can be directed to control endogenous genes but also synthetic promoters for broad utility. Some features that are often sought after are inducibility to non-native signals, sensitivity and versatility of target.

For example, tetracycline transactivators, a widely used group of synthetic transcriptional regulators, utilize modified versions of the bacterial Tet repressor protein (TetR). These modifications include tTA (tetracycline transactivator) and rtTA (reverse tetracycline transactivator), which are fused with the viral activator VP16. Under specific conditions, relating to inducer doxycycline concentration, these can bind to a particular DNA sequence, the Tet operator (TetO), in promoters (**Figure 1.5**) (Loew et al., 2010; Zhou et al., 2006).

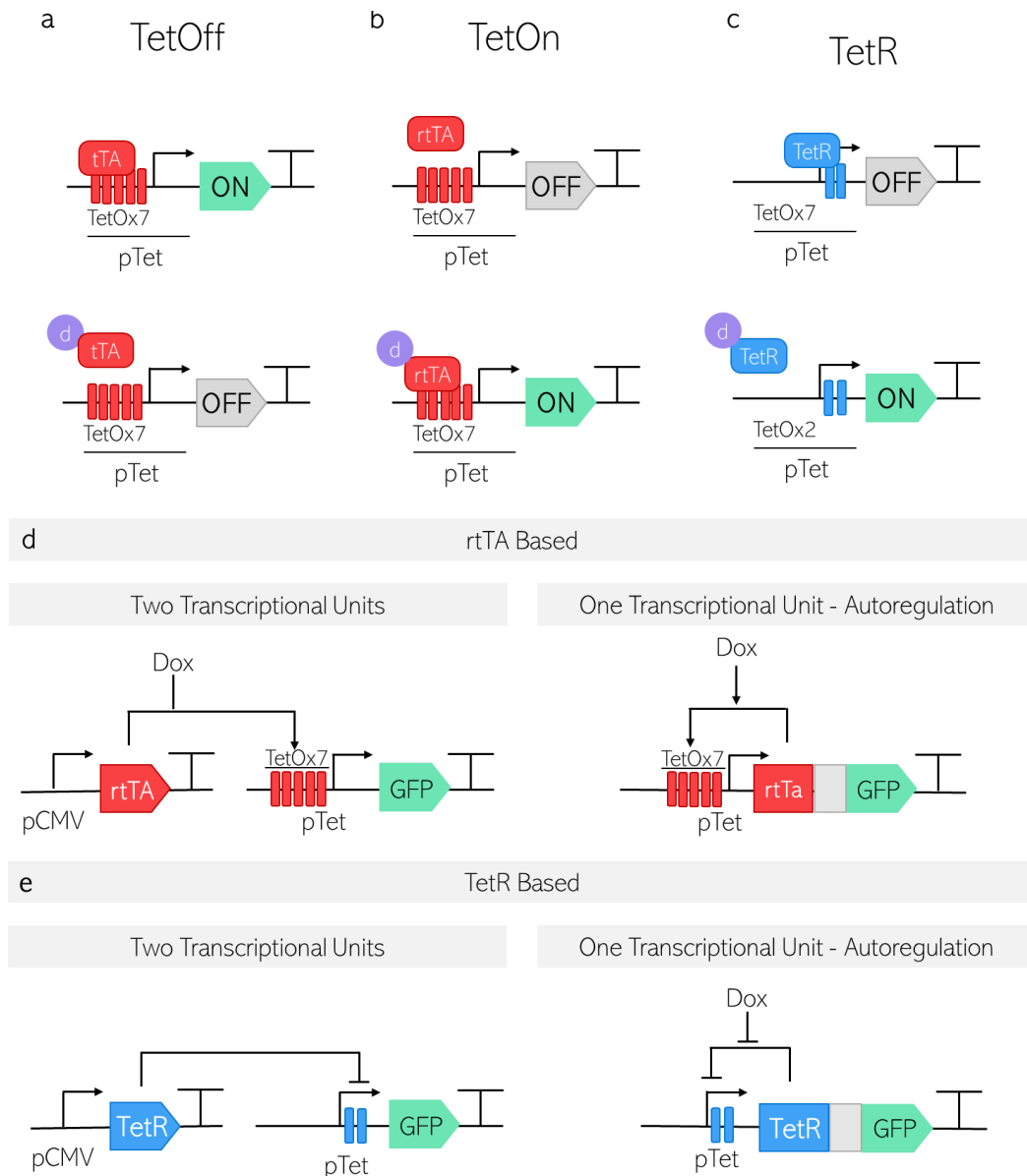


Figure 1.5: Mammalian transcriptional regulation using Tet-based systems. **a.** TetOff system uses tTA TF which activates expression in the absence of a dox inducer. In the presence of dox, tTA loses its ability to bind with operator sites, reducing transcription. **b.** TetOn relies on TF rtTA which only induces transcription in the presence of dox. **c.** Bacterial TetR can be used to block transcription in mammalian cells while the addition of dox causes TetR to detach from binding sites in promoters. **d.** rtTA based circuits where rtTA TF is expressed alone (left) or in concert with a reporter gene (right). In the autoregulation loop, the induction of the system with dox induces the completion of the

rtTA TF, allowing its binding to its operator sites and inducing further expression of more rtTA and reporter. In the two transcriptional unit version, a stable but high amount of rtTA is constitutively expressed. e. TetR based circuits where TetR TF is expressed alone (left) or in concert with a regulated reporter gene (right). In the autoregulation loop, the induction of the system with dox induces the completion of the TetR TF, allowing its unbinding from operator sites and inducing further expression of more TetR and reporter. In the two transcriptional units' version, a stable but high amount of TetR is constitutively expressed which is alleviated in the presence of dox.

In the repressible system, often called TetOff (Gossen & Bujard, 1992), tTA normally binds with the operator causing expression from the target. This occurs in the absence of a small molecule inducer doxycycline (dox). The presence of dox in culture induces conformational changes in the tTA TF which loses the ability to attach to DNA, reducing transcription at sites (**Figure 1.5.a**). This is such that, the more dox added to the culture, the more tTA molecules become bound in the OFF state unable to bind TetO DNA. At high concentrations of dox, all tTA becomes unable to bind with the cognate promoter and gene expression is mostly silenced.

In contrast, the activatable system referred to as TetOn (Gossen et al., 1995) requires dox binding to rtTA TF for DNA binding and gene activation (**Figure 1.5.b**). As more dox is introduced into the system, the TF's capacity to bind TetO sites in promoters increases. This continues only until the maximum capacity is reached, which is determined by various factors, including the strength of the promoter and the cellular context. Both TetOff and TetOn systems act on the same DNA sequence but vary in action by the use of the different TFs which gain or lose the capacity for DNA binding upon binding dox.

Another variation of Tet TF use in mammalian systems uses TetR without appended activator VP16 to repress a different mammalian promoter by steric hindrance. Dox induces conformational changes in TetR, dislodging TF and repression on this promoter (**Figure 1.5.c**) (Yao et al., 1998). Therefore, the more dox added to the system, the less TetR available for binding operators and the greater the output from the transcriptional unit.

Tet-based TF systems have been extensively used in a variety of applications, due to their reversible and tuneable control of gene expression, including the generation of transgenic mice with inducible gene expression (Saunders, 2011).

More recently, synthetic transcriptional regulators based on CRISPR/Cas9 systems have been developed (Qi et al., 2013). These regulators use a modified version of the Cas9 protein deficient in nuclease activities called dCas9, which is usually additionally engineered to be fused

to or attract transcriptional activator or repressor domains (Chavez et al., 2016; Li & Zhou, 2021; Wang et al., 2022b). The major advantage of dCas9 systems stems from the ability of effectors to be directed to almost any DNA sequence of interest using gRNAs. Therefore, these systems attract to a target DNA region transcriptional activator or repressor domains. This approach has been used to develop a wide range of synthetic transcriptional regulators with varying levels of specificity and activity (Jusiak et al., 2016). It is of note that dCas9 affinity and cooperativity might be case dependent, with many studies screening several guides to achieve the required gene expression control (Wang, La Russa & Qi, 2016; Feng et al., 2023). Despite the simplicity of engineering dCas9 systems for use, detrimental off-target effects are possible (Bae, Park & Kim, 2014; Wang, La Russa & Qi, 2016).

Several groups have also developed synthetic transcriptional regulators with improved specificity based on other modular classes of DNA binding proteins, such as transcription activator-like effectors (TALEs) (Thakore & Gersbach, 2016). In some cases, TALEs have also conferred better cellular responses than dCas9 regulators (Gao et al., 2014). While the development of these takes much longer because protein engineering is required for DNA targeting, they provide more targeted results (Waryah et al., 2018).

Zinc fingers (ZFs) have also been used as synthetic transcriptional regulators. Each finger often recognises 3-4 DNA base pairs (bp) but linking several can increase specificity and therefore reduce off-target effects (Kim & Kini, 2017). While ZFs can be laborious to design, they benefit from smaller sizes compared to TALE and dCas9 effectors, improving delivery to cells. These characteristics form the basis of why zinc finger domains are used in the development of characterised suites of TFs (Donahue et al., 2020).

The landscape of synthetic modulators of transcription is therefore rich in artificial TFs, the user must compromise between off-target effects, simplicity of design and size for best use for their specific applications.

1.2.2.2.3 Utilising TFs to engineer genetic circuit features in mammalian cells.

TFs, whether engineered or natural, are commonly used in engineered mammalian circuits as cell processing modules. These modules mediate sensed messages through cells eliciting responses (Mansouri & Fussenegger, 2022). For example, a commonly used sensing component is engineered through extracellular synthetic receptors that are directly fused to

synthetic TFs to actuate regulated endogenous or transgene expression in circuits (Manhas et al., 2022).

The choice of which TF to use is primarily driven by the desired effects within the circuit. Some factors to be considered include off-target effects, simplicity to develop (especially in cases where iteration is required), size, inducibility and response curve. Moreover, how the TF is expressed can impact the performance characteristics of circuits so that the steepness of response, or time to reach the steady state changes (**Figure 1.3**). This is such that the same TF can lead to very different responses depending on how they are implemented in circuits.

For example, identical components of the TetOn system can be configured in either one or two transcriptional units, leading to different responses. In the two-transcriptional unit setup shown on the left side of **Figure 1.5.d**, the gene-activating TF rtTA is expressed from a constitutive promoter, which subsequently activates another gene controlled by the rtTA responsive pTet promoter. Upon changing the organisation of the same components such that the activator is expressed under its own promoter (auto-activation loop), the circuit's response changes so that induction is ultrasensitive, which means it goes from off to on within a narrow induction range but takes a longer time for response to achieve steady state (**Figure 1.5.d**) (Hermsen, Erickson & Hwa, 2011).

Changes in output expression from circuits containing TFs depending on their configuration in circuits can also be seen when comparing TetR effects on another transcriptional unit (two transcriptional units) versus when used in an auto-repression circuit (**Figure 1.5.e**). While the two-unit system benefits from lower background expression, the auto-repression circuit yields a more uniform linear expression for a broad range of inducer concentrations (Nevozhay, Zal & Balázsi, 2013). Therefore, understanding the predicted circuit function is important for part and circuit choice.

Overall, the development of methods for synthetic transcriptional regulation for use in mammalian cells is a rapidly progressing field, with new tools, circuitry and applications being developed frequently. For example, a recent paper described programmable control of GOIs using dCas9 with 25-fold stronger activity than the strong mammalian promoter pEF1a in multiple mammalian cell lines (Chen et al., 2022).

These tools and others have potential for a wide range of applications, including in basic research, as well as in biotechnology and gene therapy. The diverse uses of TFs of natural or engineered origin facilitate a comprehensive response to many cellular conditions. While

currently capacity for tuning responses is governed by expression level dictated by promoter strength, there is room for fine-tuned handles on response in engineered cells. Methods that improve flexibility in expression levels of TFs can further lead to more tuned responses in cells required in some cases where only a small change in expression levels is required to alter cellular function (Lee & Young, 2013).

1.2.2.3 RNA Control: Post-Transcriptional & Translational Regulation

Engineering approaches have also been adopted for the regulation of RNA in gene circuits. However, there are some additional considerations to be taken into account when intervening in gene expression at the RNA level, with these considerations often utilised as benefits for taking the RNA approach. Firstly, RNA has a shorter half-life than DNA or proteins, meaning a faster turn-around of molecules. This makes it more suitable for applications where rapid detection and response to events in real-time are required (Ortega, 2023). Furthermore, when instructions are provided as RNA to cells, stimulation is not long lasting, like with DNA genomic integration, and modifications are also not inherited from generation to generation. RNA also often only needs to reach the cell's cytoplasm, whereas DNA must reach the nucleus to exact its effects (Spagnou, Miller & Keller, 2004). Because of these properties, there are fewer ethical implications in interventions that change RNA sequence and levels (Guerriaud & Kohli, 2022).

Other benefits include smaller cargo size, and properties facilitating delivery compared to proteins (Marschall, 2021). RNAs have consistent physicochemical properties and negative charge compared to proteins which vary greatly in size, structure, hydrophobicity and charge, requiring careful consideration of protein vehicles for delivery. Moreover, there is an amplified effect with using RNA meaning fewer copies need to be delivered for response. For example, one mRNA molecule can act as the template to produce many proteins prior to its degradation. All in all, RNA intervention provides many benefits encouraging engineering at this level (Zhu et al., 2022).

This area has received renewed interest due to major developments in this field leading to the unprecedented development, approval, and widespread administration of mRNA as vaccines for the treatment of the infectious disease SARS-CoV-2 (World Health Organization (WO), 2021). The delivery of encapsulated mRNA molecules triggers immune responses in preparation in case of infection. The delivery of mRNA as therapies is now being considered for many other

applications, such as for monoclonal antibody therapeutics conferring simpler delivery and less immunogenicity compared to antibodies expressed in mice (Van Hoecke & Roose, 2019).

Cell engineering at the RNA level frequently entails the modification of messenger RNA (mRNA), which are produced through transcription. Transcription of mRNAs allows cells to convey genetic instructions to ribosomes, guiding the synthesis of proteins. This can be by impeding translation or impacting the stability of mRNAs, thus affecting the number of proteins that can be produced. One common way that mRNA is naturally and synthetically controlled in cells is by the addition of small RNAs (sRNA) complementary to the target mRNA leading to their binding in the cytoplasm and degradation or inhibition of translation in a process termed RNA interference (RNAi) (**Figure 1.6.a**) (Mukherji et al., 2011). MiRNAs (microRNAs) expressed from mammalian genomes regulate gene expression post-transcriptionally and have been found to have major roles in tuning gene expression, signalling and disease states (Arora et al., 2013; Donati, Ciuffi & Brandi, 2019). Their specific expression patterns in disease states make them ideal biomarkers for diagnosis and many have created biosensors and classifier circuits for their detection (Arora et al., 2013). Due to their flexibility in sequence, the applications of RNAi also extend into therapeutics where antisense oligonucleotides (AGOs) and siRNAs (silencing RNA) against mRNA transcripts of interest have been approved for therapeutic use (Anon, 2017; Guerriaud & Kohli, 2022). As this approach relies on the binding of relatively short sequences with complementarity, there is a chance of off-target binding at unintended locations with increasing size of genomes (Qiu, 2005).

Another class of RNAs that can be engineered are ribozymes, which catalyse reactions (**Figure 1.6.c**). While these can be used similarly to miRNAs or AGOs to encourage the degradation of RNAs, ribozymes have various other possible functions from the ligation of RNAs to intron splicing and folding (Puerta-Fernández et al., 2003). RNA tools have also been developed as agonist and antagonist aptamers, facilitating, or preventing cell interactions respectively (**Figure 1.6.b**) (Kang & Lee, 2012). Notably, RNA aptamers have been generated to exhibit selectivity in binding to specific TFs or regulatory proteins, offering a means to modulate their activity (Zhu et al., 2022). Despite the broad potential in the regulation of cells, RNA aptamer development for mammalian cells remains difficult especially due to issues related to intracellular folding, degradation and function (Hou & Jaffrey, 2023).

Moreover, a combination of ribozyme/aptamer strategies have been used to craft riboswitches and circuits with unique features (**Figure 1.6.e**), to allow precise control of gene expression in mammalian cells. For example, a riboswitch can be designed by harnessing the self-cleaving functions of ribozymes with the inducible functions of aptamers. The resulting switch allows

inducible gene expression control when added to untranslated regions of mRNA transcripts (Mustafina, Fukunaga & Yokobayashi, 2020; Xiang et al., 2019; Fukunaga et al., 2023).

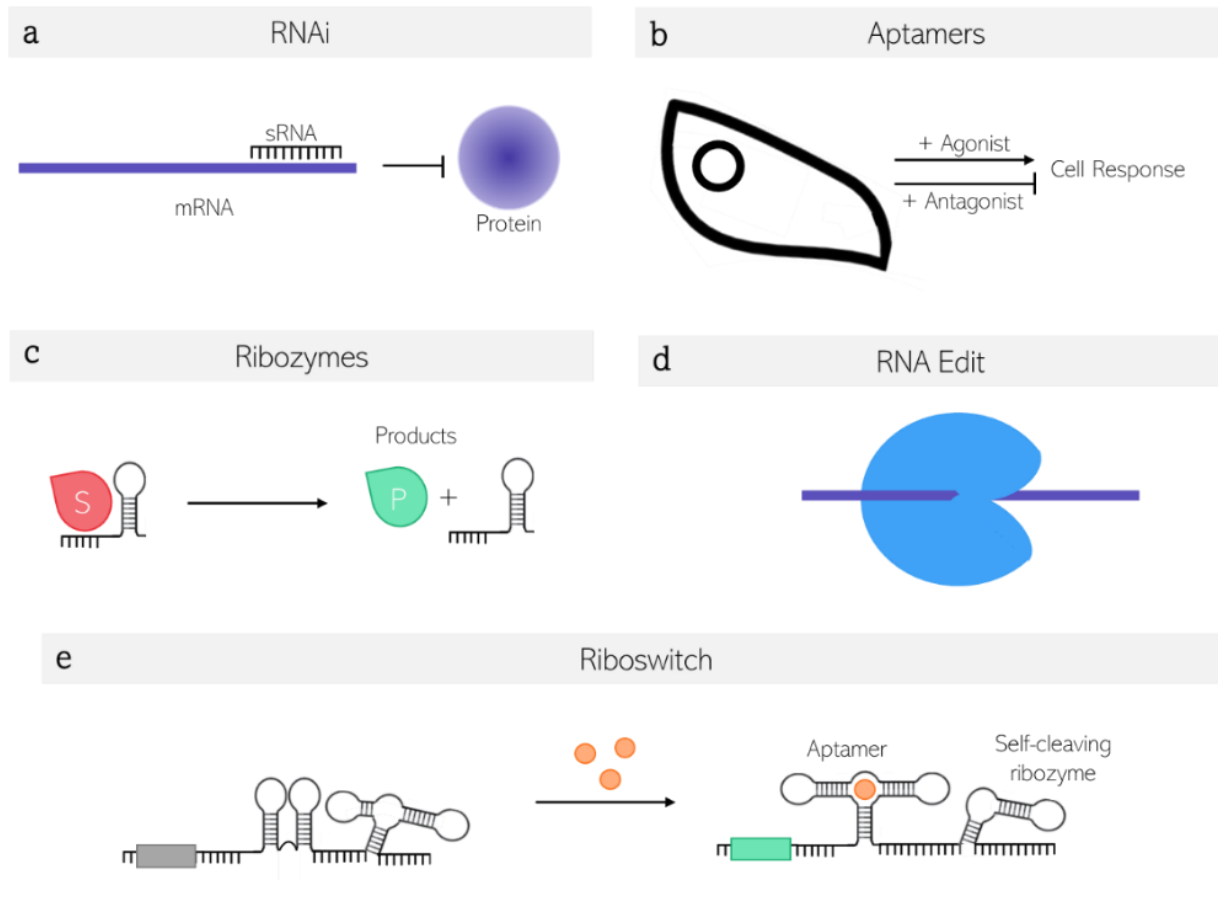


Figure 1.6: Methods of RNA regulation. **a.** RNA levels can be regulated in cells by miRNAs that bind regions within mRNA transcripts and repress translation. **b.** RNA aptamers can also be designed and built to promote or prevent cell responses. One example of this is by binding membrane receptors. **c.** Ribozymes bind to substrates and promote the production of the product. **d.** Moreover, RNA sequences can be cut or edited using CRISPR 13 technologies. This requires a gRNA for targeting. **e.** Aptamer and ribozyme technologies can be combined into transcripts to produce switches that require the cleavage of a self-cleaving ribozyme and the binding of a ligand to an aptamer for the expression of genes of interest.

Most recently, this research area has seen the development of RNA editing using CRISPR. Cas13 family of proteins have been engineered to bind RNAs and edit nucleotide sequences, or cleave specific RNA molecules, guided by gRNAs (**Figure 1.6.d**). As this class of intervention relies on the addition of CRISPR proteins, disadvantages due to the introduction of additional large protein components, and off-target effects must be considered (Zhu et al., 2022).

RNA engineering encompasses technologies that regulate gene expression, by changing RNA sequence or RNA availability for processes such as translation or enzymatic reactions, in relatively small packages. Easier delivery and quick turnaround time further encourage their use case in mammalian cells as therapies and beyond.

1.2.2.4 Protein Control: Post-Translational Regulation

Proteins, which serve as the primary effectors within cells, are produced through a process known as translation. This process involves the synthesis of proteins from mRNA transcripts within the cell's cytoplasm. The crucial role of translation in modulating cellular activity and function is underscored by its preservation throughout the evolution of various species, highlighting its fundamental importance in biological systems (Ganoza, Kiel & Aoki, 2002). After translation, polypeptides must accumulate, fold into proteins, and post-translational modifications (PTMs) added for functions (**Figure 1.2**).

Interventions at the protein level encompass methods of regulating protein concentrations or functions through binding, post-translational modifications or changes to amino acid sequence (Bojar & Fussenegger, 2020; Ramazi & Zahiri, 2021).

Protein capacity for binding a wide variety of signals makes them classical druggable targets, with 95% of drug targets approved for use until 2017 being proteins (Santos et al., 2017). Upon binding, proteins change structure altering protein functions. Some common targets are agonists which induce/trigger protein response or antagonists which block the potential to respond.

More recently, protein engineering has aimed to increase the repertoire of signal/function pairs to engineer new responses to protein binding (Bojar & Fussenegger, 2020). One group of proteins which is often engineered for this purpose are membrane receptors. These represent the largest group of classical druggable targets due to their positioning on the cell surface and bind agonists or antagonists for various downstream cellular effects (Sriram & Insel, 2018). The rational design of receptors has led to breakthroughs such as in CAR T-cells and the reprogramming of cell behaviour (Javdan & Deans, 2021).

In other efforts, researchers have developed a technique known as conditional protein splicing used to activate proteins in a controlled manner (Wang et al., 2022a). This process cleaves protein amino acid sequences in specific ways. When the process involves the cleavage of an inactivating domain *within* a protein, the process is referred to as cis splicing. This cleavage

reconstitutes the protein, rendering it active and capable of performing its intended functions. When the induced association of two proteins is required, trans inteins (also known as split inteins) can be used. Cleavage of the inactivating domains on both proteins allows activation of the protein pair. These tools are versatile allowing the activation of proteins in response to a variety of inducers ranging from light and antibiotics to proteases (Di Ventura & Mootz, 2019).

Another way in which proteins are often regulated is through the control of protein half-life through degron domains or ubiquitin modifications. Additional features such as inducibility are used to tag proteins of interest for degradation by intracellular machinery in a controlled way. An example used in many organisms including mammalian cells is the Auxin-Inducible-Degron (AID) which provides inducible fast and reversible degradation of proteins containing AID. Tagged proteins are degraded in the presence of the inducer auxin, a plant hormone (Holland et al., 2012; Yesbolatova et al., 2020). Addition of degrons to proteins is often regular, for example by addition at the N terminal end of a protein, simplifying implementation and design of proteins with these modifications (Chassin et al., 2019).

Novel enzyme engineering is another common manipulation of proteins allowing for the creation of biocatalysts with improved efficiency, stability, and specificity (Ding et al., 2022). Novel operations can also be added to cells through the addition or engineering of new enzymes. For example, a plant PTM has been engineered to work in CHO, a cell line often used as a workhorse in mammalian bioproduction, improving product quality by improving bioactivity and half-life of produced protein products (Heffner et al., 2018; Larsen et al., 2020).

Therefore, proteins can be engineered to add novel functionality through combining domains but also by regulating and controlling protein levels through different methods.

While engineering at the protein level has facilitated large benefits in many fields, engineered proteins have complex, sometimes unpredictable, structures and therefore have led to the development of AI-assisted software for predicting this. For example, AlphaFold, an AI system developed by DeepMind, has made significant strides in predicting protein structures with high accuracy (Jumper et al., 2021).

While protein engineering has proven to be an important point of intervention, the molecular complexity of engineered proteins can put a strain on cellular resources, potentially triggering cellular stress responses (addressed more comprehensively in a later section). Moreover, the delivery of proteins into mammalian cells, required for certain applications that require dosage control, short lifetime or immediate effects, is still a significant issue. This means instructions for protein expression are more often delivered as DNA or RNA for expression within cells (Ebrahimi

et al., 2022). Specific culture conditions, cell type, protein size, efficiency and modifications must all be taken into account, complicating delivery so that delivery is often optimised on a case-by-case basis (Porello & Cellesi, 2023).

1.2.2.5 Multiple Interventions

Cells utilize a variety of regulatory mechanisms at each layer to accurately manage the abundance and activity of their genes (Black, Perez-Pinera & Gersbach, 2017). Naturally occurring dual regulation on transcription and post-transcription levels in cells is observed when miRNA represses translation of a TF, preventing the action of the TF on gene(s) of interest (GOI) (Arora et al., 2013; Bracken, Scott & Goodall, 2016; Matsuura et al., 2018; Martinez & Walhout, 2009).

As a result, implementing interventions at multiple regulatory layers could potentially enhance the performance of engineered circuits. This is further exemplified in instances where engineered interventions at a single regulatory layer did not result in expected outputs as other regulatory stages either interfered with or compensated for these interventions (Perl et al., 2017). Perl et al, showed that interventions that change transcript levels are often buffered so that they do not always result in different protein levels in many mammalian cell types.

In this direction, researchers can construct more complex biological circuits that act at several points of the central dogma. Such circuits, including some multi-layered/level circuits, can have added benefits compared to those that only consider one regulation approach (Gao, Wang & Wang, 2023). Multi-level circuits refer to those composed of multiple interconnected genetic modules such that the output of one module serves as the input for the next for the implementation of complex genetic operations or the regulation of multiple genes.

One approach uses logic gates which rely on more than one input to help ensure that a cellular response is only triggered under very specific conditions, thereby reducing the risk of unintended consequences (Doshi et al., 2020; Wang et al., 2019). Logic gates developed by Doshi et al in 2020 integrate transcriptional responses, by TF, with post-transcriptional responses by miRNAs into logic gates for classifying specific cell types with smaller DNA sizes compared to other circuits for similar classifying applications (Doshi et al., 2020).

Multi-layered circuits can be developed that incorporate different regulatory levels. A study in bacterial strains showed that multi-layer circuits regulating transcription and translation had

favourable responses such as low basal expression, high maximal expression, and a steep transition between these two states (Greco et al., 2021a).

Moreover, Wang et al developed an AND gate which requires the response to miRNAs for the maturation of gRNA for dCas9 targeting and TF activation of targets (Wang et al., 2019) (**Figure 1.7**).

Together these examples describe a few of the multitude of options for the development of multi-level circuits for improved circuit effects. But these remain difficult to achieve and optimise as circuits rely on well characterised orthogonal modules and a suite of compatible input and output modules (Brophy & Voigt, 2014).

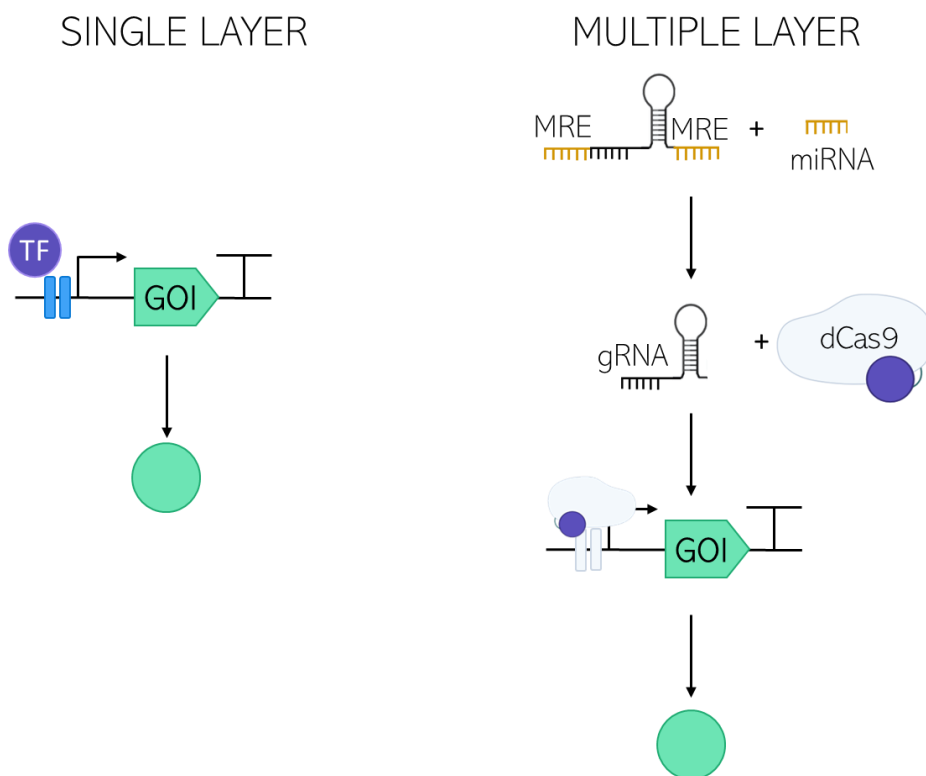


Figure 1.7: Single-layer vs multi-layer circuits. An example of a single-layer circuit (where a TF binds to a specific DNA sequence and regulates the expression of a downstream gene) in comparison to a multilayer circuit (where the output of one layer, i.e. the miRNA binding on an RNA to release a mature gRNA, serves as the input for the next module, the dCas9 based gene regulation of output).

Therefore, engineering principles can be applied to biological processes, transforming them into controllable systems. This involves manipulating the central dogma at transcription, post-transcription, translation or post-translation. The incorporation of engineering rhetoric and

variable tools in cell engineering represents a promising avenue for advancing biotechnology and propelling us into the era of the fourth industrial revolution (Alejandro Lavopa and Michele Delera, 2021).

1.2.3 RESOURCE AWARE ENGINEERING

A major consideration increasingly becoming clear in synthetic biology is that engineering cells by adding in more and more components may not necessarily yield improved performance. Cells operate with finite cellular resources and machinery required for their own gene expression needs. This means that engineers must account for stress inflicted on the cell as a by-product of any engineering that introduces new genes and components that will require resources and machinery to operate. This places a burden on engineered cells leading to cell's differential gene expression profiles, unpredictability in circuit responses, loss of the expression of the engineered components over time or even cell death, reducing the impact of engineered systems outside of the short timescale of lab experiments (Borkowski et al., 2016).

Competition for cellular expression resources poses a significant problem in many engineered model organisms. Bacterial studies have explored different facets of this problem. For example, resource monitors have been developed to detect resource competition (Ceroni et al., 2015) while other studies have explored the cellular responses to conditions of resource competition (Zhang et al., 2021). The significant impacts that these conditions have on engineered systems have led to an array of methods for mitigating these issues for improved function of engineered bacterial cells (Grob, Di Blasi & Ceroni, 2021). In yeast, heterologous protein expression also consumes cellular resources triggering stress responses (Kastberg et al., 2022). Studies have identified potential sources of burden to mitigate against to increase overall heterologous protein production usually limiting yeast strain use in bioproduction applications. To combat effects such as reduced cell fitness and protein titres, engineering strategies have developed more burden resistant strains for bioproduction applications.

Resource competition has also been shown within mammalian cells. A common example of resource constraints in mammalian cell engineering is observed when co-transfection markers are used in experiments. These plasmid designs express a reporter in addition to circuit components which confirms transfection success and allows estimation of intracellular copy number of transfected plasmids. However, this process can impact the expression of circuit components due to resource allocation to reporter expression cassette (Di Blasi et al., 2021). A

partial solution used by Brown et al., 2014 proves transfection success by transfecting additional samples with constitutively expressed reporter as a control for transfection performance (Brown et al., 2014). Assuming transfection performance is similar across samples in an experiment, this control indicates whether the process of transfection was successful, but it does not provide an understanding of intracellular plasmid copy number.

In bioproduction, where mammalian cells are modified to function as biological factories, additional modifications to the cell line are often implemented to enhance the cell's ability to overexpress products and improve yield, usually achieved by reducing cell response to overexpression stress (Dangi et al., 2018; Tihanyi & Nyitray, 2020). For example, miRNAs which do not require the cell's translation machinery for their expression have been used in bioproduction to broadly regulate expression in cell lines to improve cell proliferation (Fischer et al., 2017). These techniques aim to improve the cell's performance in conditions of overexpression, thereby prolonging cell life and yield, but they do not free up or increase resources.

Other methods of response to resource stress conditions have seen mitigation strategies developed into genetic circuits. These circuits often regulate burden through feedback loops designed to self-regulate control for optimal performance of the cell and the circuit (Lillacci, Benenson & Khammash, 2018; McBride, Grunberg & Del Vecchio, 2021; Frei et al., 2022). The prevalent use of feedback loops within these circuits highlights these circuits' features which allow monitoring of cellular conditions and adjustment of expression respectively to effectively manage the system's load. This mechanism ensures that the host organism's limited resources are not overwhelmed by the genetic circuit, thereby maintaining a balance between the circuit's operation and the host's survival. These endeavours enhance the stability and reliability of synthetic biological systems, paving the way for more complex and robust applications in genetic engineering.

Lastly, another way scientists can reduce the effects of resource competition is at the design stage by choosing genetic components with favourable properties. The work by Di Blasi et al. in 2023 has made strides in this area. They compared the required resources of a library of mammalian genetic parts and identified designs with improved performance compared to a resource "monitor" (**Figure 1.8**). The monitor constitutively expresses a reporter and the burden inflicted on the monitor (detected by a reduction in its expression) in the presence of other expression cassettes is measured. This is such that the burden is identified by how drastic the effect is on the monitor. More burdensome competitor plasmids will have larger impacts on the monitor expression whereas less burdensome constructs will not impact the monitor

expression as much. Interestingly, the group also identifies certain conditions where constructs express similar amounts of the output (therefore they should use the same number of resources) but with varying effects on burden, showing the complexity of this process in cells (Di Blasi et al., 2023).

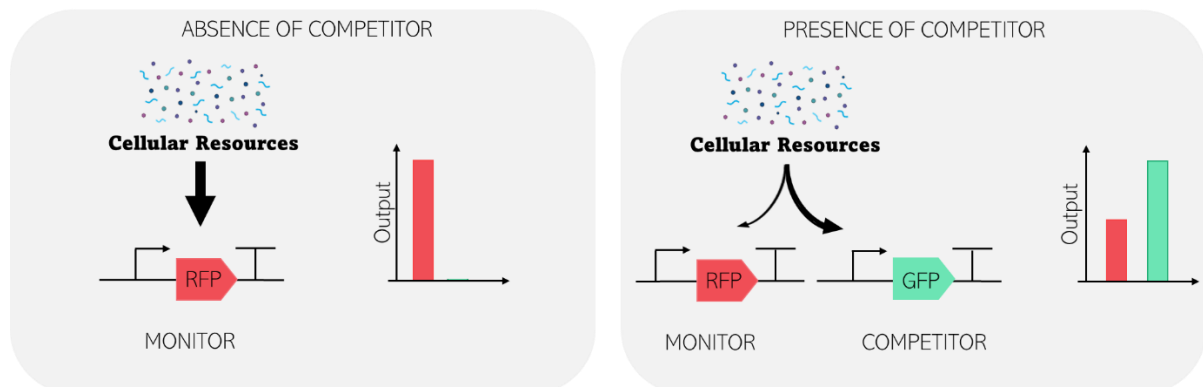


Figure 1.8: *Design of Mammalian Resource Competition Sensor. Finite cellular resources which are usually only allocated to the expression of a monitor constitutively expressing RFP (left) will have to compete for resources with the competitor plasmid constitutively expressing GFP marker(right). Depending on the competitor’s ability to sequester resources, the amount that can be allocated to the monitor will change. This change in RFP fluorescence is reflective of the burden in the system.*

Methods and mechanisms that drive potent effects in cells but with minimal use of host resources are likely to be the future of cell engineering. By reducing the resource load used by designs, resource-aware cell engineering can avoid the effects of resource competition, causing unintended interference with cellular processes, and suboptimal function of engineered components. In the field of bioproduction, taking into account resource usage will ultimately improve the yield of products, reducing the cost of production.

1.2.4 ETHICAL CONSIDERATIONS OF ENGINEERING BIOLOGY

However, while mammalian cell engineering is commonplace and has advanced our knowledge of many biological processes in controlled lab environments, there are still challenges and barriers to overcome in terms of public acceptance, ethical concerns, and regulatory frameworks for the use and release of genetically modified organisms (GMOs) in real-world applications.

Ongoing dialogue across various stakeholders including scientists, policymakers and the community is required to untangle the complex process and improve the safe use of GMOs in the environment and in the clinic (Tellechea-Luzardo et al., 2022).

One such complexity is the difference in perception between scientists and the public on the risk of the products produced. While scientists agree that users need to feel safe, the risks are considered manageable (Visschers & Siegrist, 2018). Other growing public concerns discussed by social scientists surround the disclosure of what could be perceived as deeply personal information in the form of DNA data. People are worried about how this information could potentially be used to their detriment. Additionally, there are apprehensions about potential discrimination arising from the costs associated with implementing these technologies. These concerns underscore the need for careful consideration and consistent public engagement in the handling of these topics (Howell et al., 2020).

Moreover, due to the large breadth of research topics engineering biology can be applied to, revisiting and adapting policies to further ensure safe implementation of technology is an ongoing topic (Wang et al., 2023). This can be through the establishment of robust risk assessment and management strategies to monitor the safety and efficacy of GMOs to prevent adverse effects on the environment or human health (Macfarlane et al., 2022).

Engineering biology requires careful considerations along the process from lab testing and beyond, with feedback from stakeholders especially when they require extensive sequencing or alterations of DNA. Further limiting permanent changes to genetic material, which have some safety concerns associated with them, and better control measures would improve public perception and transition of technologies into everyday products (Sahin, Karikó & Türeci, 2014).

These considerations should be considered in the development of next-generation tools for applications including cell and gene therapies to ensure their responsible use. Scientists can further address this by prioritising features such as enhanced precision, improved safety, and the ability to control gene expression dynamically.

1.2.5 NEXT-GENERATION CELL ENGINEERING TOOLS

Given the above discussed aspects, there exists a significant opportunity for the refinement of genetic tools for genetic regulation with predictable performance. Next-generation tools would benefit from having low resource costs to reduce side effects caused by the expression of engineered components, no permanent genome editing to allow reversibility and reduce safety

concerns, tuneable for different response characteristics and broad regulation of natural and artificial targets for flexibility and control. In this context, miRNAs and TFs emerge as particularly promising candidates for next-generation cell engineering regulation.

MiRNAs and TFs possess several attributes that align with the requirements of next-generation tools. Firstly, they are naturally occurring regulators of gene expression, which means they have evolved to perform their functions with specificity and efficiency (Hobert, 2004). These properties can be harnessed for the rewiring of endogenous regulation. TFs and miRNAs exert regulation on targets with pleiotropic characteristics (see **Figure 1.9.a**), allowing a single molecule to target its binding sites in multiple positions for broad regulation. For example, the TF P53, known for its role in maintaining genomic integrity, binds a variety of DNA targets in genomes to regulate effects such as DNA maintenance/repair or programmed cell death (Yu et al., 1999). Moreover, their cooperative nature (**Figure 1.9.b**) increases the range of outputs and facilitates multiplexing via the use of multiple adjacent/neighbouring targeting sites.

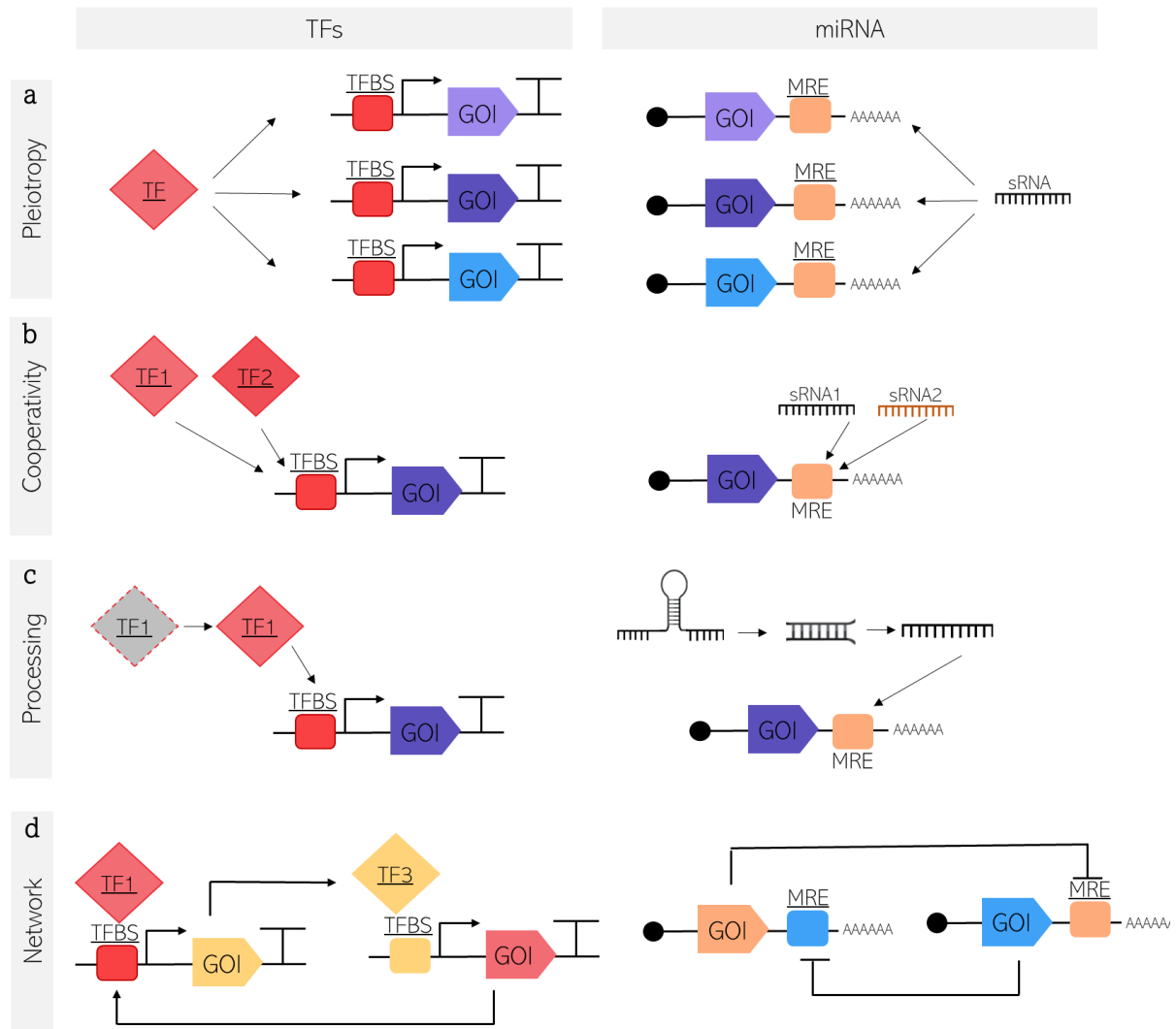


Figure 1.9: TFs and miRNAs have similar functions in gene regulation. **a.** miRNAs and TFs are pleiotropic. This means that TFs can bind several target TFBSs and control their expression. Similarly, a miRNA can bind MREs within different mRNA transcripts and reduce their translation. **b.** TFs and miRNAs are cooperative. This means that more than one TF or miRNA can bind their targets simultaneously, improving effects. **c.** Both regulatory molecules require processing prior to action at the target. For TFs, this may give the protein the ability to enter the nucleus or bind other proteins forming a complete complex. For miRNAs, this includes processing by proteins to reach the final mature conformation. **d.** Regulatory molecules can be used in more complicated networks allowing feedback and control.

In addition, when combined, miRNAs and TFs make up the largest groups of regulatory molecules in mammalian cells (Hobert, 2008) and synthetic versions with desired characteristics can be engineered (Quarton et al., 2018; Bhatt, García-Díaz & Foight, 2023). This makes miRNAs and TFs particularly versatile candidates for use in the modular engineering of cells either in native form or with artificial/synthetic versions. These regulators also require

processing (**Figure 1.9.c**) where TFs must often be activated by signal transduction pathways to be able to localise to the nucleus or bind TFBSs, and miRNAs also undergo processing from primary miRNA into precursor miRNA then finally mature miRNA capable of exacting regulatory roles. This means that the control of processing adds a further handle to control tool effects such as output expression. For example, TFs may bind repressor proteins in the cytoplasm which prevents their nuclear localisation and therefore their effects on the genome (Gao et al., 2022).

TFs and miRNAs can also be used as modules of larger, more complex circuits, regulatory networks or loops (**Figure 1.9.d**) resulting in different expression profiles and control.

Another favourable feature is that both miRNAs and TFs are tuneable by the quantity expressed. This tuneability can also be extended to different target sequences with different binding affinities, displaying the flexibility of targets and responses. When tools are further transiently transfected into cells, this regulation ensures reversibility and flexibility of the engineering process for fine-tuning and adaptation to intracellular conditions.

In the following chapters of this thesis, we will delve into the methodology, findings and discuss the implications of using miRNAs and TF regulation as next-generation cell and gene engineering tools with features that can be used to advance genetic regulation required for engineering biology applications with improved predictability, versatility, reversibility, scalability and reduced resource costs.

1.3 AIMS & OBJECTIVES

In the pursuit of advancing cell engineering tools with improved characteristics, this project is structured to address aims and objectives across three distinct result chapters and topics. Each chapter contributes a methodology towards the goal of refining tools for enhanced cellular control using bio-inspired approaches to miRNA and TF-based gene regulation. In all cases, our tool development is done considering resource demand as a design criterion, to better achieve predictability and reliable performance using these tools.

In Chapter 2, we lay the groundwork for genetic control in cells by developing methods to design and assemble DNA tools. We aim to design a new way to generate repetitive DNA sequences, inspired by the role natural genomic repeats have in native gene regulation. Our objective is to assess various aspects of DNA assembly methods (e.g. chassis, the time required for assembly, length reliability and variability of the repeats generated) to determine the best method for the custom design of repeat sequence elements. The best performance was achieved by developing an iterative assembly method we called CTRL, specifically tailored for generating repetitive DNA constructs for applications in mammalian cells. This work is essential for enabling the study of repeat functionality in the context of mammalian cell engineering. The ultimate goal of this chapter is to contribute to the advancement of synthetic molecular toolboxes for non-TU-based control of gene expression in mammalian cells.

In Chapter 3, we set out to display the uses of plasmids containing TF decoy sites for engineering gene expression in mammalian cells. The plasmids which contain TFBS repeats are developed using the CTRL method from Chapter 2. We develop the use of these plasmids in mammalian cells by observing whether increasing the concentration of TFBSs reduces the activity of TFs on sites required for the expression of GOIs. We further endeavour to compare the effects of different formats of decoy plasmids changing intracellular plasmid copy number or number of repeats per plasmid. We use these in mammalian cells for altering gene expression of outputs of circuits under the control of synthetic TFs, observing success by detectable changes in response curves due to decoy plasmid presence. Moreover, we use a similar workflow to show that TF decoy plasmids can be used to sequester native TFs required for directing cell fate or changing actions of another TF.

In Chapter 4, we set out to develop the use of miRNA in circuits for gene expression control. We initially explore using miRNA to repress gene expression outputs by post-transcriptional methods. We then also explore the use of miRNAs in the activation of gene expression by

altering transcription from DNA by control at promoters. We explore the response curves obtained from using these methods and try to improve characteristics such as dynamic range for improved useability in engineering biology applications.

Taken together, this thesis extends our knowledge of promising methods to manipulate gene expression in mammalian cells. Through the approaches taken here, we aspire to contribute valuable insights and methodologies to progress the landscape of cell engineering tools for use in mammalian cells.

1.4 NOTE TO THE READER

Publication related to this work:

Zouein, A., Di Blasi, R., Ellis, T. & Ceroni, F. (2021) Genetic Toolkits to Design and Build Mammalian Synthetic Systems. *Trends in Biotechnology*. 39 (10), 1004–1018.
doi:10.1016/j.tibtech.2020.12.007.

Manuscript in preparation:

Zouein, A., Ellis, T. & Ceroni, F. (2024) TF Decoy Plasmids as Competitive Inhibitors in mammalian cells.

2 ASSEMBLY OF HARD TO CLONE REPETITIVE SEQUENCES

2.1 INTRODUCTION

Repetitive DNA sequences, ubiquitous in the human genome, are an active subject of study due to their potential implications in various biological processes and diseases. Interest in repetitive sites spans many research topics from genome engineering to a foundational understanding of diseases. For example, many have tried to decipher the roles of genomic short tandem repeats (STRs), which occupy 5% of the human genome and have key roles in genome stability and gene regulation (Noskov et al., 2011; Horton et al., 2023; Depienne & Mandel, 2021). While repeats may be important for genetic tuning, altering expression in some instances by 70-fold, more is not always better and the expansion of repeats due to high mutation rates is implicated in many diseases (Depienne & Mandel, 2021; Paulson, 2018).

There is a wealth of toolkits and workflows designed to facilitate the assembly of mammalian TUs for the expression of circuit components, unlocked by enzymes and temperature cycling which we have previously reviewed (Zoueiri et al., 2021). These toolkits reduce building time for faster testing by outlining specific procedures and providing characterised parts. For example, Golden Gate toolkits such as the EMMA toolkit (Martella et al., 2017), use the power of Type IIS restriction enzymes, which cut at specific distances from their recognition sites, in a 'one-pot' reaction which cycles between digestion and ligation steps to assemble transcriptional components and units in a predefined order.

Given the various roles of repetitive sequences play including gene regulation, there is a need for specialized toolkits for reliable and efficient assembly. The existing toolkits for the assembly of TUs provide a promising starting point for the development of such specialized toolkits. Yet, less is currently available for the development of plasmids with a defined number of repetitive units. This is because the assembly of repetitive units is associated with higher chances of mutation and recombination which requires careful consideration in assembly strategies (Guye et al., 2013). Recombination, a method of DNA repair, causes the exchange of DNA between similar or identical DNA sequences, plentiful in repetitive plasmids (Li & Heyer, 2008). To

reduce the effects of this in cloning methodologies, the development of recombinase-free bacterial strains (DH5 α , DH10 β , NEB Stable) to reduce the impacts of recombination on the assembly of required plasmids is now commonly used in laboratories (Clark, 1973). Other expansions or deletions of repetitive regions are induced by cellular replication machinery during DNA propagation which has trouble replicating repetitive sequences leading to “replication slippage”, which produces products with additional or reduced number of repeat units (Richards, 2013).

A toolkit for assembling repetitive DNA sequences would speed up experiments and take into consideration the specific difficulties with working with these systems, such as the higher chances of mutation and recombination (Chu et al., 2011; Wan et al., 2020; Horton et al., 2023).

Another challenge that a toolkit for the assembly of mammalian repetitive DNA must consider is the eventual loss of validation of assemblies using short-read and long-read sequencing. A common way in which scientists confirm and validate the success of rounds of cloning is through the sequencing of assembled products. The emergence of long-read sequencing services has improved the capacity for resolving repetitive sequences compared to short-read sequencers (De Roeck et al., 2019). But short-read sequencers, which rely on assembling short-reads into a final sequence, still provide better base-pair accuracy compared to long-read sequencers (Oehler et al., 2023).

To date, few toolkits for repeat assembly exist which facilitate investigation of specific repeat types (

Table 2.1). For up to 20 repeats iterative capped assembly has proven an interesting approach for assembly of TAL (Transcription Activator Like) proteins. Here, repeats are added sequentially to a growing chain attached to a surface (Briggs et al., 2012). Other approaches multiply sites using Type IIS enzymes in cycles reaching repeat lengths of up to 500 bp (Scior et al., 2011) or 16 repeats (Pollak et al., 2019) sufficient for their poly-glutamine or plant circuit applications respectively.

Further improvements on repeat assembly tools would make a wide variety of repeat lengths, be easily adapted for specific applications across polypeptide repeats, genome part assembly (centromere, STR, telomere) and gene expression control while also being backwards compatible with other toolkits to allow diversity of application.

Although DNA printers are promising to significantly improve the field of synthetic biology, reducing the need for laborious rounds of cloning at a decreased cost per synthesized base pair, these companies are often unable to synthesize long sequences including repeats at the lengths required as many rely on hybrid approaches where at least part of the process relies on classical cloning and its associated drawbacks (Hoose et al., 2023).

By unlocking higher-throughput reliable methods to assemble these regions, the significant bottleneck of repeat assembly can be overcome, for many studies including engineering regulation of native or synthetic genes. This chapter is dedicated to the development of a methodology capable of assembly of repetitive sequences. Utilizing readily available starting materials, these methods aim to construct repetitive plasmids with no constraints on size, or sequence while ensuring accuracy, efficiency and simple validation. We will describe how these techniques have been previously employed and discuss our advancements and modifications for the assembly of repeat units. Lastly, we will describe our development of a toolkit designed for the streamlined construction of repetitive plasmids for various applications and our verification processes to ensure stability in light of current limitations.

The effectiveness of these methods is assessed based on several key aspects. Primarily, the capacity to produce intact, stable plasmids with repeats up to 10 kb, without loss of repeat content due to recombination or mutation. Additionally, the method's capacity to accommodate various repeat designs to meet diverse needs by amplification of an initial unit. Finally, the method's simplicity in design, ease of implementation, and duration to perform to create plasmids with a variety of repeat numbers.

Table 2.1: *Repeats occur in several natural settings such as peptide repeats, plant circuits, STRs and centromeres. These differ by the length of the base unit, how many repeats and total length of the repeat.*

Type	Repeat Size (bp)	Number of Repeats	Total Length (bp)
Peptide Repeats (polyQ)	3 (1 codon)	10-200	Up to 600
Plant Circuits	1600	16	up to 25,600
STRs	1-6	5 -50	up to 300
Centromeres	177 (α -satellite)	1920-27000	340 Kb-4.8Mb

2.2 RESULTS AND DISCUSSION

Workflows for the assembly of repetitive methods, like methods for assembly of expression cassette(s) included an assembly step which produces the sequence of interest (**Figure 2.1.a**), a propagation step where assembled DNA is introduced into a microorganism for selection and replication which improves plasmid yield (**Figure 2.1.b**) and a final validation step to confirm the success of the process (**Figure 2.1.c**). Many techniques can be adapted for the assembly of repetitive sequences. The methods we trialled can generally be subdivided by the scale at which the method generates repeats, either library-based approaches that generate a varied number of repetitive sites at once, or iterative approaches that multiply repeats in cycles/stages. These methods vary in time to perform, repeat number variability and host for propagation used. Quite often there is a trade-off between the time taken to assemble a certain number of plasmids and the specificity of yielding a certain number of repeats.

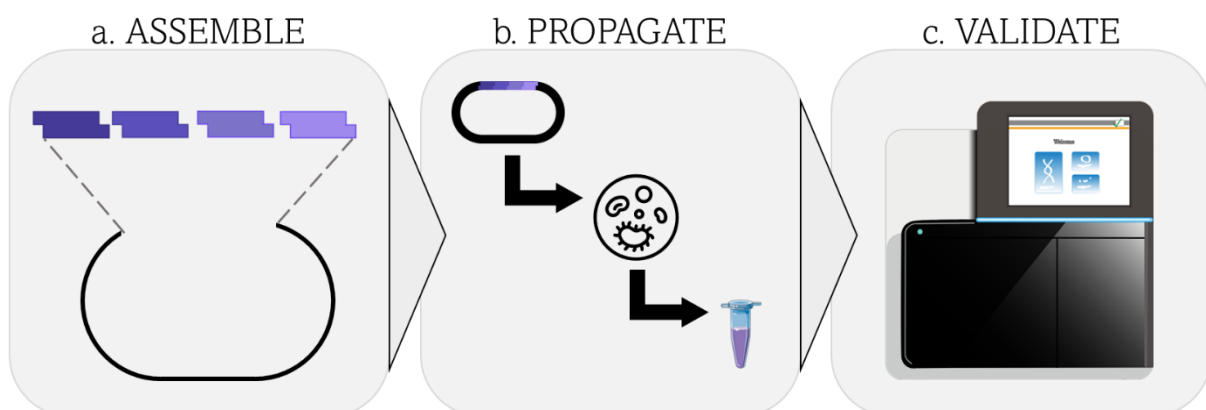


Figure 2.1: Common molecular biology workflow for the assembly of desired DNA sequences for engineering biology. **a.** Engineers can choose from a suite of assembly toolkits which combine standardised parts using a predefined workflow. **b.** The assembled plasmid is then propagated in a microorganism for selection and to increase DNA yield. **c.** The assembled DNA is then validated for correct DNA sequence using short-read or long-read sequencing or test restriction digests. This process yields DNA plasmids of interest that can then be used in experiments of choice.

2.2.1 LIBRARY-BASED APPROACHES TO REPEAT ASSEMBLY.

Library-based approaches are designed to produce more than one plasmid product at a time, each with varying repeat sizes. The assembly methods we trialled can be summarised by two key stages. The first stage of the assembly generated repeats while the second added repeats to a destination vector compatible for amplification and selection in the host of interest. In the first step, the various methods combine the required ingredients to generate a pool of repeats in series (concatemers) with the correct context for the second stage where the repeats are assembled into complete plasmids. The destination plasmids contained regions to facilitate selecting correctly assembled clones (also used to estimate success rate) and regions required for their propagation (for example resistance cassettes).

These methods promise a high throughput method of generating a variety of plasmids at once, a beneficial characteristic of engineering biology. The lack of iterative cycles may also be favourable as it reduces costs, time to implement, and chances of errors and mutations developing and being propagated/dragged on throughout other cloning stages. We tested three methods of repeat generation namely rolling circle amplification (**Figure 2.2**), redundant annealed oligos ligation (**Figure 2.4.a**) and sequential non-template PCR (**Figure 2.6**).

Initial repeat units used within this section contained 3 repeat units with 2 unique spacers (**6.5 List of Materials Used, P.1536.5**). These were synthesized as duplexed oligonucleotide pairs with single stranded overhangs complementary to the following oligonucleotide pair. The limitations on the length and complexity of the DNA sequence are therefore determined by the synthesis method in this case by the manufacturers' sequence complexity score.

2.2.1.1 *Rolling Circle Amplification (RCA).*

Rolling circle amplification (RCA) is an isothermal reaction where a primer, enzymes and NTPs, are used to amplify a circular template resulting in products which are concatemers of the original template. This method has been established for many applications including as a signal amplification module for biosensors (Ali et al., 2014). Our method for assembly of repetitive sequences used RCA to generate an array of concatemer products that was then incorporated into a plasmid using homologous recombination in yeast, a process often termed Transformation Assisted Recombination (TAR) (**Figure 2.2**). The preference for yeast as a host in extensive repeat assemblies (up to 140 kb) for incorporation into human artificial

chromosomes (Ebersole et al., 2005; Noskov et al., 2011), is likely due to enhanced recombination efficiency and simplicity in handling using well-established protocols.

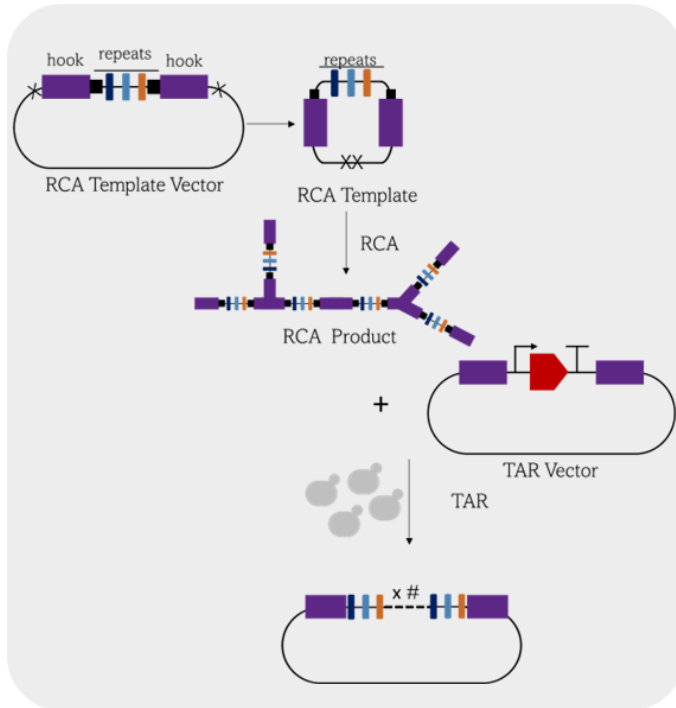


Figure 2.2: Depiction of key steps of repeat generation using Rolling Circle Amplification (RCA) and Transformation Assisted Recombination (TAR) in yeast. Library-based approach to the assembly of repetitive sequences using RCA for repeat production and yeast-assisted recombination for propagation. An in-house built RCA template vector is used to insert DNA repeats into the right context. The RCA template vector is then used to generate the RCA template, a subsection of the original vector by digestion of flanking restriction sites depicted as X. The template is combined with the rest of the required components for RCA. The first step of RCA denatures the DNA template so that random primers can bind to and extend the template into a branched product such as the one depicted in the figure. The RCA products and TAR vector are co-transformed into yeast where recombination can occur and plasmids with different lengths of concatemers are formed. After these plasmids propagation and selection in yeast they can be transformed into bacterial strains.

We therefore similarly used yeast to generate DNA repeats. Our adapted strategy was based on a paper where this method was used to assemble artificial centromeres to understand centromeric repeat function in mammalian cells (Noskov et al., 2011). Within this paper the authors described a method of amplifying initial small DNA fragments, about 100 bp in size, to generate products up to 140 kb in size in a process that takes 2 weeks to perform and

complete. These characteristics made this a promising avenue to pursue in the rapid development of many repetitive plasmids.

Our workflow followed the paper's instructions with the following modifications. Firstly, the repeats are synthesized by oligo annealing or DNA synthesis and incorporated into an RCA template vector (AZ_PL_81) which introduces the repeat of interest into the right context, with flanking hooks for recombination into the destination vector. The repeat unit number which is 3 is dictated by the oligonucleotide providers restrictions where the maximum possible size is 60bp. The hooks were chosen to be 100 bp long and unique from sequences within the yeast and human genome to reduce chances of off-target recombination, especially into the yeast genome. The hook sequences chosen did not show similarity with these genomes when searched in BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Other than the three repeat units, and hook sequences, the RCA template vector, designed in-house, also contains restriction sites for the same enzyme, denoted as X in **Figure 2.2** and **Figure 2.3.a**. These restriction sites, flanking the repeats and hooks, are used to circularise and create the RCA template absent of any sequences we do not wish to concatemerize such as those for antimicrobial resistance or replication. The circularised RCA template which is 350bp is confirmed by running on a 2% agarose gel, where the circularised template (**Figure 2.3. b**, well 3) runs slower than its linear equivalent (**Figure 2.3.b**, well 2). Once the generation of the RCA template is confirmed, the circularised template is subsequently purified on a column (QIAGEN) and eluted in 20 µL nuclease-free water. The template is then introduced to the rest of the RCA reaction components (TempliPhi™) as outlined by the manufacturer (Cytiva). Briefly, the RCA template is denatured at a high temperature. The single stranded circular templates are then able to anneal to random primers and extend from these locations. After this, a small subset of the reaction product was run on a gel (**Figure 2.3.c**) to observe the success of the reaction by a smear which are indicative of different repeat sizes generated. The rest of the RCA products were diluted and combined with the TAR vector (AZ_PL_82) which contains the same hooks (denoted as purple boxes in **Figure 2.2**) flanking a drop-out cassette for selection. This mix is transformed into competent yeast strains BY4741 and VL6-48N, for recombination to occur in yeast. VL6-48N is hypothesised to have enhanced recombination efficiency compared to BY4741 which is the community standard strain. White clones which have recombined to exclude the drop-out cassette are picked from transformations into both strains, plasmid DNA was then isolated from yeast colonies and transformed into competent DH5α bacterial strain to increase plasmid DNA yield. The procedure took about two weeks to complete.

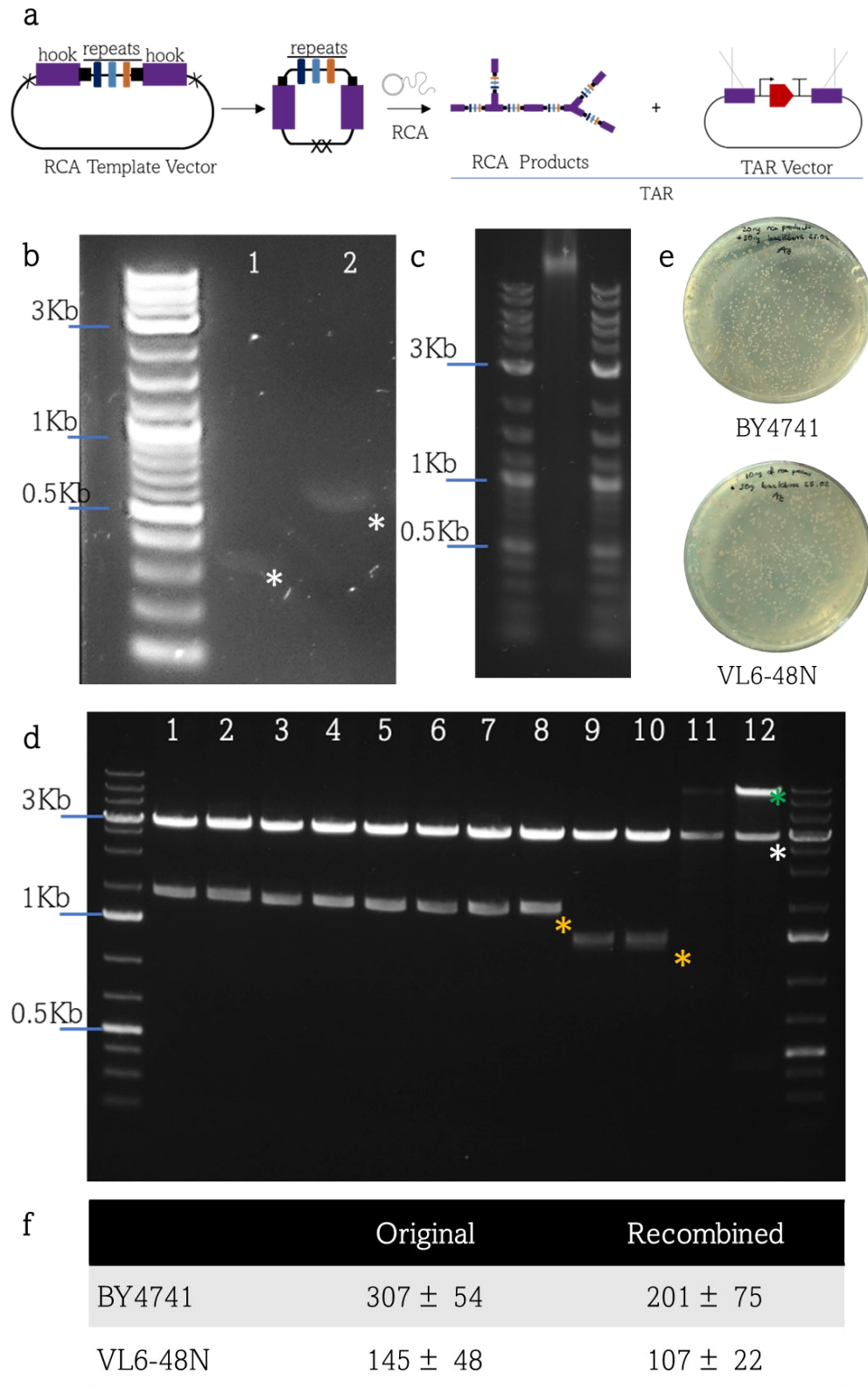


Figure 2.3: RCA TAR cloning key stages and analysis **a.** Repeats are introduced into an RCA template vector before miniaturisation into the RCA template by digestion and ligation. The RCA template combined with the RCA reaction components to form the branched RCA product which can then be used in transformation-assisted recombination into yeast. **b.** The RCA template before (band indicated by * in well 1) and after circularisation (band indicated by * well

2) compared to 1Kb+ ladder with major bands annotated. **c.** Sample of RCA products on 1% agarose gel where a smear of varied sizes of repeat units was expected. **d.** Testing recombination of a stable repetitive construct in DH5 α (*E.coli*) in well 12 with backbone at 3Kb (band indicated on gel by a white *) and repeat fragment at 10Kb (band indicated on gel by a green *). This is compared to different colonies on the same transformation plate into BY4741 (yeast), wells 1 to 9 or VL6-48N (yeast) wells 10 and 11 which were recombined and therefore lost a number of repeat units (indicated by bands at the loci of the orange *). **e.** Typical results of TAR plates where blue colonies are those which did not recombine to include a repeat fragment while white colonies may have incorporated a repeat fragment generated by RCA. **f.** Typical results of TAR by colony numbers of original template or recombined vectors.

The repeat content of clones was then assessed by restriction digestion. BY4741 yielded more white colonies compared to VL6-48N (**Figure 2.3.f**), while VL6-48N yielded longer repeats (data not shown). Nonetheless, the largest repeat sizes we were able to sequence using this method was 3 repeats (roughly 300 bp in length) compared to the 140 kb sizes reported by Noskov et al. This may have been due to the reduction of the scale of the RCA reaction, our protocol's RCA reaction had a total volume of 10 μ L yielding around 1 μ g of DNA whereas the paper reported 100 μ L RCA reactions with more than 4 μ g of DNA produced. Moreover, our RCA products did not appear to smear much on agarose gel indicating there may not have been much variability in the sizes of repeats generated. Despite taking measures to reduce the chances of unwanted intermolecular and intramolecular recombination by including neutral spacers between repeats, the higher rates of recombination in these yeast strains may have led to the loss of the repetitive DNA in final constructs. This is further exemplified in **Figure 2.3.d**, where the repetitive plasmid TetOx320, containing similar spacers between TetO sites, was transformed into competent DH5 α *E. coli* where it is stable as well as into BY4741 and VL6-48N yeast strains. The plasmids were then purified and digested with restriction enzyme with recognition flanking the 320 repeats. While BY4741 produced more colonies in total, a similar product is observed in 8 out of 9 clones picked, where predicted repeat fragment size was reduced from 10 kb mark down to 1.2 kb (**Figure 2.3.d** well 1-8). There were fewer VL6-48N colonies to pick from and of these one showed the repeat fragment at 1 kb (**Figure 2.3.d** well 10). The backbone remains stable under all conditions at around 3 kb proving the repeats are prone to excision out of the plasmid likely by recombination.

Our initial approach fell short of achieving the desired repeat sizes, generating only 3 repeats instead of the kilobase scale. Additionally, we were unable to produce plasmids of varying repeat lengths, likely due to recombination in the yeast strains used. These challenges led me to investigate alternative library-based strategies, which I will describe in the following sections.

Given that the RCA method of library preparation, which is reliant on recombination, may result in the loss of repetitive elements, we decided to transition to bacterial hosts lacking recombinases for our subsequent attempts at generating repeat libraries. The following explored methods differ in their method of generating repeats but not in the formation of the complete plasmid stage.

2.2.1.2 Redundant Annealed Oligo Ligation:

The next approach for repeat array library formation used the ligation of annealed oligos containing repeat units and redundant overhangs (**Figure 2.4.a**). This method promised a quick way to create a library of repeats with low design complexity. For this, 4 pairs of oligonucleotides were designed such that the internal pair contained compatible overhangs (top panel **Figure 2.4.b**). In the initial step of the process, the internal pairs are annealed, phosphorylated, and ligated for different times and concentrations to promote the correct formation of different lengths of concatemers from the original oligonucleotides containing repeat units (**Figure 2.5**). The second step then ligated head and tail duplexed oligonucleotides that contain restriction sites for digestion and ligation into a compatible destination vector. The destination vector contained a drop-out cassette which expressed bacterial RFP, such that if the vector had not taken up a repeat array the colony would appear red when cultured.

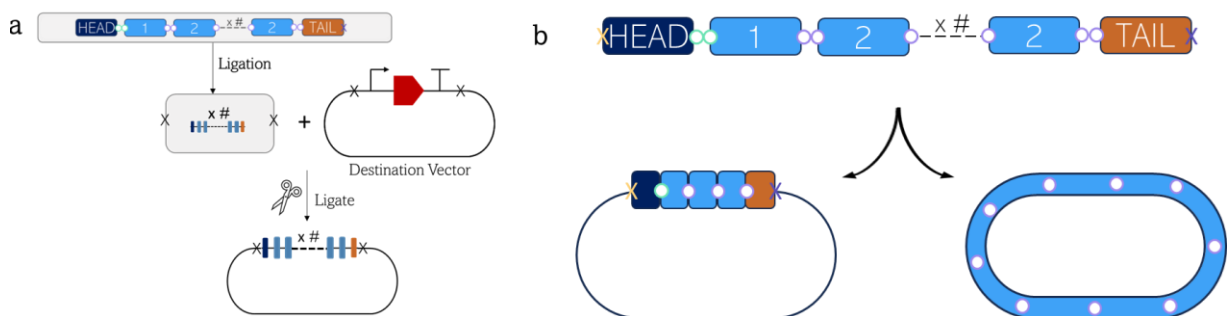


Figure 2.4: Depiction of steps and expected results of repeat plasmid assembly using redundant oligonucleotides. **a.** Oligonucleotides with redundant overhangs are annealed and ligated to create a pool of repetitive sequence lengths. Finally, the pool of repeats is ligated into a bacterial destination vector for propagation in bacteria. **b.** Oligonucleotide pairs made up of a head, tail or two internal oligonucleotides with redundant overhangs labelled 1 or 2 are designed. The ligation of the duplexed oligonucleotides can generate linear concatemers with flanking head and tail oligonucleotides as required (bottom left) or the internal duplexed oligonucleotides can bind to each other forming circular concatemers on their own (bottom right).

A disadvantage of this approach which likely reduced the plasmid diversity obtained by this method was the chance the redundant overhangs ligated to produce circular DNA products unable to bind the head and tail oligos and subsequently unable to combine into the destination vector required for its propagation in bacteria (**Figure 2.4.b**). To encourage intermolecular ligation different concentrations of oligos in the initial ligation were trialled (**Figure 2.5**).

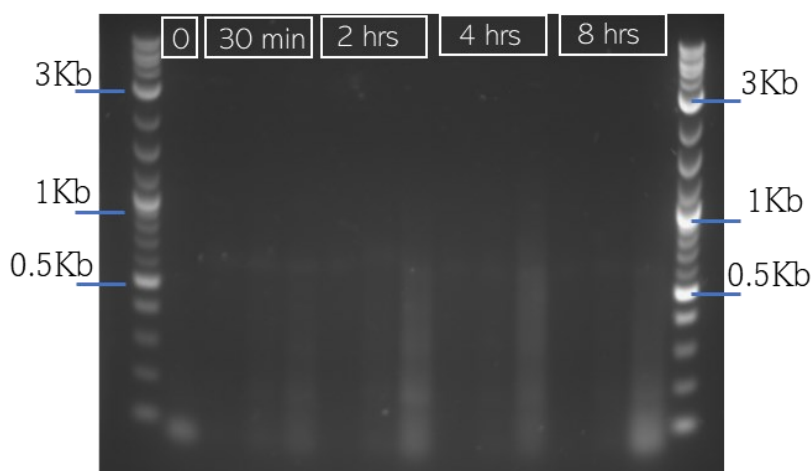


Figure 2.5: Agarose gel image showing resulting sizes of ligated oligos ligated for different durations and concentrations to increase the proportion of linear concatemer products over circularised concatemer products. Each set contains ligation products of reactions containing 0.2µg, 0.4µg or 1µg of ligated oligos in succession.

Despite testing various conditions oligos did not combine sufficiently to produce the diversity of overall sizes we required, a size range of between 1 kb and 10 kb. For each time point (30 minutes 2 hours or 4 hours of ligation) 0.2, 0.4 or 1 µg of total oligos was used per reaction but none of the conditions resulted in increasing length of repeats, hinting at a preference for forming more circular products.

Another consideration that may have reduced the efficiency of this procedure may be the accuracy and length of purchased oligonucleotides. Using our provider oligonucleotides over 40 nucleotides make up less than 80% of non-purified products and further purification is recommended to improve this (Nan Pazdernik, 2015).

Furthermore, once the process concludes, it offers no avenue for expanding repeats. So, while this method is one of the fastest tested, with the wet-lab component taking three days to complete, short repeat numbers and no method to further improve diversity in subsequent steps are significant limitations to this approach.

Therefore, the method for repeat array library formation through ligation of annealed oligos promises a quick way to create a library of repeats but faces limitations such as the formation of circular DNA products and insufficient diversity in overall sizes.

2.2.1.3 Non-template PCR.

The last method attempted for the assembly of repeats with a library preparation stage was using non-template PCR. The non-template PCR method was adapted from Chu et al., 2011, where non-template PCR was used for the creation of polypeptide repeats more than 2 kb in length.

The procedure consisted of a minimal sequence of four PCR reactions (**Figure 2.6, Table 2.2**). The first PCR combines initiation primer and elongation primer with PCR reagents to form a double stranded DNA sequence containing initiation primer and elongation primer sequence. The second PCR then used the products of the first PCR reaction as a template along with elongation primer. With each cycle, the elongation primer can bind its complementary sequence within the template and extend the number of repeat units which are depicted as the orange fragment in **Figure 2.6**. The third PCR encourages the formation of products with more varied lengths and combines the PCR product from the second PCR as template along with the elongation primer. The fourth and final PCR uses the third PCR product as a template along with a termination primer. This primer has homology with the template and upon hybridisation and extension during PCR a product containing a restriction site annotated as X in the figure is formed. This collection of double stranded DNA containing varied repeat lengths and flanking restriction sites is then digested and ligated into a compatible backbone for bacterial propagation. Our adaptation combined standard Phusion (NEB) reaction components as this polymerase has 50 times less chances of errors compared to Pfu polymerase (used in the paper) with cycle numbers from the paper (**Table 2.2**) as these were optimised to encourage the generation of DNA with various lengths. The use of a polymerase with better fidelity is essential as unlike the paper, where the authors chose different codons to reduce sequence similarity, we must maintain the DNA sequence and its similarity for our goals. This made this approach challenging because the final products are prone to nonspecific annealing or the formation of secondary structures (Chu et al., 2011). A schematic of reaction components and cycling conditions used are summarised in **Table 2.2**.

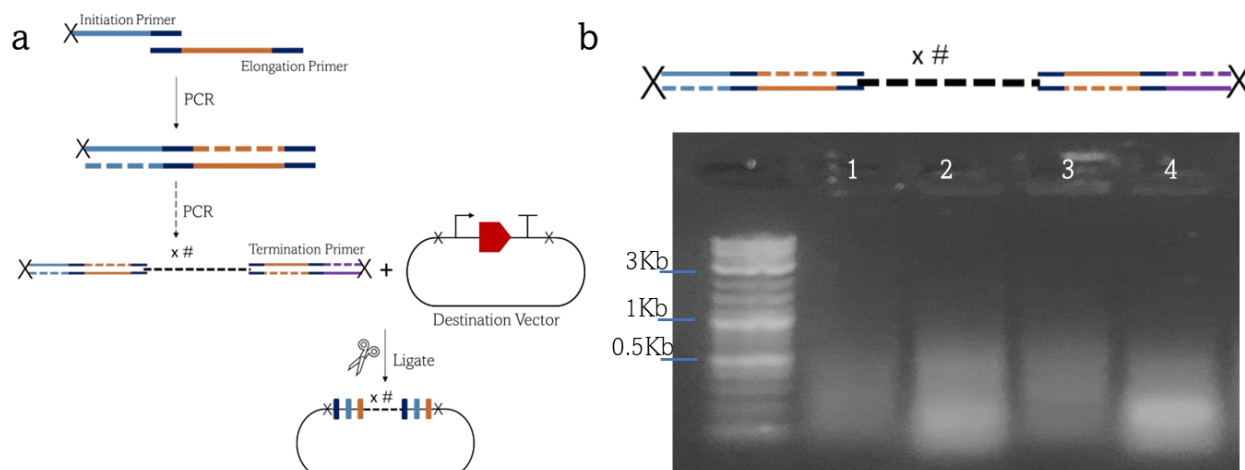


Figure 2.6: Library-based approaches to the assembly of repetitive sequences using non-template PCR **a.** Non-template PCR is used to generate the repetitive sequence pool through an initiation PCR, extension PCR (which we repeat twice) and lastly a termination PCR. Solid lines indicate primers whereas dotted lines indicate extension products. The initiation and termination PCRs added restriction sites which are cleaved and assembled into a bacterial destination vector replacing a drop-out cassette **b.** Portion of PCR products from non-template PCR workflow for repeat generation. The first well contained the PCR with initiation and elongation primers. The second and third wells contained PCR products using the amplicons from the previous well and the elongation primer. The fourth well contained the products of the 4th PCR which uses 3rd PCR as the template along with termination primer.

Table 2.2: Summary of non-template PCR primers and conditions used.

	PCR #1 (Initiation PCR)			PCR #2, #3 (Elongation PCR)			PCR #4 (Termination PCR)	
Reaction Components								
	1x	95°C 2 min	1x	95°C 2 min	1x	95°C 2 min		
Cycling Conditions	30x	95°C 1 min	5-15x	95°C 1 min	30x	95°C 1 min		
		55°C 1 min		55°C 1 min		55°C 1 min		
		72°C 2min		72°C 2min		72°C 2min		
	1x	72°C 2min	1x	72°C 2min	1x	72°C 2min		

The first reaction combined an initiation primer containing a restriction site upstream and homology downstream to an elongation primer. The PCR is expected to extend from the region

in common forming double-stranded amplicons. The subsequent 2 PCRs combined the purified product from the previous PCR as the template along with the Elongation primer, to create longer repeats. The last PCR took the purified product of the previous PCR as a template and used a termination primer with homology upstream and a restriction site downstream. This reaction was used to introduce a restriction site for subsequent digestion of products and ligation into the receiver vector.

A portion of the products of each PCR in order are shown in **Figure 2.6. b**. While the repeat lengths appeared to be longer than those resulting from the previous method the target size of repeats did not significantly change between cycles. This suggests that the method may not be as effective at generating various repeat lengths, a key requirement of our adopted method. Moreover, other challenges of using this approach were the significant dependence of the method on PCR which may have introduced significant mutations, especially at higher repeat lengths. This would have become a bigger limitation as the capacity to verify products reduced with longer repeat lengths and identification of mutations would have been difficult. The method also had a more difficult design stage especially when scarless repeats are required making it difficult to implement for some repeat applications such as in the study of genomic repeat expansion.

Therefore, while the non-template PCR method, adapted from Chu et al., 2011 had the potential to generate various repeat lengths, it faced challenges such as nonspecific annealing, formation of secondary structures, and significant mutations at higher repeat lengths complicating its use in the assembly of repetitive DNA.

2.2.1.4 Comparing Library-based Approaches.

The various approaches in this section contained a step that generated various repeat numbers. The different methods can accommodate initial repeat units of any sequence, but user must consider the best source and method of generation of their repeat unit. For example, PCR of a long repetitive sequence might lead to more mutations than if this were to be excised from an available sample containing the initial repeat by restriction digestion. Therefore, the methods allow for choice of initial unit sequence based on convenience. The initial repeat unit architecture was identical in these methods but varied in the additional flanking sequences for incorporation into a final plasmid. For the RCA based method this was hooks for recombination while for redundant annealed oligo ligation and non-template PCR this was restriction sites for ligation into a compatible backbone. The methods did not provide a method of assembling intact

varied lengths of repeat units that were backward compatible, some of the key features we require.

2.2.2 ITERATIVE APPROACH TO REPEAT ASSEMBLY: DEVELOPMENT OF CTRL

In the preceding section, we explored various library-based approaches for repeat assembly. The methods, while promising, faced challenges in achieving the desired repeat sizes and diversity required. Given these challenges, we turned our attention to iterative approaches to repeat assembly. These methods aimed to address the limitations encountered in the previous techniques and enhance the assembly of repeats by encouraging the assembly of repeat plasmids one at a time.

Iterative enzymatic approaches, where gene assembly is performed in cycles of cloning to achieve longer or more complex systems from gene part modules, confer benefits to these assemblies. Most predominantly, when compared to the above methods for repeat assembly where many products are sought after, iterative assemblies seek one reaction product, and reactions are set up to go to completion resulting in a product that is rich in one specific, defined product. This contrasts with library methods which yield a mixture of products. While the assembly of repeats with iterative approaches is slower, where a new repeat plasmid will take about three days to assemble, they provide more reliably a predetermined result. This is because there is no stage where multiple repeat lengths are generated, instead, the repeats are domesticated with overhangs and multiplied within each cycle. Furthermore, the use of recombinase-free bacterial hosts in these assemblies further reduces the chances of recombination. These design choices mean that when verification - often through sequencing - is lost at higher repeat sizes, restriction digests may be sufficient to verify product and cloning success. While iterative approaches are commonplace for the assembly of genes, less is available for the assembly of DNA repeats for use in mammalian cells.

In this direction, we built a method for the iterative assembly of repeats for mammalian cells. We “lovingly” termed this method CTRL (Cloning Troublesome Repeats in Loops). CTRL multiplies repeats in cycles, either by a factor of two (through BioBrick prefix and suffix) or by a factor of 4 using Golden Gate principles (**Figure 2.7**). CTRL allows customisation to facilitate the purposes of the researcher, allowing changing backbones, repeat lengths and linkers.

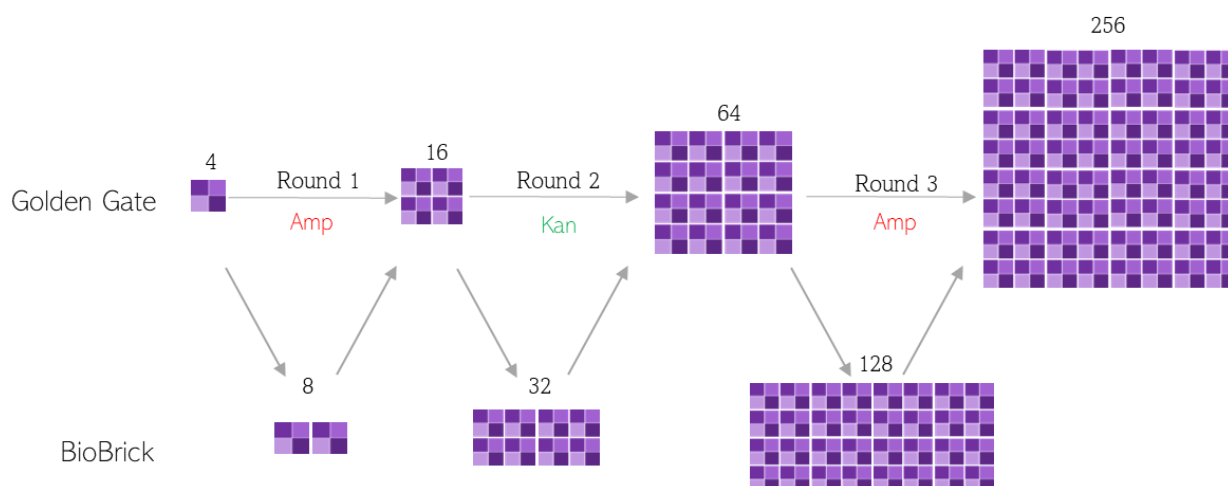


Figure 2.7: CTRL allows assembly of repeats using Golden Gate and Biobrick iterative techniques. After their domestication into Ampicillin backbones restriction sites for Biobrick can be used to multiply repeats by 2 or Golden Gate restriction sites used to multiply the repeats by 4.

2.2.2.1 Design & Validation of CTRL

The strategy begins with the design stage, where it is crucial to ensure that the DNA sequences do not contain the restriction sites used in this method, namely Bsal, Bsmbl, and Bbsl. This is because their presence can impede the chances of obtaining the desired assemblies. The designed DNA repeat sequences of interest, which can be sourced by PCR, oligonucleotide annealing, or by g-block synthesis, must be designed with flanking overhangs A and B (**Table 2.3**) for introduction into ampicillin backbones 1 through 4. Following this, the process of domesticating DNA repeats into ampicillin backbones 1 through 4 takes place where four Golden Gate reactions are set up using internal Bsmbl sites, as shown in **Figure 2.8**.

In our experience, the most successful sources for domesticating DNA repeats have been PCR or g-block, consistently resulting in products 100% of the time. This initial integration of repeats ensures the efficient execution of the rest of the strategy.

Table 2.3: Summary of 4 base pair overhangs revealed during CTRL. Overhangs after digestion by Type IIS enzymes and respective fidelity of ligation into receiver vector.

Enzyme		Overhangs					Fidelity
Kan	Bsal	GCAA	CAGA				100%
	BsmbI	TCAG	TAGT	CCTT	GCGA	TGCT	
Amp	Bsal	GCAA	TTAC	GCTC	TCCT	98%	
	BsmbI	CAGA	AGCA	CTGA			
MTK	BbsI	CTGA	AGCA				100%

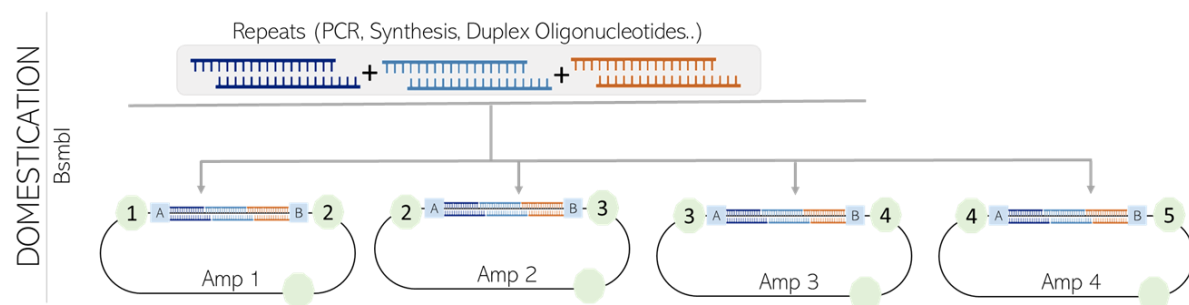


Figure 2.8: Domestication of repeats into CTRL. Repeats were obtained by synthesis or PCR and designed to have overhangs A and B compatible with ligation into Amp backbones 1 through 4. Domestication introduces repeats into the correct context for further cycling and multiplying of repeat units as required.

Once domesticated the repeats gain Bsal sites (annotated as 1 and 2 in **Figure 2.8**) compatible with assembly into Kanamycin backbones such that inserts of Ampicillin backbone positions bind together and to internal Bsal sites in Kanamycin backbones (**Figure 2.9**, top grey block). The resulting Kanamycin plasmids, each from a separate Golden Gate reaction, had four times the number of repeats and Bsmbl sites available for the next cycle into Ampicillin (**Figure 2.9**, bottom grey block).

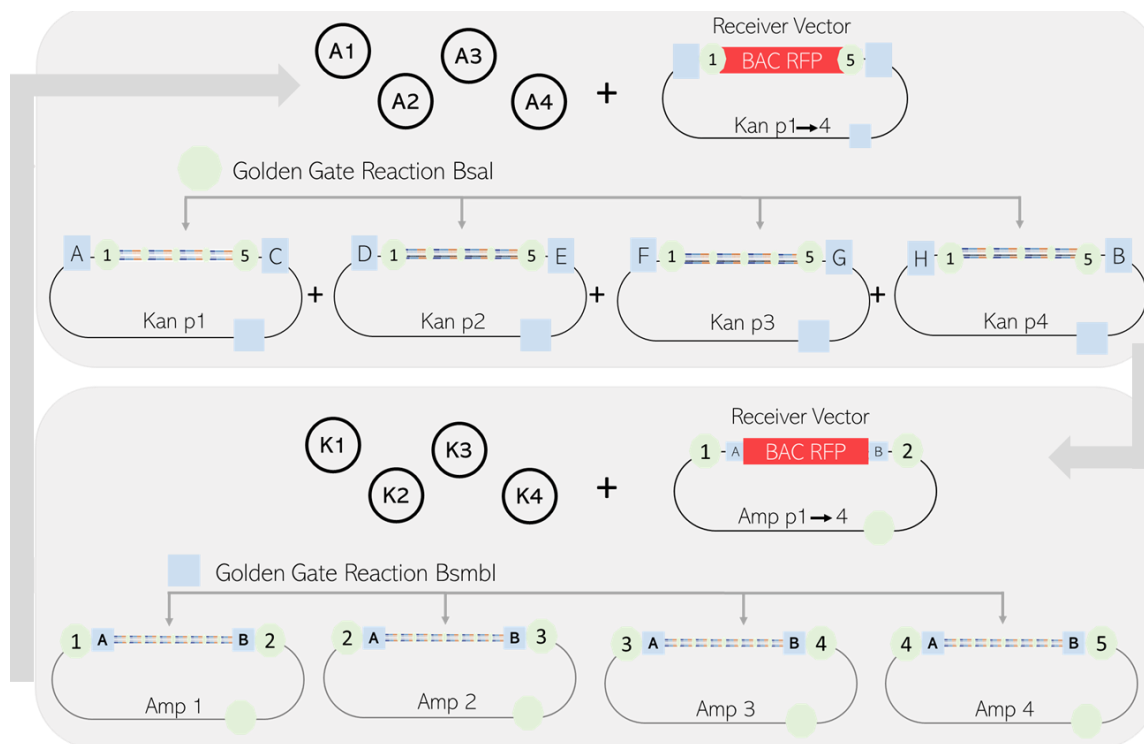


Figure 2.9: CTRL Golden Gate functionality. Once repeats were domesticated into the system, plasmids can be used in Golden Gate reactions where parts in Ampicillin backbones assemble into Kanamycin receivers and parts in Kanamycin backbones assemble into Ampicillin receivers replacing transcriptional unit coding for the strong expression of RFP in bacteria.

Therefore, Ampicillin positions act as assembly parts into Kanamycin receivers, using Bsal sites and Kanamycin positions act as assembly parts into Ampicillin receivers using BsmBl sites. With each Golden Gate cycle, the repeat number is multiplied by four. The eight plasmids (Amp and Kan receiver vectors) were designed with maximal chances of ligation fidelity of overhangs (Pryor et al., 2020). The optimised chosen overhangs and associated predicted ligation fidelity are listed in [Error! Reference source not found.](#). By applying Golden Gate methodology to repeat assembly, CTRL benefits from scalability, automatability, and efficiency. Four backbones were chosen as we believe this improves toolkits characteristics. This is because the more backbones the fewer cycles required to generate larger repeats, but repeats must assemble into all positions to move onto the next golden gate cycle. To assist with making repeats with intermediate sizes we also include Biobrick prefixes and suffixes in all eight backbones.

As all backbones contain Biobrick prefix and suffix, a plasmid domesticated into CTRL can be digested with EcoRI + SpeI as well as PstI + XbaI and ligated together along with a backbone digested with EcoRI and PstI, to produce a plasmid with double the repeat number separated by a scar (denoted as /) formed from the ligation of XbaI and SpeI overhangs. This scar sequence

is no longer cleavable by the same enzymes, preventing the separation of repeats (**Figure 2.10**).

Moreover, Biobrick parts can also be used for the addition of transcriptional units that may be required in applications. For example, the addition of a transcriptional unit coding for the expression of a constitutive reporter as a marker of transfection efficiency (**Figure 2.12**). To further improve the utility of the resulting plasmids, ligation into MTK backbones is also possible. This enables users to combine repeats as an MTK part into any MTK backbone of their choosing. To date the MTK backbones available facilitate specific experimental requirements such as integration for mammalian cell delivery and can be purchased from Addgene.org (Fonseca et al., 2019, 2020). Therefore, we have designed flexibility into this toolkit to allow users to combine repetitive inserts with other characteristics of their choice.

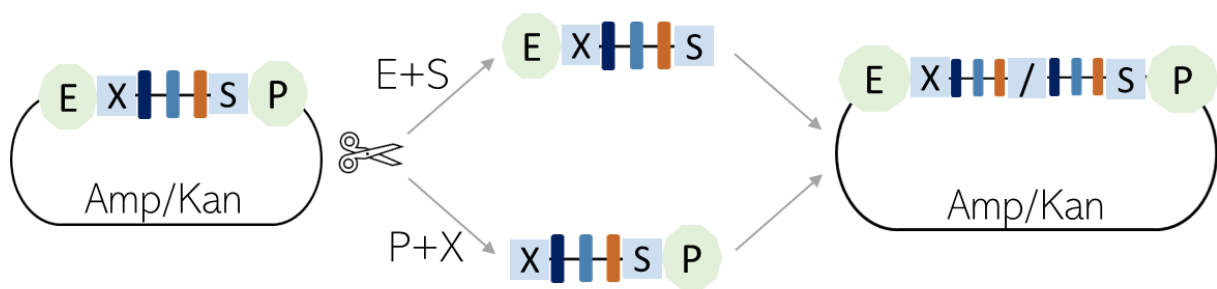


Figure 2.10: Use of the Biobrick function of CTRL. Repeats domesticated into any CTRL vector can be digested using *EcoRI*+*SpeI* (E+S) and *PstI*+*XbaI* (P+X). The two digestion products can be ligated together along with a receiver vector digested with *EcoRI*+*PstI* (E+P) to form a new plasmid. The ligation of *XbaI* and *SpeI* overhangs forms a sequence not digested by either of these enzymes, denoted as “/” in the figure.

After assembly, products were transformed into competent recombinase deficient bacterial strains (DH5α) where assembled colonies are distinguished from the receiver vector by the absence of the drop-out cassette, which codes for strong expression of RFP in bacteria (**Figure 2.9**). This is such that red colonies have not taken up repeat inserts and were not selected for further validation. We used this system to assemble 3 different repeat sequences from 4 initial repeats to 256 repeats (4^4 , over 10 kb in size) using CTRL’s Golden Gate and Biobrick functions. Biobrick assembly yielded 100% correct assemblies, especially when a different resistance cassette is used as the destination vector. Golden Gate assemblies yielded 94% assembled colonies, with typical numbers of colonies per assembly in the different

positions listed in **Table 2.3**. There was no significant difference in the number of correct assemblies based on repeat sequences we tested, suggesting wide utility for many mammalian sequences. Moreover, CTRL can be used for the assembly of repeats without scars using Golden Gate functionality, unlike most methods discussed previously. To achieve this Type IIS overhangs outlined in **Error! Reference source not found.** should be changed to fulfil sequence restraints, although this will likely alter ligation fidelity and should be checked using appropriate online tools. Biobrick functionality does insert scars, and therefore may not be suitable for applications requiring scarless assembly.

In conclusion, the designed CTRL strategy, with its iterative approach allowed flexible assembly of repetitive inserts with increased specificity and lengths. This presents a promising solution for repeat assemblies needed for a variety of mammalian cell studies, overcoming many limitations of previous methods and opening up new possibilities for research in various contexts.

Table 2.4: CTRL Cloning results. Counts of white colonies with correct assembly (number of repeats) using Golden Gate functionality. Averaged results from 3 independent assemblies of different repeat sequences.

KAN			AMP		
Position	Assembled	Percentage	Position	Assembled	Percentage
p1	26 ± 4	96%	p1	51 ± 23	95%
p2	31 ± 3	100%	p2	29 ± 7	86%
p3	57 ± 23	100%	p3	29 ± 10	82%
p4	36 ± 20	97%	p4	32 ± 3	93%

2.2.2.2 Characterisation of final repeat assembly:

Validation of repeats can often be problematic due to their repetitive nature. While short-read sequencing is usually the gold standard used to confirm assembled plasmids due to its accuracy, this method was not sufficient for sequencing any of the plasmids we assembled containing more than 32 repeats. This is in line with our Sanger sequencing provider's guide for sequencing purified DNA where the maximal expected read length is 1.4 kb. The limitations of this approach arise due to the dependence of the method on fragmented DNA for sequencing before assembly into a final sequence map. On the other hand, long-read sequencing such as those provided by Oxford Nanopore, is becoming more reliable, and accessible, and promises to overcome the shortcomings of short-read sequencing, enabling the sequencing of repetitive regions (De Roeck et al., 2019; Nanopore, n.d.). Using full plasmid sequencing service (Full circle labs) we were able to increase sequencing of repeats up to 5 kb. To date, with the available resources, we can sequence confirm up to 128 repeats built with CTRL consistently, with full reads of 258 repeats being sequenced about 70% of the time. Additionally, in contrast to RCA and non-template PCR methods previously discussed, the use of the CTRL method does not depend on enzymatic DNA amplification, significantly minimizing the likelihood of sequence mutations. This is evidenced by the fact that no deviations from the anticipated sequences were detected when CTRL assemblies underwent DNA sequencing. This means that test restriction digest with flanking restriction enzyme recognition sites can be used to confirm assembly sizes of obtained assemblies (**Figure 2.11**)

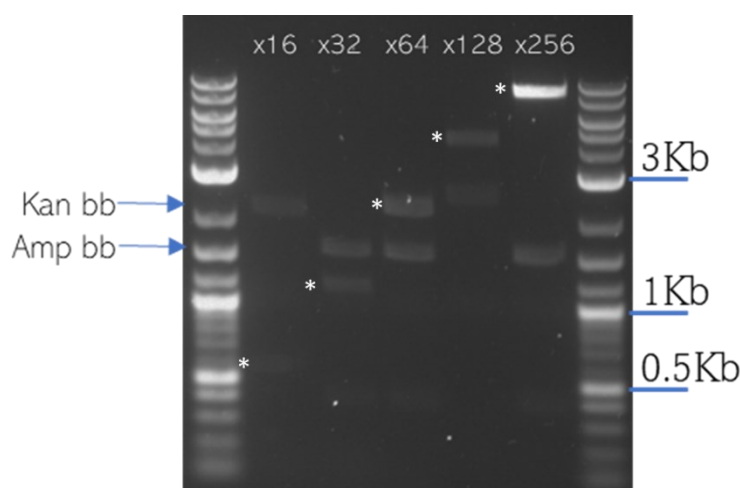


Figure 2.11: Restriction digest confirming the expansion of a repeat from 16 sites to 256 sites. Band containing repeats is indicated using *. This is facilitated by changing between CTRL backbones with Kanamycin resistance (Kan bb) or Ampicillin resistance (Amp bb).

2.3 CONCLUSION

Throughout this chapter, we investigated both library and iterative methods for the assembly of plasmids containing DNA repeats. These repeats could prove useful for various mammalian cell studies.

Library methods offered a promising avenue for DNA repeat assembly, as they quickly yield plasmids containing varied repeat lengths within one round of cloning. However, the majority of methods predominantly relied on polymerases, which increased the likelihood of errors due to their struggle with amplifying repetitive regions. This necessitates the need for accurate sequencing when using these methods. However, validating through sequencing presents its own set of challenges. Often, short and long-read sequencing are unable to provide accurate sequencing of repeats, in our hands large repeats were sequences only 70% of the time (data not shown). This leaves validation by restriction digest of transformed clones which can give a general idea of repeat size but not any small mutations. The field of sequencing is improving at an exponential rate, with long-read sequencing being of high interest. This is exemplified by Illumina's recently releasing a long-read product (Illumina, n.d.). With improvements in this field, we will further be able to consistently verify longer repeat regions.

Other considerations for the use of library methods include the formation of scars between repeats and based on our experience, did not yield the desired repeat sizes warranting further optimisations.

On the other hand, we have validated CTRL, a novel toolkit for the iterative assembly of DNA repeats. The CTRL protocol combined the principles of Golden Gate and Biobrick for the assembly of repeat arrays up to 10 kb without repeat excision by recombination. We anticipate that it could be used for even larger DNA array sizes, although this has not been tested yet.

Iterative methods, while slower compared to their library counterparts are designed to consistently result in the required product. Each completed cycle results in a specific multiplication of the repeat number. The use of Golden Gate functionality multiplies repeats by four with 94% of colonies containing the correct product whereas the use of Biobrick functionality multiplies repeats by two with 100% of colonies containing the correct product.

Furthermore, the absence of polymerases in this workflow and the use of recombinase deficient DH5 α strains reliably produce products that can be validated for correct repeat fragment by

restriction digestion and agarose gel electrophoresis, as the suite of plasmids we made did not develop mutations.

CTRL also facilitates additional features such as the capacity for scarless assembly using Golden Gate methodology, easy screening using drop-out cassettes, and versatility of use with diverse mammalian applications by substituting backbones (**Figure 2.12**). We believe that the development of these features into CTRL will significantly empower the field of DNA repeat research.

Our journey through these methods underscores the complexity of repeat assembly and highlights the need for innovative solutions. We tested solutions based on workflows extensively used in engineering biology labs to ensure useability.

This work contributes to the broader scientific endeavour of understanding the effect of repetitive genetic sequences on common laboratory enzymes and workflows, with potential applications in various fields.

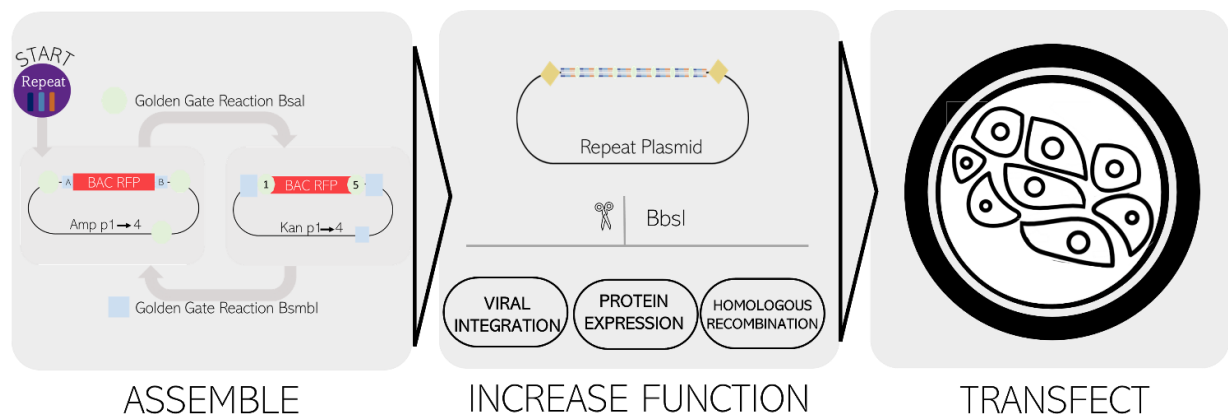


Figure 2.12: Typical implementation of CTRL workflow. Repeats are multiplied to the desired amount using Biobrick or Golden Gate as needed. To facilitate their use the backbones of resulting repeat plasmids can be changed using BbsI sites exposing sites compatible with MTK backbones. Lastly, the plasmids can be transfected into mammalian cells and observed for their responses.

3 TF DECOYS AS COMPETITIVE INHIBITORS OF GENE EXPRESSION

3.1 INTRODUCTION

Transcription Factors (TFs) play a pivotal role in managing many cellular functions. They guide the transcription machinery to specific locations within the vast DNA genome through their binding to Transcription Factor Binding Sites (TFBSs). In this capacity, TFs ensure efficient resource utilization and orchestrate cellular responses according to the plan set out in DNA.

In mammalian cells, these interactions are shaped by TF characteristics such as pleiotropy, cooperativity, and promiscuity. These TF characteristics lead to broad, tuneable, reversible regulation of genes and phenotype. These characteristics are what make TF engineering a strategy often used for applications in cell engineering, where precise intracellular TF control is required, enabling the creation of cells with desired characteristics (Naqvi et al., 2023).

Traditionally this is done by engineering protein abundance (Naqvi et al., 2023; Joung et al., 2023), post-translational modifications (Whitmarsh & Davis, 2000; Henley & Koehler, 2021; Lehal et al., 2020), and subcellular localisation (Whitmarsh & Davis, 2000). However, these techniques are often laborious. They also often rely on the expression of metabolically expensive proteins or require in-depth knowledge of the protein of interest to apply protein engineering techniques.

Alternatively, synthetic TFs such as dCas9, TALEs and ZFs, have been engineered to target GOIs (Waryah et al., 2018). These methods try to achieve specific activation or repression of a few GOIs where off-targets are discouraged, and broad regulation as observed by native TFs is a significant engineering challenge. Moreover, some synthetic TFs, especially those with bacterial origin, may be immunogenic in mammalian hosts further reducing use in applications such as in cell and gene therapies (Bhatt, García-Díaz & Foight, 2023).

Another approach to engineer gene regulation using TFs involves altering the number of TFBSs in cells. Zones of genomic repeats, termed tandem repeats often contain TFBSs for TF docking naturally modulating gene regulation in many organisms. Flooding cells with these TF decoys creates competition for TFs with active binding sites (**Figure 3.1**). This competition leads to a decrease in expression at these active sites. To create synthetic versions of these one requires information on TFBS sequence which is readily available in many databases such as HocoMocco (Vorontsov et al., 2024), JASPAR (Castro-Mondragon et al., 2022) or Factorbook (Pratt et al., 2022) and assembly methods capable of creating arrays of repeats.

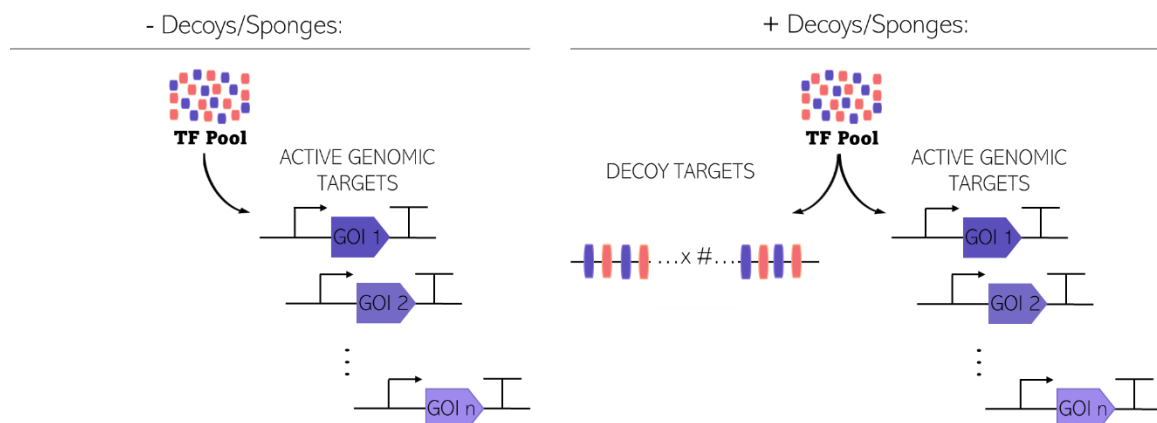


Figure 3.1: Principles behind functions of tandem repeats. In the absence of decoy TFBSs, the TF pool binds active genomic targets for GOI expression. In the presence of decoy targets the same TF pool will bind decoy and active targets reducing time on active sites and therefore the amount of expression from genes.

In many prokaryotes, an increase in genomic repeats is naturally used to alter gene expression by causing frameshifts to protein coding genes (Rando & Verstrepen, 2007). Synthetic TF decoy plasmids, often called DNA sponges, containing TFBS repeats have also been used in prokaryotic *E. coli* strains to engineer target genes and improve characteristics such as output, and dynamic range without using traditional control loops or causing additional burden on the population (Wan et al., 2020; Potvin-Trottier et al., 2016). Another study in bacteria investigated how gene expression timing can be encoded in a gene's regulatory DNA and found that the presence of decoy plasmids could significantly alter the timing response of gene expression (Ali & Brewster, 2022).

Repeats are also found in eukaryotic genomes dictating many important functions. For example, increased adherence of baker's yeast (*Saccharomyces cerevisiae*) is dictated by a gain of repeats in cell surface genes (Rando & Verstrepen, 2007). Engineering tracts of repeats in

yeast DNA is of high interest, with targeted insertion of genomic repeats being sought after for diverse projects. Conclusions as to the functions of genomic repeat positioning and length on higher-order cell gene regulation and genome attributes are often drawn as a result of the work of engineering repeats into genomes (Lee & Maheshri, 2012; Lancrey, Joubert & Boulé, 2018). Moreover, the addition of synthetic TF decoys in yeast was shown to favourably alter the behaviour of gene circuits in yeast, increasing the steepness of response (i.e. sensitivity), especially in conditions where synthetic repeats had a higher binding affinity for TF of interest (Lee & Maheshri, 2012).

Additional interesting applications occur when considering applying TF decoy strategies in multicellular organisms. TF binding, which is combinatorial, cooperative, and promiscuous in these organisms dictates expression patterns in cells influencing cell decisions and differentiation state (Joung et al., 2023; Lambert et al., 2018). Within genomes, the length and sequence of tandem repeats, often in non-coding regions, define cellular phenotype and disease states (Depienne & Mandel, 2021). Studies have also used TFBSs as methods of gene expression regulation. For example, increasing access to tandem repeats in genomes, using epigenetics, was used to decrease target gene expression (Janssen et al., 2000; Liu et al., 2007). Double-stranded TFBS oligonucleotides have also been used as a synthetic way to engineer free TF levels in cells but can be difficult to engineer as binding affinity is limited, especially since mammalian TFs bind cooperatively, and linear oligonucleotides are actively degraded intracellularly by exonucleases (Crinelli et al., 2004).

In the intricate landscape of gene expression regulation, engineering the level of free TF through TF decoys could provide a disruptive tool for changing expression profiles within mammalian cells simply through DNA addition. The proposed technology, which involves the transient exposure of TF decoy sites containing TFBS repeats, as plasmids in mammalian cells, could potentially offer a more effective approach to mammalian TF regulation. This is due to key design choices that account for TF binding characteristics in our TF decoy strategy.

Firstly, mammalian TF promiscuity and cooperativity, which makes it challenging to engineer linear oligonucleotides binding affinity, is addressed in TF decoy plasmids by the addition of sites in series to encourage cooperative binding. Also, TF decoys can be designed using a repeated sequence from a specific regulatory region of the DNA such that combinations of TFs that bind to a genomic target region upon a certain stimulus no longer bind as frequently. This accounts for the intricacies of TFs that form different complexes under conditions for various effects.

Secondly, TF decoys, being circular plasmids, are expected to be more resistant to degradation within cells compared to linear DNA like oligonucleotides which are prone to degradation by

exonucleases (Sasaki & Kinjo, 2010; Oliynyk & Church, 2022). This ensures the competing TFBSs remain in the nucleus long enough to exact observable effects.

Thirdly, as the TF decoy element of constructed plasmids does not require the expression of any components for effects, the plasmids should not add competition for cell's gene expression resources and work as soon as cells internalise these plasmids.

So, the use of TF decoys may provide reversible, tuneable, and resource-efficient tools for engineering gene regulation control in mammalian cells. To evaluate the effectiveness of these plasmids in practice, we apply this approach to various synthetic and native TFs. To display their capacity to compete for TF binding, we use previously published circuits with characterized output reporter expression profiles in the presence of TF. This allows us to compare conditions to determine the effects of TF decoys in cells. As the published circuits used a fluorescent reporter as an output for measurement of circuit response, we used flow cytometry for population level studies but also cell level resolution of effects on gene regulation. This additional resolution into cell characteristics allows the quantification of cellular responses through readily available flow compatible stains and antibodies. This makes it a valuable tool in our methodology, offering precise and detailed insights into the impact of our approach.

The validation of TF decoys in mammalian cells is carried out using well-established cell lines, HEK-293T and CHO-K1. These cells are transiently transfected with decoy plasmids of varying lengths and concentrations, and the effects on cell response are observed (**Figure 3.2**). The Chapter explores how TF decoy plasmids can be utilized to engineer or understand the responses and interactions of genetic circuits, induce change to cell fate, and observe natural TF crosstalk, highlighting a few of the wide array of applications this methodology could be applied to.

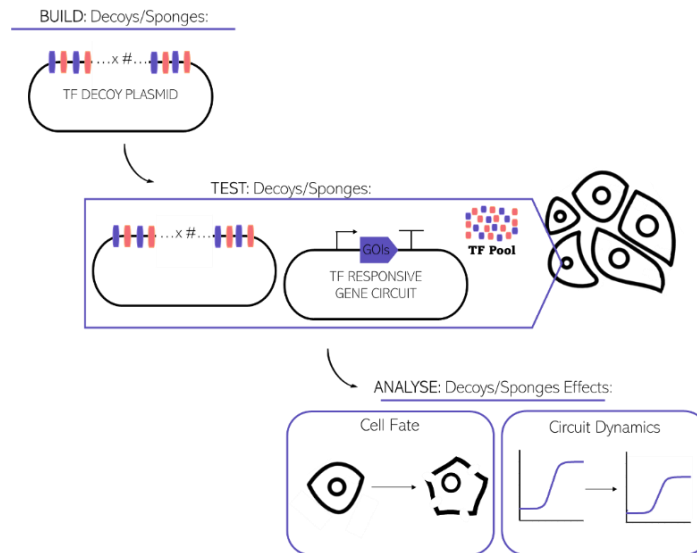


Figure 3.2: Using TF Decoys in mammalian cells to assess capacity to competitively inhibit gene expression from circuits. Firstly, decoys are built using CTRL described in Chapter 2. After validation, these plasmids are then tested for their ability to alter gene expression of endogenous or synthetic gene circuits with fluorescent outputs. The resulting cells are observed for differential characteristics by flow cytometry such as cell fate or altered circuit performance.

3.2 RESULTS AND DISCUSSION

3.2.1 DESIGN OF TF DECOYS FOR MAMMALIAN CELLS

TF decoys described throughout this chapter were designed to contain neutral spacer sequences between TFBSs. Neutral spacers are sequences that have been shown to not recombine in recombinase free bacterial strains (Wan et al., 2020) improving stability required to improve the DNA yield, required for mammalian transfections. All plasmid backbones were chosen with low copy number in DH5 α bacterial strains to further encourage plasmid stability and prevent plasmid loss during cell division (Watve, Dahanukar & Watve, 2010). High copy number plasmids, particularly those with repetitive sequences, are more susceptible to mutation and recombination. As such decoy plasmids in these backbones would be more unstable and potentially compromise the yield of plasmids required for experimental results. Therefore, by using low copy number decoy plasmids, we can reduce these risks. Furthermore, decoys were regularly verified for loss of repeats by restriction digest and retained their expected sizes. Another design feature added to decoy plasmids is neutral spacer sequences which ensure less recombination when propagating plasmids in bacteria but also added spaces between TFBS sites for reduced chances of steric hindrance between TFs. Considering that the spacer regions are at least 20 base pairs long (

Table 3.1), , and taking into account the α -helical structure of DNA, the subsequent

Spacer Identifier	Sequence (5' -3')	Size (bp)
NS 1	GCTCATACTTGCCCAGAGTCTTT	23
NS 2	TAGCAACCTTTGTCAAAAGTGATCGCT	27
NS 3	ACTAGCTTAAGTCAATCCAATGCGTTCAGGTG	32

transcription factor binding site (TFBS) is positioned at least two major grooves away from the initial one. This arrangement helps to minimize the impact of steric hindrance. All plasmids used in this study were assembled using CTRL, except TetOx320 (gift from Dr Wang Lab, University of Edinburgh).

Spacer Identifier	Sequence (5' -3')	Size (bp)
NS 1	GCTCATACTTGCCCAGAGTCTTT	23
NS 2	TAGCAACCTTTGTCAAAAGTGATCGCT	27

3.1: Neutral spacer DNA sequences utilised between TFBSs of TF decoy plasmid designs.

3.2.2 SEQUESTERING ORTHOGONAL TFS IN MAMMALIAN CELLS

3.2.2.1 Tet Transactivator Decoy Plasmids: Experimental Set-up

Decoy plasmids targeting Tet transcription factors (TFs) have been shown to enhance the performance of genetic circuits in bacteria and yeast (Wan et al., 2020; Lee & Maheshri, 2012). In light of these findings, we sought to investigate whether similar enhancements to mammalian rtTA-responsive circuits can be observed using Tet decoy plasmids, comparing the dynamics and trends observed.

To remain consistent, we opted to use rtTA (reverse tetracycline-controlled transactivator) as our primary TF target. Derived by engineering the Tet repressor protein native to bacteria, Tet transactivators including rtTA have many applications in mammalian systems and are a common feature in genetic circuits which require external induction or dose-dependent expression (Stanton et al., 2014). Extensive research has refined the system improving DNA sequence specificity and protein interactions (Zhou et al., 2006; Loew et al., 2010). The system is also versatile as promoters can be activated by using rtTA protein (TetON) or repressed using tTA (TetOff) TFs. Moreover, as this circuit is orthogonal to the native system, effects of crosstalk from interactions or interference with native pathways are assumed minimal (Stanton et al., 2014).

Therefore, we attempt to discover whether an orthogonal but relevant TF rtTA in mammalian engineering can be diverted from the expression of a transcriptional unit driven by TF binding at its sites upstream of the promoter using TF decoys. Observing any changes in response characteristics on the circuit or the system (**Figure 3.3**).

In the TetON system, used in **Figures 3, 4** and **5** of this chapter, constitutively expressed rtTA cannot bind its cognate promoter (pTet) which contains seven TetO sequences for rtTA binding upstream a minimal CMV promoter. In the presence of the inducer dox, the rtTA protein undergoes a conformational change and dimerises which allows its binding to TetO DNA

triggering the activation of the promoter and subsequent expression of gene or genes of interest. This is depicted in **Figure 3.3.a** as active sites under 'Gene of Interest'.

This construct was used to induce the expression of a GFP reporter and measure the capacity of co-transfected plasmids to compete for binding of the rtTA TF in the presence of dox (**Figure 3.3.a**). We then measured the output from the reporter 48 hours after co-transfection using flow cytometry. The singlet population was gated, and changes of expression were determined by changes in the geometric mean of the BL1 channel with appropriate emission and excitation for measurement of GFP.

To ensure observed effects are due to the competition between active and decoy TFBSs we observed effects on GFP when control systems not containing functional TFBSs for rtTA binding (pEmpty, TetO scr and pEF1α) are co-transfected (**Figure 3.3.b**).

Firstly, samples co-transfected with an empty plasmid (pEmpty), which only contains sequences for bacterial propagation (identical to TF decoy plasmids) and no TetO sites acted as the baseline fluorescence for co-transfections. Any reduction of fluorescence in this condition is not attributed to TF competition.

Secondly, a control plasmid containing 48 scrambled sites (Tetscr x48) with identical GC and AT nucleotide content to TetO but jumbled in sequence, so that TF should not recognize or bind showed similar average fluorescence to the empty condition, indicating neutral spacers and non-TetO sequences do not cause reduction of fluorescence at active sites including due to TF competition.

Thirdly, we compared the fluorescence output of conditions where the GOI was co-transfected with a plasmid containing the strong mammalian promoter (pEF1a), capable of binding transcriptional machinery but not rtTA TF, and this showed no decrease in reporter expression. This indicates that any decrease in reporter expression observed in pTet or TetOx7 conditions is likely due to the specific interaction of rtTA with its target TetO sequence, rather than a general effect of competition for transcriptional machinery such as RNA polymerase.

On the other hand, transfecting equimolar amounts of plasmids containing functional 7 TetO sites in the presence or absence of minimal promoter successfully reduced target expression by up to 38% at 0.01 ng/μL dox (**Figure 3.3.b**).

The collective evidence from these controls indicated that the rtTA system specifically regulates gene expression of the GOI. The observed decrease in fluorescence output in the presence of

pTet or TetOx7 decoy plasmids was therefore as a result of these elements competing for the rtTA transcription factor.

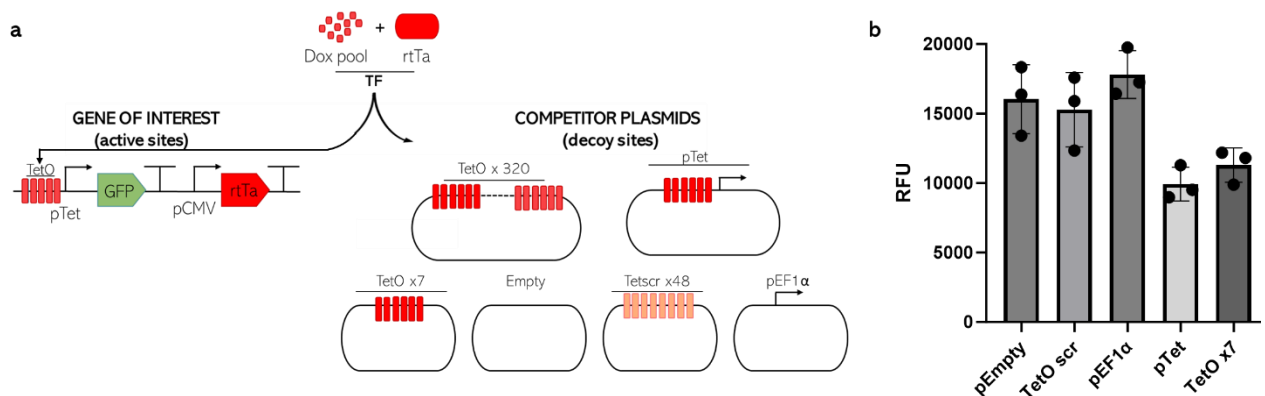


Figure 3.3: Testing principle of intracellular TF competition. **a.** Experimental set-up where various plasmids are co-transfected to test capacity to compete for binding of a finite pool of TF (0.01 ng/ μ L doxycycline, rtTA complex) with active sites driving the expression of a GFP reporter. The more successful the competitor the less fluorescence will be measured from the GFP transcriptional unit. **b.** Flow cytometry results show the effect of transfecting an empty plasmid, a plasmid with scrambled TetO sites or a strong mammalian promoter compared to plasmids with 7 TetO sites in the presence (pTet) or absence (TetOx7) of the minimal CMV promoter. Number of biological repeats for each sample and Two sided P-values are reported in 7.2 Statistical Test Summary P. 163.

3.2.2.2 Tet Transactivator Decoy Plasmids: Effects on Dose Response

To better understand the impact of TetO-containing decoy plasmids on the reduction of the expression of the fluorescent gene of interest, and to characterize the dose-response patterns of various levels of TF, we examined the influence of longer arrays of TFBSs in TF decoy plasmids. Our aim was to comprehend their effect across a spectrum of inducer concentrations in two distinct cell lines, namely HEK-293T and CHO-K1.

Both cell lines were co-transfected with a GOI construct and TF decoy plasmids, followed by fluorescence measurement 48 hours after dox induction (**Figure 3.4** and **Figure 3.5**). As dox is a critical component required for rtTA TF to be able to bind TetO DNA, its presence on the x-axis acts as a proxy for increasing TF activity.

In the absence of a TF decoy plasmid, the expression of eGFP increased with the concentration of doxycycline until it reached a plateau around 28,000 RFU in CHO-K1 and 17,000 RFU in HEK-293T the maximum capacity of this system. Beyond this point, further increases in dox do not lead to an increase in eGFP expression (black curves in **Figure 3.4**, **Figure 3.5**). In both

cell lines, the presence of intracellular TF decoy plasmid led to a reduction in GFP expression and a lower GFP expression at the plateau. This is such that the plateau decreased with an increasing number of decoy TetO sites. This observation suggests that the plateau is determined by the concentration of free rtTA/doxycycline complex, such that operators are less saturated in TF for less time in the presence of TF decoy. Interestingly, decoy did not reduce basal expression from the GOI in the absence of dox, suggesting basal expression may not be attributed to rtTA binding in the absence of the inducer. So, the more TFBSs added to TF decoy plasmids, the lower the dynamic range of the circuit becomes. These findings offer valuable insights into the operational mechanics of the circuit, which are governed by the binding characteristics of rtTA, and promoter capacity.

Moreover, the reduction in reporter output was such that TF decoy plasmids reduced expression similarly across all dox concentrations. This is indicated by the error bars in **Figure 3.4.b** and **Figure 3.5.b**, showing a consistent fold change between the presence and absence of the decoy at all dox concentrations. In these figures the mean fluorescence of all dox concentrations used was plotted for each concentration of decoy. For example, in CHO-K1 (**Figure 3.4**), equimolar amounts of TetOx7 to the GOI construct reduce expression by an average of $7.6\% \pm 1.3$ compared to no decoy conditions, across dox concentrations tested (**Figure 3.4**). This indicates a stable and uniform impact on gene expression at all concentrations of dox. This shows how TF decoys can be used to produce robust and predictable regulatory effects on GOIs.

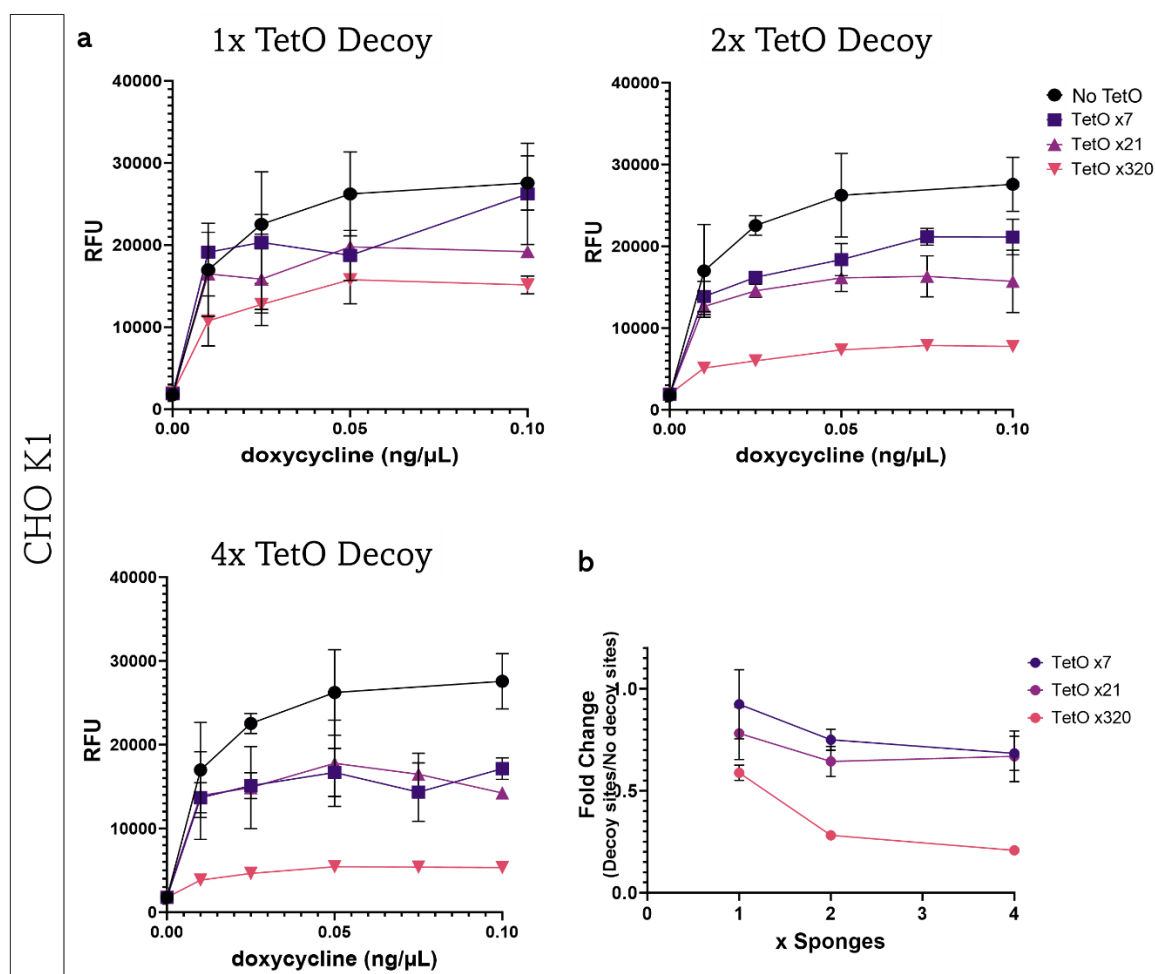


Figure 3.4: Compiled analysis of samples from at least three biological replicates acquired by flow cytometry of cells co-transfected with pTet-eGFP and TF decoy plasmids containing 0, 7, 21 or 320 TetO sites in CHO-K1. **a.** fluorescent output from pTet-eGFP with increasing active TF levels expressed as increasing amounts of doxycycline. 1xTet decoy displays results from transfecting equimolar amounts of pTet-eGFP to decoy plasmids. 2xTet decoy displays results from doubling decoy plasmids in transfection mixture and 4x Tet decoy quadruples decoy plasmids in transfection. The trend shows increasing the quantity of TFBSs (via the addition of more plasmid or more sites) corresponds to a reduction in fluorescence (RFU) from pTet-eGFP. Points represent the geometric mean while error bars represent the standard deviation of at least three biological replicates. **b.** Average fold change between samples in the presence vs absence of decoy plasmids (Calculated as: RFU Presence of TetO decoy divided by RFU absence of TetO decoy but a plasmid of equal size. Each dot corresponds to the average fold change for samples of all doxycycline concentrations with the particular decoy and concentration).

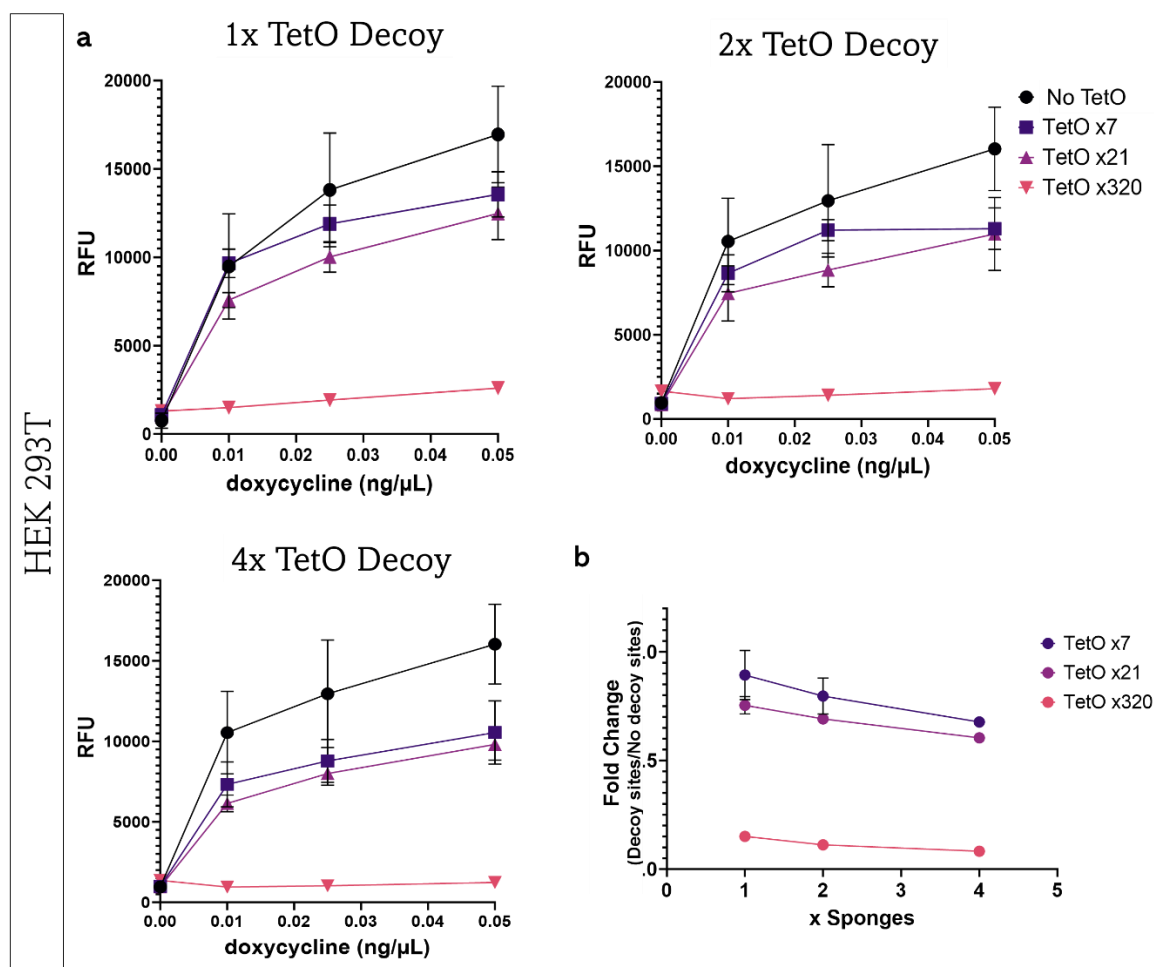


Figure 3.5: Compiled analysis of samples from at least three biological replicates acquired by flow cytometry of cells co-transfected with pTet-eGFP and TF decoy plasmids containing 0, 7, 21 or 320 TetO sites in HEK-293T. **a.** fluorescent output from pTet-eGFP with increasing active TF levels expressed as increasing amounts of doxycycline. 1xTet decoy displays results from transfecting equimolar amounts of pTet-eGFP to decoy plasmids. 2xTet decoy displays results from doubling decoy plasmids in transfection mixture and 4x Tet decoy quadruples decoy plasmids in transfection. The trend shows increasing the quantity of TFBSs (via the addition of more plasmid or more sites) corresponds to a reduction in fluorescence (RFU) from pTet-eGFP. Points represent the geometric mean while error bars represent the standard deviation of at least three biological replicates. **b.** Average fold change between samples in the presence vs absence of decoy plasmids (Calculated as: RFU Presence of TetO decoy divided by RFU absence of TetO decoy but a plasmid of equal size. Each dot corresponds to the average fold change for samples of all doxycycline concentrations with the particular decoy and concentration).

3.2.2.3 *Tet Transactivator Decoy Plasmids: Comparison Across Species*

Tet TFs, which rtTA belongs to, have previously been sponged in bacteria (Wan et al., 2020) and yeast (Lee & Maheshri, 2012). In bacteria, Tet decoy plasmids were used to sequester the Tet repressor (TetR). The study developed synthetic TF decoys for 3 different bacterial TFs including those for sequestration of TetR which ranged from 1 TetO site to 320 TetO sites. TF decoy plasmids effectively modified gene circuit responses in a straightforward and adaptable method enhancing the performance of synthetic gene circuits by significantly decreasing basal leakage, enhancing system output amplitude, tuning dynamic range, and alleviating host growth inhibition. Unlike in the study in bacteria, our data set did not show basal expression reduction, therefore a marked increase in dynamic range of our system was not observed. This may be due to the different TFs used in these two studies, rtTA used in our study was engineered to bind one of seven operators in mammalian pTet in the presence of dox (see **Figure 1.5.b**) whereas TetR binding of inducer facilitates its dissociation from operator sites in the bacterial promoter (more similar in function to the system depicted in **Figure 1.5.c**). Moreover, specific engineering efforts have further improved features for mammalian applications by minimizing background expression (Loew et al., 2010) cutting rtTA binding to TetO in the absence of dox.

In the published yeast study, tTA TF was used with a promoter containing 7 TetO sites (similar to **Figure 1.5.a**). The group assembled up to 240 TetO sites with no spacer sequences. These decoy sites competitively bind to TFs, influencing target gene expression by inhibiting gene expression from the tTA activator in the absence of dox. This sequestration of tTA on decoy sites transformed the graded dose—response of active promoters into a sharper, sigmoidal-like response. The altered response was attributed to tighter TF binding to decoy sites compared to promoter binding sites, through modelling and chromatin immunoprecipitation measurements. The increase in the Hill slope in response to the dose of dox (i.e. for sharper ON/OFF dynamics) that was reported in yeast was not observed in our tests in mammalian cell lines where slopes from TF decoys with longer arrays do not consistently increase (curves were fitted using GraphPad Prism with non-linear regression [Agonist] vs response). Further understanding the affinity of the TFs in mammalian cells and comparing decoy plasmids with and without spacer regions could help understand the mechanism behind these differences, which could simply be a feature of working in different organisms. The results shown here demonstrate that TetR TF sequestration leads to some different responses to circuit dynamics in the different circuits but none-the-less provided a method of regulating gene expression in a tuneable way by changing the number of TetO sites per plasmid or the amount of plasmid transfected into the HEK-293T or CHO-K1 cells.

3.2.2.4 Tet Transactivator Decoy Plasmids: Repeats in Series or Parallel

I next sought to use the data to investigate the effects of varying plasmids with a different number of repeats while varying the amount of these plasmids transfected into cell lines. We set out to use this dataset to examine the optimal configuration to obtain the desired impacts of decoy sites. Specifically, whether the total number of repeats dictates response, or if the arrangement of repeats in series (increasing sites per plasmid) or increasing total repeats in parallel (by increasing intracellular copy number) provide different characteristics to the competitive inhibition. Gaining these insights can guide future design decisions. For example, whether there is a need to generate larger arrays of repeats, which are harder to assemble and transfect, or multiple copies of smaller plasmids are sufficient to obtain desired circuit characteristics.

To explore the relationship between the number of binding sites in series and the resulting reduction in gene expression, we focused on the effects of TetOx21, with three times the number of sites compared to TetOx7. Assuming a linear relationship between the number of binding sites in series and the reduction in gene expression, a 3-fold reduction in expression is expected with TetOx21 compared to TetOx7. This relationship held true for equimolar transfections of TetOx7 and TetOx21 in HEK-293T and CHO-K1 as equimolar transfection of TetOx7 and TetOx21 in CHO-K1 resulted in 7.6 % and 21.9% reduction compared to no TF decoy conditions, respectively (

Table 3.2). 21.9% is very close to the predicted 22.8% obtained by multiplying 7.6% by 3.

Table 3.2: Effect of TFBS added in series. Example results of TF sequestration in the presence of 1x, 2x or 4x TF decoy. Percent reduction in reporter fluorescence compared to plasmid not containing TetO sites (pEmpty) conditions in CHO-K1. The table shows the experimental values of percent fluorescence reduction observed in the presence of TetOx7 or TetOx21 compared to conditions in the absence of a decoy. This can be compared to values that assume a linear

	TetOx7 (Experimental) TetOx7 (Experimental)	TetOx21 (Calc: TetOx7 x3) TetOx21 (Calc: TetOx7 x3)	TetOx21 (Experimental) TetOx21 (Experimental)
1x TF Decoy	7.6	22.8	21.9
2x TF Decoy	25.0	75.1	35.7
4x TF Decoy	31.7	95.1	33.1

reduction of fluorescence as the number of sites increases per plasmid to calculate the expected reduction in fluorescence.

However, this trend does not hold true when decoy plasmids were doubled or quadrupled in transfections

Table 3.2. Where doubling the plasmid should lead to a 75% reduction in fluorescence instead of the 35.7% reduction observed experimentally.

The observed deviation from the anticipated trend in equimolar transfections could be partially attributed to the transfection conditions, where increasing the concentration of DNA in a transfection mix does not necessarily result in a proportional increase in the number of

	TetOx7 (Experimental)	TetOx21 (Calc: TetOx7 x3)	TetOx21 (Experimental)
1x TF Decoy	7.6	22.8	21.9
2x TF Decoy	25.0	75	35.7
4x TF Decoy	31.7	95.1	33.1

internalized plasmids.

We similarly sought to explore the relationship between the number of binding sites in parallel on the reduction in gene expression. In this direction, we increased the amount of plasmid used in transfection mixes adding equimolar, double, or quadruple the amount of TF decoy. For HEK-293T (**Table 3.3**) transfected with 1x or 2x TetOx7 TF decoy, the observed percent reductions are close to the expected ones. This was not the case for the transfection of 4x TetOx7 decoy or TetOx21 in terms of the link between the expected and observed values.

Table 3.3: Testing cumulative effect of using more plasmid in transfections. Example results of TF sequestration in the presence of TetOx7 as compared to TetOx21. Percent reduction in reporter fluorescence compared to plasmid not containing TetO sites (pEmpty) conditions in HEK-293T. The table shows experimental and expected/calculated values of percent fluorescence reduction observed in equimolar, double or quadruple concentrations of TF decoy with 7 or 21 sites in transfection mix compared to pTet-eGFP plasmid.

	TetOx7 (Exp)	TetOx7 (Calc)	TetOx21 (Exp)	TetOx21 (Calc)
1x TF Decoy	10.7		24.6	
2x TF Decoy	20.3	21.4	30.9	49.2
4x TF Decoy	32.3	42.8	39.6	98.4

Engineering decoy plasmids to contain a reporter of transfection could improve the resolution of the effects of increasing the copy number of decoy plasmids. While this approach would provide a method to visualise and normalise for variation in the efficiency of transfections, other possible

confounding effects may complicate analysis. For example, enhancers which increase the frequency transcription of a GOI are often made up of TFBSs (Kang et al., 2019). Therefore, the TFBS array within TF decoys may act like an enhancer (Lim et al., 2024) in their ability to attract transcriptional machinery to the decoy plasmid. The inclusion of a constitutively expressed reporter on TF decoy plasmids could therefore lead to the reporter's preferential expression. This can cause a strain on limited cellular resources leading to less expression from the eGFP construct. Therefore, results obtained in the absence of transfection marker include less interaction with cellular components which may differ from effects observed in the presence of transfection marker, requiring careful consideration in implementations.

Finally, some differences occurred when transfecting the same systems into HEK-293T and CHO-K1. HEK-293T appeared more sensitive to TetOx320, exhibiting a larger change in expression compared to conditions with no decoy (**Figure 3.5.b**). Therefore, the answer to whether TFBS achieve better repression when placed in series or parallel may be cell line dependent as well as TF dependent.

In this section, we demonstrated how TF decoys can be used to understand the functionality of circuits containing the rtTA TF of the TetON system in mammalian cells. This showed for the first time in mammalian cells (to our knowledge) the effectiveness of altering TFBS concentration using TF decoy plasmids in achieving repression of target GOIs.

3.2.2.5 Tet Transactivator Decoy Plasmids: Modulating & Investigating Genetic Circuits

The versatile TetR-based TFs have become indispensable elements in the realm of genetic circuits. Their strategic use as part of various circuits, often incorporating feedback, has produced circuits with diverse characteristics, including sensitivity, detection range, and response time (**Figure 3.6.a**) (Afroz & Beisel, 2013). In this section, we delve into the modulation of Tet TF containing genetic circuits through the utilization of TetO decoy plasmids, to understand the components that underlie the functionality of circuits depicted in **Figure 3.6.b,d**. This investigation outlines how engineers can implement the use of TF decoy plasmids to fine-tune the outputs of various genetic circuits or study circuit mechanisms.

To advance in this direction, we chose two previously published mammalian circuits that employ Tet TFs to achieve distinct dynamics and properties. We chose a previously developed positive feedback loop (Di Blasi et al., 2023) (**Figure 3.6.b**) where the Tet protein rtTA and a reporter

are both expressed downstream of the Tet responsive promoter (pTet), forming a multicistronic transcriptional unit (We used 2A, where DNA with instructions for a single mRNA molecule but translates into more than one protein by utilizing a mechanism known as ‘ribosome skipping’). This is commonly referred to as an auto-activation circuit or positive feedback loop (PFL), where the output of a transcriptional unit triggers its own expression (**Figure 3.6.b**).

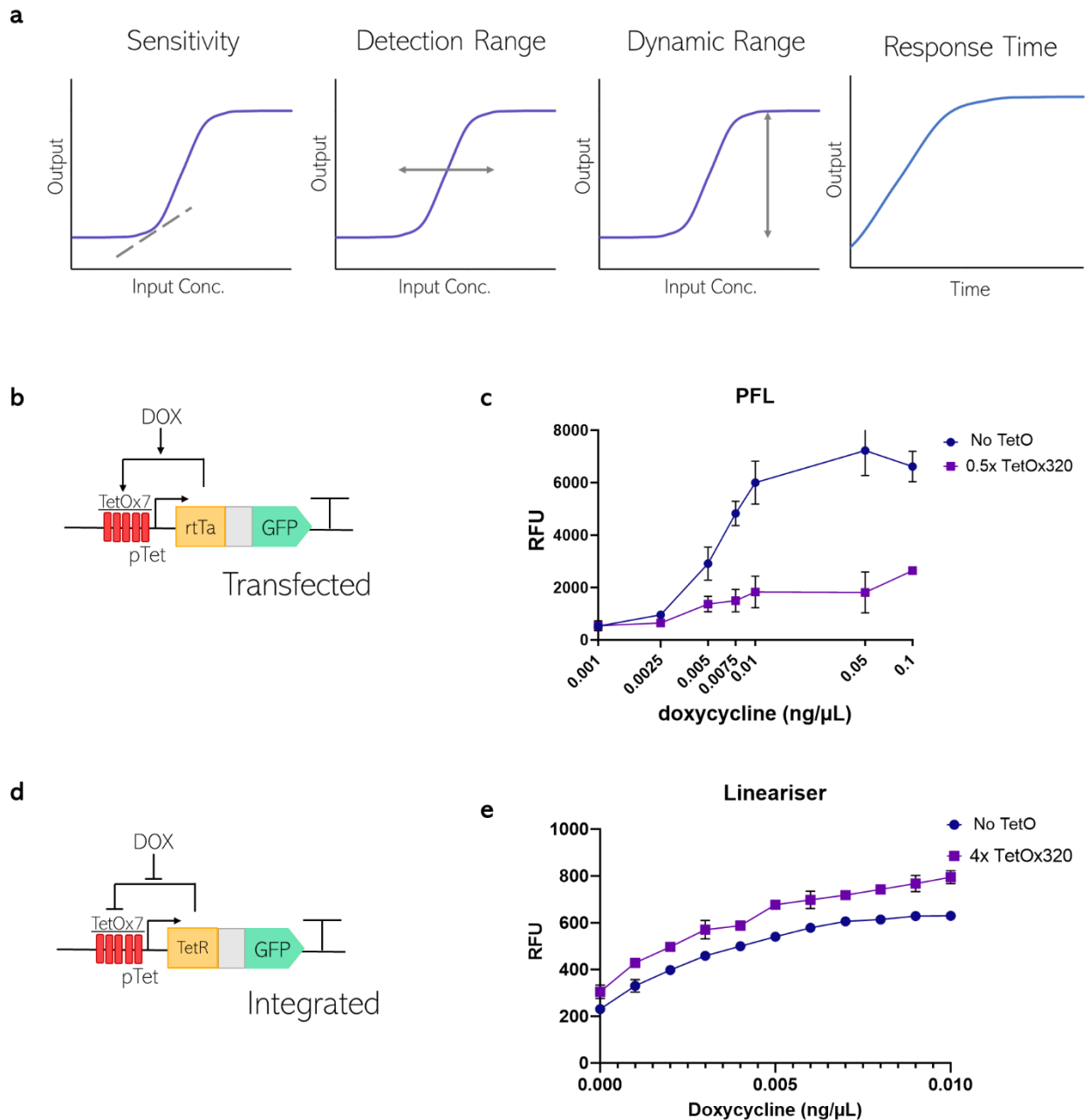


Figure 3.6: Free TF effect on autoregulatory circuits **a.** Features of genetic circuit components come together to achieve circuit dynamics which include sensitivity, detection range, dynamic range, and response time. Effects of TF decoy plasmids on genetic circuits may change these. **b.** Previously reported design of a PFL where induction by doxycycline induces expression of rtTA and GFP reporter from the same mRNA transcript with a 2A between. **c.** Effect of half the amount of TetOx320 decoy compared to PFL in a transfection. Using a TetOx320 with this

system suppressed expression from the loop circuit and reduced the sensitivity and range of the switch, key features of PFLs. **d.** Previously reported design of an NFL integrated into HEK-293T cells where induction by dox reduces repression inducing expression of TetR repressor and a reporter with a more linear sensitivity giving rise to the name of this circuit “lineariser”. **e.** Effects of 4x TetOx320 decoy on an integrated NFL (lineariser). Decoys did not drastically perturb this circuit.

Here the addition of dox binds to basal rtTA and induces the expression of more rtTA and a GFP reporter (TetON configuration) (**Figure 3.6.a**). The resulting response curve showed a steep increase in GFP expression, upon the addition of dox, governed by rtTA concentration until the TFBSs are saturated (**Figure 3.6.c** blue line). This is often referred to as ultrasensitivity. While small increases in dox lead to sharp responses, the response often takes longer to occur than the two transcriptional unit system shown in **Figure 3.3.a**. These features were consistent with characteristics of autoregulatory PFLs (Afroz & Beisel, 2013). The addition of decoy TFBSs should create competition for TF binding, reducing the expression of rtTA and GFP. Indeed, the resulting curve plateaus at around 2,000 RFU whereas in the absence of TF decoys, it was much higher, reaching 6,000 RFU. Effectively very little TF decoy plasmid, compared to doses used in other experiments, was required to reduce expression in this circuit by 75% and subsequently the dynamic range of this circuit was altered.

The second feedback loop circuit we tested with added TF decoy, is a genome-integrated negative feedback loop (NFL) built and provided to us by Prof. Balázsi’s group (Nevozhay, Zal & Balázsi, 2013; Szenk, Yim & Balázsi, 2020). This is an auto-repression circuit that expresses TetR and reporter from a multicistronic unit. This is designed such that the TetR TF binds TetO in the absence of dox and inhibits its expression. Upon doxycycline addition, which acts to derepress the system, TetR undergoes a conformational change that prevents TetR/promoter coupling allowing expression of more Tet repressor and reporter (**Figure 3.6.d**). As a result, circuit response to dox was less steep, had a larger detection range and cellular distributions were narrower (**Figure 3.6.e** blue line) (Afroz & Beisel, 2013). When decoys were added, even at the highest doses (4x TetOx320) which achieved up to 90% repression of fluorescence in **Figure 3.5**, did not result in as much a change in output fluorescence, although a small significant increase in reporter expression is detected. This trend is consistent with the author’s two-gene mammalian ‘lineariser’ design (which can be likened to a decoy with equimolar operator sites), where results were used to prove detection range could be applied to more than one gene (Nevozhay, Zal & Balázsi, 2013). In the team’s work, the second gene expression unit which is separated from the first with 2A, appears to have slightly higher expression than the other.

Utilizing a TF decoy module, for its ability to sequester TFs, proved helpful in elucidating the underlying mechanism of this circuit function. The observed linearised response is governed by control of TetR expression within the circuit such that in the presence of decoys, de represses repressor at the promoter, causing the expression of more repressor, preventing abrupt surges in expression from the unit while subtly elevating the plateau of expression. This mechanism appears to exert a modulating influence on the response, persisting even in conditions with the highest amount of TetOx320 decoy plasmid (4xTetOx320).

A more detailed examination of the lineariser circuit displayed in (**Figure 3.6.d**) reveals the influence of this circuit on cellular behaviour (**Figure 3.7**). Each histogram in **Figure 3.7.a** illustrates the cell count distribution of circuit output GFP for a certain concentration of dox, such that histograms in the absence of dox are at the bottom and max dox (0.01 ng/ μ L) is at the top of the panel. The histograms on the right represent cells in the presence of the decoy, while those on the left indicate cells without the TetOx320 decoy. The histograms show a similar distribution across the GFP axis, indicating there is no obvious change to GFP expression, but a decrease in the number of events when the decoy is present, especially at higher dox concentrations. This is further illustrated in **Figure 3.7.b**, where the number of singlet events is plotted for dox concentrations, both in the presence and absence of the decoy. It is noteworthy that in contrast to the presence of TF decoy the event count remained relatively stable when a plasmid not containing TetO sites was transfected. However, fewer events can be detected in the presence of decoy plasmids, leading to an 80% reduction in cell count in samples where the decoy was transfected. So, de-repression of the NFL promoter by decoy plasmids was cytotoxic.

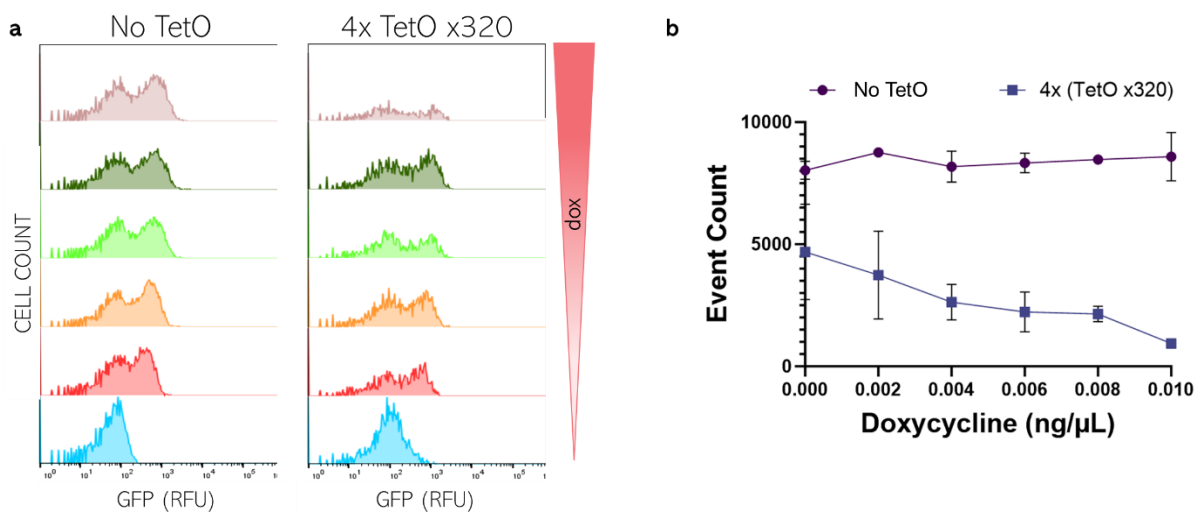


Figure 3.7: Effect of TF decoy plasmids and NFL (Lineariser) co-transfection on cells. **a.** Histograms of cell counts vs GFP spread in the presence (right) or the absence (left) of TetOx320 decoy for a range of dox concentrations. This displays a reduced

*number of events in the presence of TF decoy plasmid indicated by increased noise in peaks and lower in scale as and more of the dox inducer. **b.** Change in event count, indicating cell death, with increasing dox concentrations in the presence or absence of TF decoy plasmid.*

These observations aligned with findings from other studies on auto-repression circuits (Chaturvedi et al., 2022). In these circuits, relieving repression on the promoter within an auto-repression circuit can trigger a toxic overexpression of proteins. This suggests that auto-repression negative feedback loops respond to drastic changes in free TF levels, such as the addition of many decoy TFBSs, by overexpressing the repressor in a harmful attempt to regulate the response.

In summary, the incorporation of decoy plasmids, in conjunction with positive and negative feedback loop circuits, highlighted distinctive features of these circuits. This exploration offered insights into the roles of free TF concentration on positive and negative feedback loop circuits, displaying how the reduction in free TF due to decoy presence destroys the ultrasensitivity of PFLs and increases output fluorescence in NFLs although not in as drastic a way as NFL works to dampen effect by expressing repressor. The addition of a TF decoy module to PFLs could act as a method of modulating or tuning the dynamic range in these circuits. While the NFL results show it is possible to sequester a repressor (TetR), further testing of decoys with a two-transcriptional unit configuration (similar to **Figure 3.3.a**) may be used to alleviate repression and activate output from transcriptional units.

Future work with this experimental set-up could observe whether decoys and modulation of TF off-rates offer methods that can be used to achieve tuneable levels of noise in gene transcriptional responses, similar to the reported computational work by Ali & Brewster, 2022. This would provide a means to fine-tune and control the stochasticity in gene regulatory networks.

3.2.3 SEQUESTERING NATIVE TFS IN MAMMALIAN CELLS

Mammalian TFs play a crucial role in the regulation of gene expression and are essential for various biological processes, including cell growth, differentiation, and response to environmental stimuli. Unlike in cases of TFs with prokaryotic origins such as TetR and rTA, eukaryotic TFBSs tend to bind cooperatively and are less sequence-specific (Zhao, 2023). This means that cooperative TF binding exacts large regulatory impacts from low affinity TFBSs. With

the goal of achieving more intricate signal processing required for cell decisions engineered tools have tried to recapitulate cooperative features into synthetic systems (Bashor et al., 2019; MacCarthy et al., 2024).

In natural cell systems, TF activity on mammalian genomes is regulated through repeat sequences in non-coding regions of the genome, sometimes spanning hundreds of base pairs, containing TFBS repeats that regulate free intranuclear TF levels and thereby tune gene expression. A high percentage of these repetitive genomic regions known as STRs contain lower affinity repeats (compared to active sites) but still maintain higher binding to preferential TFs compared to random sequences (Horton et al., 2023). Recent work has also shown the intricate relationship between the affinity of TFs, controlling TF stoichiometry and differentiation state, making avenues to engineer synthetic STRs for cell engineering interesting (Nair et al., 2023).

Inspired by STRs in genomes, we sought to create TF decoy plasmids capable of sequestering specific native TFs or TF complexes to engineer responses in cells. This method is tuneable by controlling repeat length or copy number as seen in decoy plasmids tested for synthetic TFs. We therefore next investigate the use of TF decoy plasmids for the strategic sequestration of native TFs in mammalian cells, guided by our understanding of TFBSs and STRs. This work could open up many applications within cellular engineering, offering a powerful and tuneable approach to modulating gene expression and cellular responses.

3.2.3.1 Mammalian TFs Can Be Sequestered

As a primary TF target for sequestration, we chose NRF2, a basic leucine zipper TF implicated in cellular response to oxidative stress. Aberrant expression of NRF2 leads to several diseases in many tissues (Cho, 2013; Kuosmanen et al., 2016). Under normal conditions, NRF2 is bound to inhibitory complexes that reduce the movement of the TF to the nucleus and actively degrade NRF2 in the cytoplasm (Ganner et al., 2020). When cells are exposed to oxidative stress, the inhibitory complex is modified such that it can no longer bind NRF2 allowing NRF2 to freely accumulate in the nucleus and target genomic regions containing antioxidant response elements (ARE) in regulatory elements of genes (**Figure 3.8.a**).

In the decoy designs we assembled, three unique neutral spacers (**Table 3.1**) interspace four TFBSs in the synthesized DNA used for domestication into CTRL. Once introduced into CTRL backbones, with each cycle the neutral spacers also multiply in number. These spacers are on

average 27 bp long and function to reduce recombination when propagating plasmids in bacterial strains and also facilitate optimal protein binding by spacing out TFBSs. Neutral spacers were also checked not to contain homology to NRF2 binding sites and were assumed to be neutral by transfecting other neutral spacer containing plasmids with no impacts on the cells (**Figure 3.3.b**, Control graphs **Figure 3.8c and d**).

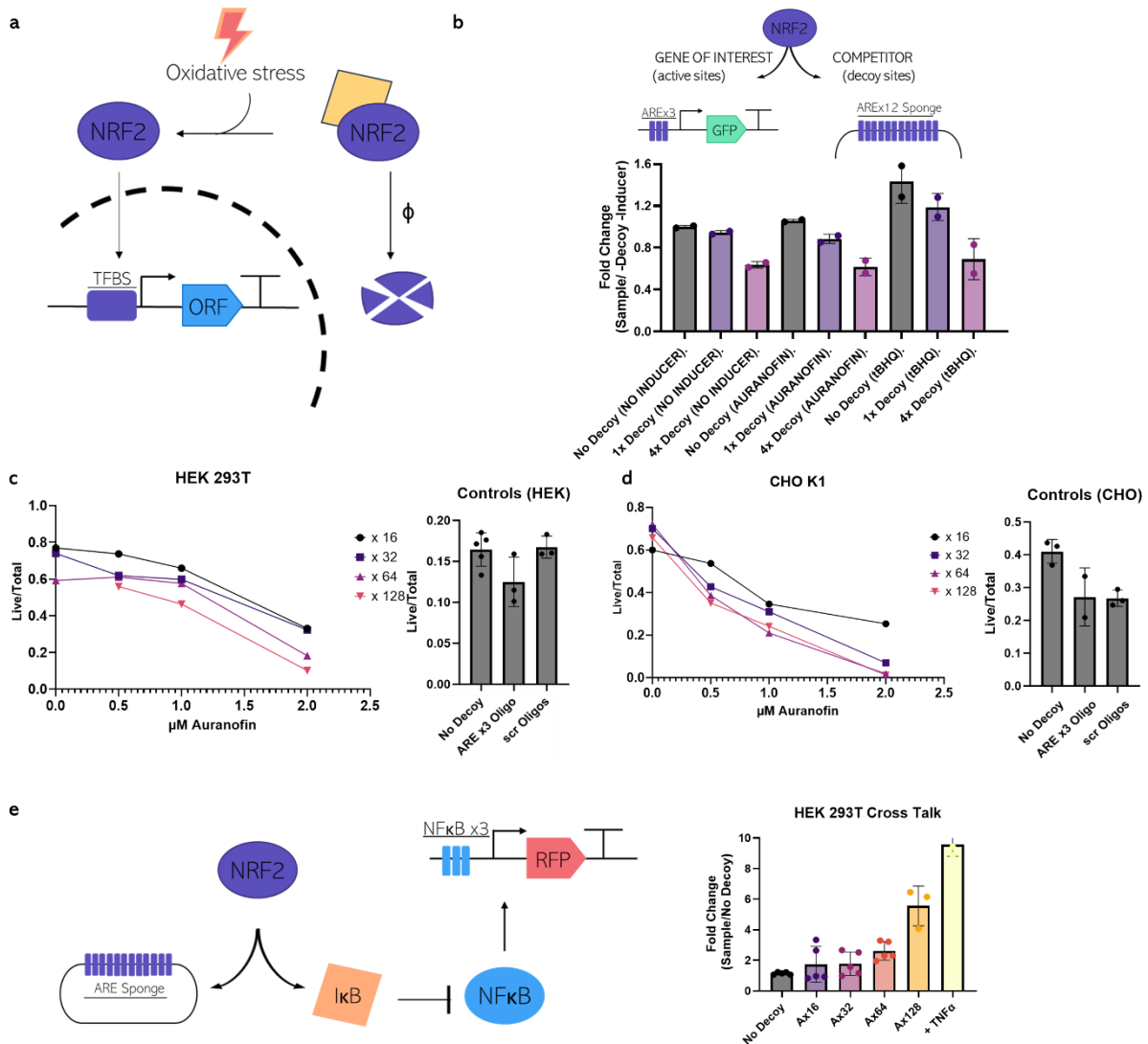


Figure 3.8: Regulating native gene regulation of NRF2 using decoy plasmids. **a.** NRF2 is naturally actively degraded in the cytoplasm and cannot normally enter the nucleus. Upon cell exposure to oxidative stress, in our studies simulated by exposing cells to auranofin or tBHQ, NRF2 is released and is then able to travel to the nucleus where it regulates gene expression by binding ARE in the genome. **b.** Competition for TFBSs between active sites, the expression of GFP from an NRF2 responsive promoter containing 3 ARE sites or decoy containing 12 ARE sites. Plotted fold changes are sample values divided by average value of condition where no inducer and no decoy are present. In the absence of inducers of oxidative stress, higher doses of decoy plasmids reduce the expression of GFP. In conditions where stress is induced with

Auranofin or tBHQ, the basal expression from ARE promoter increases in the absence of TF decoy, while the increase of TF decoy copy number reduces GFP expression per cell. **c.d.** Live/Dead staining (Invitrogen), shows increasing the number of decoy sites per plasmid reduces the viability of cells and therefore capacity of cells to respond to oxidative stress by binding native ARE sites in the genome of HEK-293T cells (c) or CHO-K1 cells (d). Control graphs are of proportion of live events of samples with 0.5 μ M of Auranofin. **e.** NRF2 and NF κ B TF crosstalk such that NRF2 increases expression of I κ B the inhibitor of NF κ B. The sequestration of NRF2 should reduce the expression of I κ B therefore freeing a larger portion of NF κ B to enter the nucleus and alter gene expression of RFP under the control of a promoter containing 3 sites for NF κ B binding. Increasing the number of ARE sites on decoy plasmids increases RFP fluorescence per cell. TNF α is a known inducer of NF κ B, confirming pNF κ B-RFP functionality. Fold change of detected RFP from samples divided by condition with no competitive TFBS.

Decoy plasmids containing the consensus binding sequence of ARE for NRF2 targeting (5' GTGACTCAGCA 3') with 12, 16, 32, 64, 128 and 256 sites were assembled. The smallest TF decoy plasmid containing 12 sites was transfected into HEK-293T cells in conjunction with a promoter containing 3 ARE sites upstream a minimal promoter for NRF2 driven GFP expression as a reporter of ARE-driven gene expression (**Figure 3.8.b**). Endogenous levels of NRF2 in HEK293T are reported (Johansson et al., 2017) as lower in the absence of oxidative stress and are displayed in the first column of the dataset shown in **Figure 3.8.b**.

This plasmid was used as a condition with a lower number of binding sites reducing cellular interference such as cytotoxic effects which may impact fluorescence readings. This is because high levels of cytotoxicity can lead to autofluorescence due to changes of cellular composition and structure and not due to changes in expression from the GOI (Kozlova et al., 2020). To further account for any changes in acquired events that may affect the analysis of results, we report the results as average fluorescence per cell (Mean Fluorescent Intensity, MFI) to account for any change in cell count in samples.

The decoy plasmid containing 12 sites was applied at equimolar amounts along with the pARE-eGFP cassette or quadruple the amount of ARE decoy to pARE-eGFP. In the absence of an oxidative inducer (No Inducer, **Figure 3.8.b**) some eGFP expression is detected likely from available free NRF2 in the nucleus, required for cell maintenance (He, Ru & Wen, 2020). The presence of our TF decoy plasmid reduced the capacity for GFP expression in cells as compared to the transfection of a plasmid without ARE sites but of identical size and backbone. In the presence of the specific NRF2 inducers, Auranofin and tetrabutylhydroquinone (tBHQ) which have been shown to increase oxidative stress and lead to cytotoxicity of cells (**Figure 3.8.c**), mean basal expression of GFP per cell increases, indicating an increase in active NRF2

in the nucleus while the presence of TF decoy reduces output from pARE promoter. The presence of 4x TF decoy consistently reduces cellular GFP expression on average by 40% in all inducer conditions, indicating a consistent effect in our cells.

These data show that the decoy plasmids with native TFBS can be used to reduce expression from pARE-eGFP. This is likely due to the sequestration of the NRF2 TF, which has been proven to activate similar synthetic constructs with ARE-responsive promoters in the presence of oxidative inducers (Johansson et al., 2017).

3.2.3.2 Native TF Decoy Plasmids Redirect Cell Fate

Activation of loci containing AREs ensures cells respond to oxidative stress in the environment. The knock-down of NRF2 will therefore reduce the cell's capacity to respond leading to cell death. Therefore, transfecting ARE decoy plasmids with increasing array length of ARE repeats into HEK-293T and CHO-K1 (**Figure 3.8 c or d** respectively) should show reduced viability in these cells, especially under conditions of oxidative stress.

We observe cytotoxicity, measured as the proportion of live cells in LIVE/DEAD stained samples, as more Auranofin is added to cells in the presence of TF decoy containing 16, 32, 64 or 128 sites such that increasing number of repeats led to reduced number of live cells in samples. This cytotoxicity can likely be attributed to cells being unable to respond to oxidative stress in the absence of NRF2.

Previous articles have shown the use of double-stranded oligonucleotides containing TFBSs can be used as decoys for TF binding (Crinelli et al., 2004). We tried to replicate this by annealing oligos together containing the ARE consensus sequence to compare the repression effects of plasmids vs linear oligos. This showed that the use of double-stranded oligos with the ARE consensus site (in this case, 400ng of oligonucleotide, the equivalent of 200x the amount of decoy plasmids used) does not significantly reduce live cell count compared to samples without TF decoy, or samples transfected with scrambled oligos in HEK-293T **Figure 3.8.b**. In CHO-K1 cells, ARE duplex oligonucleotides do not give any difference in behaviour compared to using a scrambled ARE sequence **Figure 3.8.c**. Whereas in both cell lines NRF2 decoys were able to deplete cell count.

This observation may be due to different binding characteristics between NRF2 TF that we were studying and NFκB TF used in these past oligonucleotide-based studies. This may hint to the TF decoy having more consistent and reliable results across different TFs as it accounts for the

cooperative binding of mammalian TFs which we achieve by adding many sites in series. Double-stranded oligos, on the other hand, are susceptible to fast degradation by exonucleases and rely only on the affinity of TF with sites (Crinelli et al., 2004). Assembling NFκB decoy plasmids with the sequence described in this paper would provide more conclusive evidence as to whether this is the case.

Therefore, our findings demonstrate that the cell's ability to respond to oxidative stress can be altered, in this experiment leading to cytotoxicity. TF decoy plasmids could potentially overcome the challenges associated with the weak affinity of some mammalian TFs by assembling repeat arrays which support cooperativity. This approach could lead to a more reliable regulation of gene expression and guide cells towards specific states other than cytotoxicity. We propose a workflow similar to the one described here, where synthetic transcriptional units are constructed for competition assays. These units must have promoters regulated by the TF, leading to either the expression or repression of the reporter. By observing the graded effect of different TF decoy plasmids on reporter expression, we can gain valuable insights into the dynamics of gene regulation and trigger cellular responses.

3.2.3.3 Native TF Decoy Plasmids Assess TF Crosstalk

We further endeavoured to see if we could use the built TF decoy plasmids to observe native relationships between TFs in cells using decoys. NRF2 regulates the activity of NFκB by inducing the expression of its inhibitor IκB. Therefore, the sequestration of NRF2 from the IκB promoter is expected to reduce expression of IκB and increase regulation by NFκB (Gao et al., 2022). To test this, we transfected cells with a promoter containing 3 sites for NFκB binding controlling expression from RFP (Johansson et al., 2017) and various sizes of ARE decoy plasmids (**Figure 3.8.e**). We observed as we increased the number of decoy sites in the plasmids, that we increased mean cell fluorescence from pNFκB-RFP. This was likely due to a reduction in IκB expression and an increase in NFκB expression. We use TNFα, a known inducer of NFκB, as a positive control showing that the pNFκB-RFP unit sensitivity to NFκB induction.

Therefore, decoy plasmids can be used for more than the graded repression of gene targets. Decoy plasmids may also be used to observe the effect of decreasing levels of a TF on other TFs regulated by the sequestered TF. This may provide insight into TF interactions in regulation especially when coupled with other techniques such as qPCR or RNAseq.

3.3 CONCLUSION

This study explores the use of repetitive decoy plasmids for applications in the field of mammalian cellular engineering. For this work, we set up competition experiments observing effects on GOIs upon the addition of repetitive TFBS plasmids with shared TFBSs. GOIs used in this work were previously developed with an established capacity for binding and responding to TF to further verify the method. The use of flow cytometry, with its high-resolution capabilities, allows for quantification of individual cell characteristics, while simultaneously providing valuable insights into the broader population. This workflow allows the study of endogenous or engineered TFs, using available plasmids or using established methods for the design of synthetic promoters responsive to a TF of interest (Saxena, Bojar & Fussenegger, 2017). We applied this workflow here for a range of demonstrations in mammalian cells, scratching the surface of the potential applications of decoy plasmids.

For engineering biology applications, the development of robust, reliable, predictable and tuneable outputs is often sought after. Artificial TFs such as Tet TFs and dCas9 TFs are frequently incorporated as sensing and actuating modules. We showed that both are susceptible to graded sequestration upon co-transfection with a variety of TF decoys, and this was not the case when in the presence of controls containing pEmpty plasmids, strong promoters or scrambled TFBS.

We tested TetO containing TF decoy plasmids along with three different Tet-based TF circuits. The TF decoys consistently sequestered rtTA or TetR TFs shedding light on the functionality of the different circuits. For example, in the two TU system, we were able to draw conclusions as to the source of basal expression in this system and observe a reduced dynamic range. This reduced dynamic range was also observed with a PFL circuit, reducing the ultrasensitive response, which is usually a key feature of such a system. On the other hand, the use of TF decoys along with an NFL system showed a slight increase in dynamic range as the circuit works to counteract the drastic effect of high concentrations of TF decoys by expressing more repressor.

Therefore, the use of TF decoys may also be used to tune the output expression of circuits, as this is successful in the case of the two TU systems and PFLs. In the case of the NFL system, disruption of the closed-loop architecture instead induces cell death. Further mathematical modelling of these systems using the binding characteristics of the heavily studied TFs could be

interesting to understand the effects of decoys on GOIs and which factors (e.g. length, copy number or sequence) are most important.

The application of these to native circuitry may provide important breakthroughs in therapeutics, for buffering systems to high expression of TFs tuning output expression or disturbing closed loops for use in a new generation of anti-viral therapies (Chaturvedi et al., 2022).

Selective differential TF expression is often used to direct cells into different differentiation states. This often targets a few specific TFs. Yet, broad regulation is often not achievable with these methods. We display that a native TF can be sequestered at TF decoys with potent impacts on cell fate. The use of TFBSs accounts for the characteristics of TF such as promiscuity and cooperativity demonstrating the use of TF decoys to manipulate cellular responses. In our hands, the use of double-stranded oligos did not impact the cells as efficiently as our decoy plasmids and we attribute this to be due to cooperativity when binding at these sites. So plasmid TF decoys may provide more reliable sequestration of TFs. Further work directly comparing previously characterised oligo strategies with TF decoys and affinity studies may further prove this hypothesis. The use of TF decoys for the study of TF interactions may also provide an improved alternative to knockouts giving a more tuneable repression for studies where the knockout leads to lethality or for higher resolution information as to cellular response to a variety of TF levels.

We have therefore shown for the first time several uses of TF decoy plasmids in mammalian cells competing to bind both engineered TFs and native TFs. Artificial TF decoys can be used to engineer or study an expression response and analyse the importance of free TF levels and robustness to this parameter. TF decoys targeting engineered TFs can also be used to alter target gene expression in a dose dependent way. Increasing the number of repeats in series and in parallel can confer repression on targets.

Here we also display some of the potential uses of TF decoys for altering the effect of native TFs in mammalian cells. They can be used to reduce expression from target genes (in this case we used NRF2 targets in the genome) and induce cell responses such as cell death or to decipher interactions between TFs. This work may have implications in therapies due to NRF2 overexpression in some cancers such as pancreatic cancer (Lister et al., 2011). This work could be further expanded to understand sequence requirements for the most favourable responses. For example, STR sequences contain no spacing between repeats, for some purposes which require finer tuning of the system, testing a variety of different affinity/cooperativity properties may be required for best decoy mediated gene regulation.

Moreover, using decoys with stronger total affinity to TFBSs in decoys compared to active sites for gene expression may change the dynamics of these systems to more digital responses, where target genes are not activated until decoy sites are saturated (Buchler & Louis, 2008).

In addition, the work opens up new avenues for further research applications of this research including the quantitation of biology, studying position effects of TF decoys in the genome, sequestering RNA polymerases for studies of gene expression burden or switching on and off availability to decoys using epigenetics. These potential applications represent a wide range of possibilities for future research and highlight the versatility of the TF decoy approach. With further work, the response curves observed from the implementation of competition workflow as in this chapter, may also allow for a better understanding of TF binding properties, a research area actively being sought by many groups.

The potential of sequestering extracellular mammalian TFs that also function as paracrine or endocrine factors may provide a way to regulate gene expression signals between cells. In this way, one could imagine TF decoys being used extracellularly to control the tumour microenvironment and signals of tumorigenesis (Ung et al., 2014). Other than modulation of gene regulation, TF decoys can also be used as synthetic STRs to understand the roles of natural STRs. Understanding STR roles and building synthetic versions could be important for the assembly of a mammalian synthetic genome. As the synthetic yeast genome project has recently seen major strides taken, scientists are looking towards mammalian ones next and the tools required to build them (Boeke et al., 2016).

Overall, the work of this chapter represents an innovation in the field of mammalian cell engineering, providing a new tool for manipulating cellular responses and opening up new avenues for research. This work lays the foundation for the development of more sophisticated TF decoy systems for various applications within the field of cellular engineering.

4 MODIFYING GENE EXPRESSION USING MIRNAS

4.1 INTRODUCTION

Mammalian cells contain a suite of small regulatory non-coding RNAs (sRNAs) which regulate gene expression throughout cells. One group of key post-transcriptional regulators of gene expression is miRNAs (microRNAs) which are estimated to regulate at least 20% of endogenous genes (Donati, Ciuffi & Brandi, 2019). MiRNAs are 22 bp single-stranded RNA sequences, expressed from introns of genes (**Figure 4.1**) or polycistronically from the genome, to regulate gene expression in eukaryotes. Their post-transcriptional activity is achieved by binding mRNA transcripts within the 3' UTR (untranslated region) at target homology regions called MREs (miRNA recognition elements) (**Figure 4.1**) and reducing their expression either by blocking translation machinery or inducing degradation of transcripts. MiRNA can also bind other regions such as 5' UTRs or ORFs (open reading frames) of mRNA or on DNA at promoters. The functions at these positionings are not as clearly investigated but may require the same machinery (Fazi & Nervi, 2008).

There are many benefits which make miRNAs an appealing tool to engineer. Firstly, regulation can be local, maintaining intracellular homeostasis and also far reaching as cell-to-cell signals circulated in vesicles (Fazi & Nervi, 2008; Friedman et al., 2009). Interestingly, cells have even been shown to tailor the miRNA messengers secreted for circulation in extracellular vesicles (Chen, Xiong & Li, 2016). Moreover, miRNAs bind to multiple targets in a dose dependent manner, for broad regulation. In addition to this feature, more than one miRNA molecule can bind the same transcript enhancing the response. Incorporation of these characteristics plays an important role in the regulatory networks they influence. One regulatory network impacted by the binding of mRNA transcripts by miRNAs is gene expression noise where miRNAs act to buffer noise, prompting downstream processes such as cell fate (Adhikary et al., 2023; Del Giudice et al., 2018). While noise buffering is required due to native low expression of molecules and can be circumvented by overexpression, these methods are often resource heavy as they require the expression of proteins and can instead be achieved using miRNA expression control

(Lo et al., 2023) (Donati, Ciuffi & Brandi, 2019). This displays a way cells cleverly moderate burden in cells.

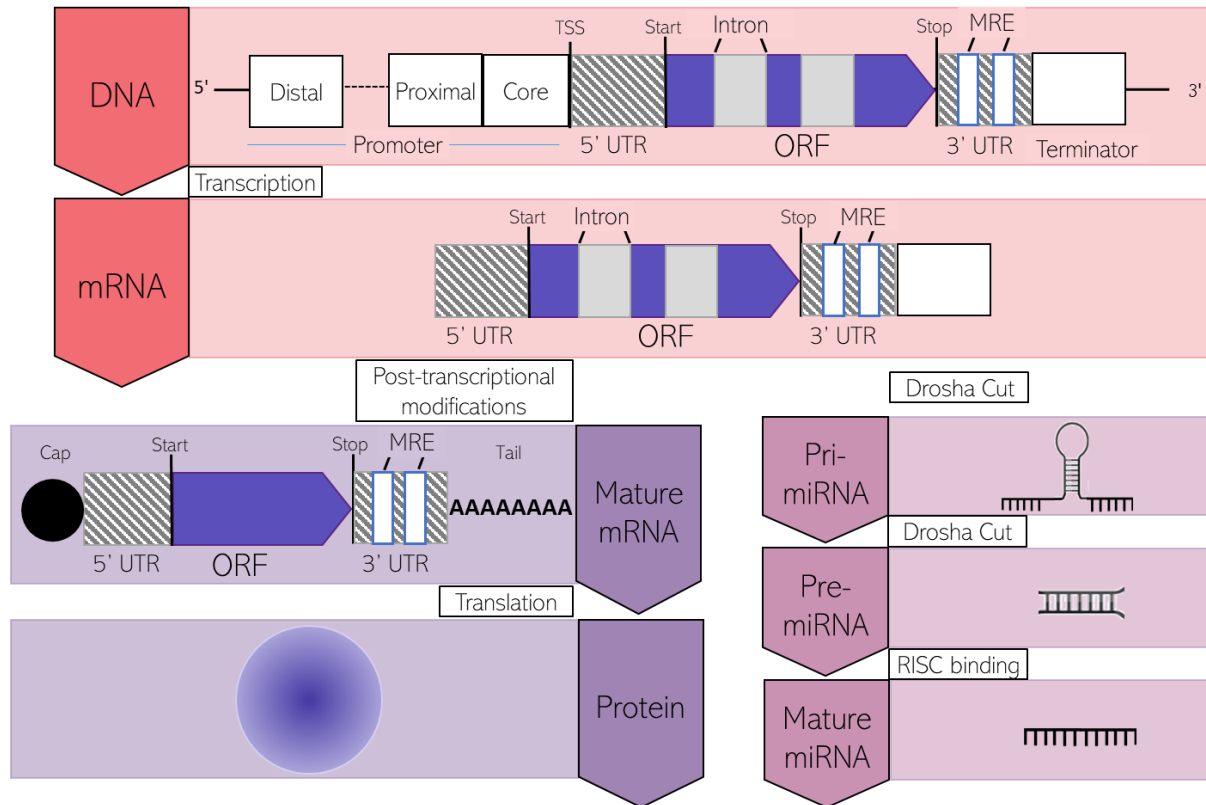


Figure 4.1: Expression from eukaryotic transcriptional unit. DNA is transcribed into mRNA in the presence of transcriptional activators and machinery. Transcripts are processed to add a 5' cap and 3' polyA tail and process intronic RNA which can contain primary miRNA (pri-miRNA). The then mature mRNA can be translated into protein or be degraded if miRNA binds within MREs (miRNA recognition elements). Pri-miRNA containing introns are processed by Drosha in the nucleus to make pre-miRNA which are further processed by RISC to exact gene expression control as mature miRNA.

Further designing synthetic miRNAs promises similar functions with fewer restrictions on the DNA sequence. This means that a miRNA of any sequence can be designed with binding to as many or as few targets as required.

These characteristics display miRNAs' critical roles in a cell's makeup justifying why their dysregulation is used as biomarkers of many diseases (Nam et al., 2014; Chen, Xiong & Li, 2016). Due to their flexibility, synthetic or native miRNAs are often used as components of diverse genetic circuits such as classifier circuits (Mohammadi, Beerenwinkel & Benenson,

2017; Sayeg et al., 2015) miRNA sponges (Ebert, Neilson & Sharp, 2007) and in feedback loops (Frei et al., 2020). Less tools exist that use activating functions of miRNAs. Recent advancements have seen miRNAs adopted for activation of genomic gene expression (Kingwell, 2021). An improvement on some of the state of the art by providing expression control without delivery of exogenous proteins or mRNA, which can be immunogenic within a small size for easier delivery. Similar to the impact of TFs in gene regulation, we see miRNAs as a strategic point of intervention to exert gene expression control, harnessing the diverse functions, small footprint and designability of miRNAs to create tools.

In this chapter, we delve into the multifaceted roles of miRNAs in synthetic biology, exploring their potential for broad and tuneable gene regulation in mammalian cells. While miRNAs are commonly used as repression modules in circuits, we believe that their capacity for activating targets at the transcriptional level represents an untapped resource in mammalian synthetic biology.

We sought to assess the multiple ways in which miRNAs can be used most effectively to control the outputs of synthetic circuits, with binding sites within 3' UTR or within promoter regions. To observe circuit effects, we use a fluorescent reporter as the circuit's output and another fluorescent reporter as a marker of miRNA expression. Transiently co-expressed components, often result in resource competition, which in turn impacts the expression of genes of interest (GOIs). This is a trend we anticipated seeing when miRNA is present. Therefore, we propose using an extra control to better differentiate the effects of miRNA and to understand how resource competition impacts the system. The use of flow cytometry allows us to quantify the reporter expression within samples. We further employ flow cytometry as a high-throughput method for screening the numerous iterations we have designed and used. The flow cytometry events data is then further analysed using our custom code to interpret (Code for Data Analysis by Binning.P.160) the effects of miRNA concentration on circuit dynamic range. This methodology allows us to understand the conditions in which gene regulation is most potent and the best way to use miRNAs for gene expression control.

4.2 RESULTS AND DISCUSSION

4.2.1 REDUCING EXPRESSION OF TARGET GENES.

4.2.1.1 Design and characterisation of circuit with miRNA expression repression

To understand the intricate dynamics and roles of miRNA regulation, we sought to design a miRNA-sensitive plasmid construct that could provide us with valuable insights into sensing microRNAs through repression of output.

To design this miRNA-sensitive plasmid construct, referred to as the sensor, we initially selected the most extensively studied MRE positioning within the 3'UTR (Lai, 2002; Fang & Rajewsky, 2011) which we assembled into a transcriptional unit expressing a green reporter (eGFP) under the control of a strong, constitutive mammalian promoter pEF1 α (The plasmid depicted in the top half of **Figure 4.2.a**). We expected to capture the miRNA effect through a decrease in signal. To facilitate design iterations of MREs within the 3' UTR with favourable outputs, we modified the 3' UTR by placing restriction sites at either end to allow for easy substitution of this region. This enabled us to vary conditions such as the number of MRE sites (multiplicity), optimal spacing between MREs for optimal miRNA binding, and explore sequence homology requirements with the miRNA of interest (**Figure 4.2.a**).

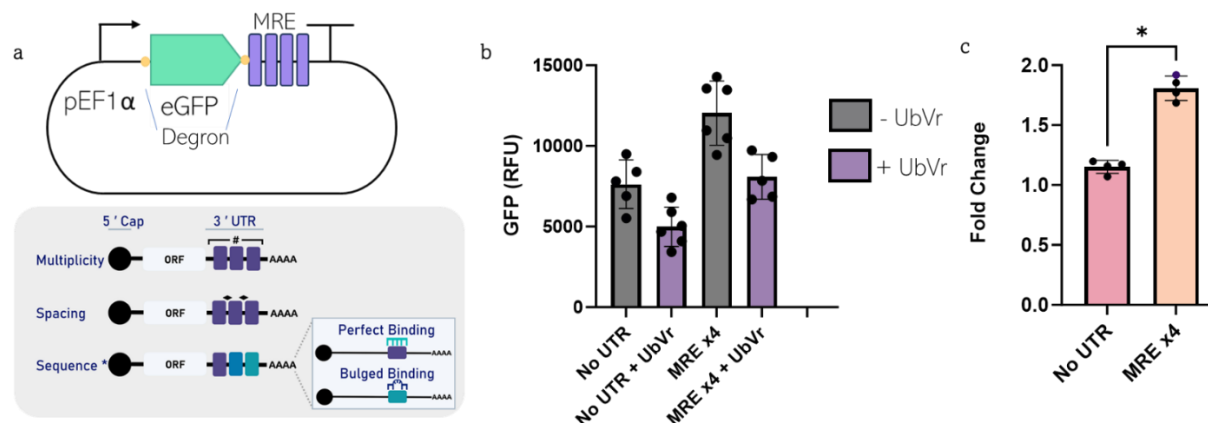


Figure 4.2: a. Design of the proposed miRNA sensor. A reporter is expressed from a strong mammalian promoter pEF1 α and degraded according to protein degrons present. MREs in 3'UTR should allow the sensitivity of the system to miRNAs. This 3' UTR region can vary in the number of MREs, the spacing between the MREs and the sequence of the MREs. **b.** Expression of sensor with 4 MREs vs no 3' UTR as a control, in the absence of miRNAs. Sensor designs with and without UbVr, a degradation tag to increase degradation rate of eGFP was also tested. **c.** Fold change was determined by dividing the fluorescence of samples in the presence of

miRNA mimic (Sigma Aldrich) by samples in the absence of mimic for constructs with no 3' UTR or 4 consecutive MREs.

Moreover, our design for a miRNA sensor did not include a transfection control, which is often a constitutively expressed reporter indicating the efficiency of plasmid transfection into cells (Huliák et al., 2012). This was to avoid resource competition and the potential overstating of miRNA effects due to the associated reduction in sensor fluorescence in the presence of a transfection control reporter (Di Blasi et al., 2021). This differs from previous studies which relied on the expression of a transfection marker on the sensor plasmid as an indicator of copy number and a measure of the success of the transfection method (Lemus-Diaz, Tamon & Gruber, 2018; Gam, Babb & Weiss, 2018; Wang et al., 2019). While we did not have a method of measuring plasmid copy number, we were able to ensure the success of the process of transfection through the parallel transfection of samples with the constitutive strong expression of eGFP (pCMV-eGFP plasmids) (Brown et al., 2014).

Constructs that were built included TUs with or without a 3' UTR. This was done to observe specific miRNA responses as well as different protein degradation signals. RNA regulation by degradation was known to occur at a faster rate than protein degradation (Schwanhäusser et al., 2011) and we thought that we might require active degradation of the fluorescent protein to observe these effects.

Firstly, to ensure the selected “No UTR” control reflected samples with no miRNA activity on the 3'UTR we transfected sensors with 4 MRE sites or none (No UTR) into HEK-293T cells in the absence of miRNA. The chosen sensor design was not compared to samples with “No 3' UTR” conditions in later experiments as the absence of a 3' UTR appeared to impact the output of the reporter, such that conditions with no UTR had lower protein expressed than conditions with a UTR, making it an inaccurate control for comparison to a system immune to miRNA effects (**Figure 4.2.b**). While it is relatively common to not include 3' UTRs in synthetic constructs built for mammalian cells the length and sequence of 3' UTRs play an essential role in mRNA stability, and the absence of a UTR in this case may have reduced translation efficiency (Tanguay & Gallie, 1996; Orlandini von Niessen et al., 2019).

To observe miRNA effects on the sensor plasmids, we purchased miRNA mimics, often used in the community to replicate the functions of miRNAs in cells (miRNA Mimics, Thermo Fisher Scientific). We co-transfected DNA constructs and miRNA mimics using Attractene transfection reagent (Qiagen).

As expected, the mimic in the absence of MREs within the 3' UTR had no significant effect on the reporter expression, showing almost no fold change between the presence or absence of mimic. Surprisingly, the addition of miRNA mimic to a sensor with 4 MREs in the 3' UTR led to a two-fold increase in output (**Figure 4.2.c**), instead of the decrease expected due to reduced translation. This may be due to limited cellular mRNA degradation machinery, as described by Cella et al., 2023. The authors suggested that miRNAs might cause upregulation in protein expression by saturating the mRNA degradation machinery, prolonging degradation wait times, and increasing the chances for mRNA translation.

Moreover, we further considered whether the presence of one protein degron (PEST – UbVr) to two protein degrons (PEST + UbVr) would help better resolve miRNA effects but decided against the use of two degrons to not inadvertently add further effects from bottlenecks elsewhere in cells.

4.2.1.2 Scanning Effects of miRNA Expression on a Repression Sensor

The balance between mRNA and miRNA levels is finely tuned in cells to determine how effectively miRNAs can control mRNA targets and therefore protein expression. This is such that the abundance of miRNA relative to mRNA determines expression control. Specifically, at low concentrations of mRNA, mRNA is saturated in miRNA and protein expression is most impacted. On the other hand, when mRNA targets are abundant, protein expression is not severely impacted by miRNA binding. Therefore, to engineer circuits that sense and respond to miRNAs, the threshold where the response is sensitive to changes in miRNA levels is sought after (Mukherji et al., 2011). As we had little visibility of relative amounts of miRNA using the mimic system above, we shifted to a plasmid system where miRNAs were expressed from the intron of a DsRed reporter (Lillacci, Benenson & Khammash, 2018). This allowed us to use the expression of DsRed as an indicator of the amount of miRNA in the system.

Therefore, in order to refine and optimize the proposed engineered circuit to be responsive to miRNAs, we aimed to study the miRNA expression level required to approach the threshold where there is sensitivity to changes in miRNA levels. We next aim to understand sensor responses to various amounts of miRNA and the range of miRNA the sensor is most sensitive to.

The DsRed containing intronic miRNA expression was obtained from the Professor. Khammash lab and modified to contain miRNA of interest using annealed oligos and promoters of different

strengths (**Figure 4.3.a**). Plasmids containing different strength promoters were assembled to proxy different intracellular miRNA expression levels. The promoters used included, constitutive promoters of varying strengths, pJB (weak), pACTB (medium), pEF1 α (strong, native), pCMV (strong, viral) and the dox inducible promoter pTet (strong at high levels of inducer, synthetic promoter).

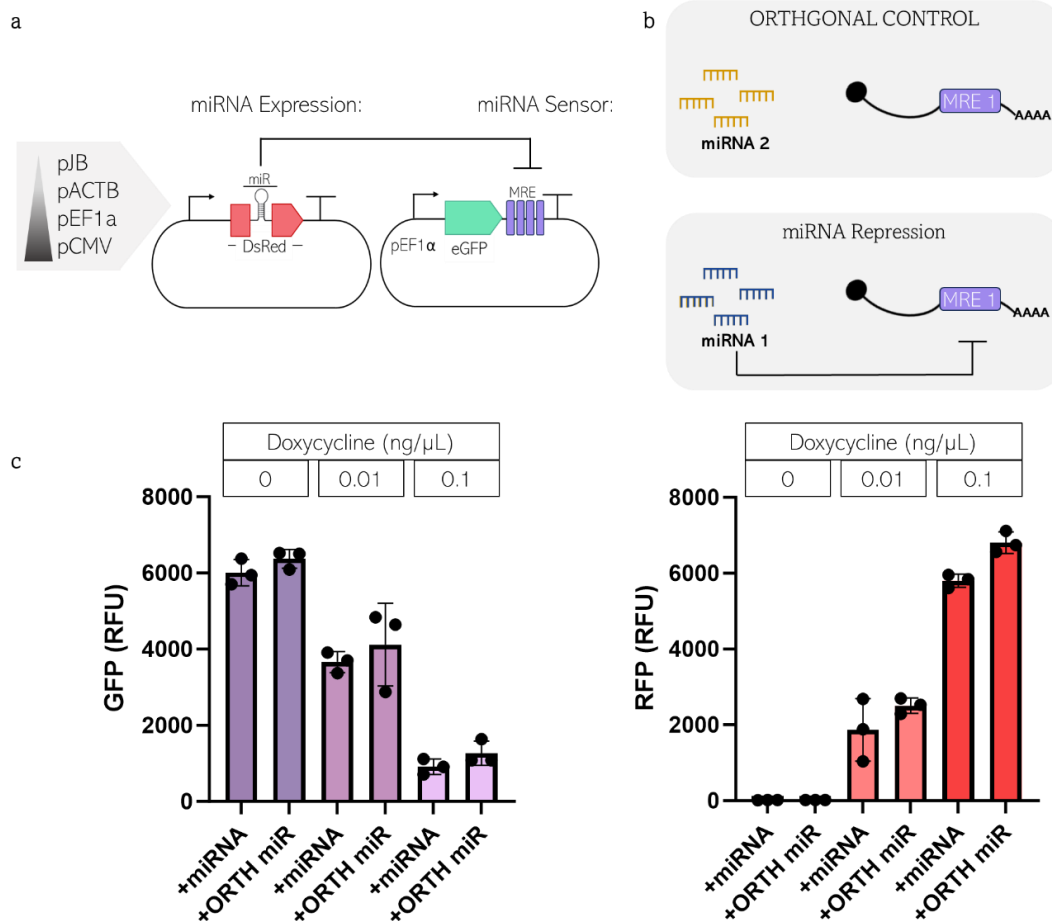


Figure 4.3: Identifying miRNA Effects. **a.** The experiment set-up transfects a plasmid for expression of miRNA of interest under the control of promoters of different strengths such that the expression of the miRNA is coupled to the expression of a red reporter (DsRed). The miRNA should bind and repress expression of the co-transfected sensor plasmid which contains MRE sites within the 3'UTR. **b.** Accounting for resource competition of co-transfected plasmids with orthogonal control. MiRNA with (for miRNA repression) or without (orthogonal control) complementarity to the MRE region of the sensor is transfected into cells. In the case of the orthogonal control transfection (top panel), the miRNA expressed does not bind to MRE and therefore should not have any miRNA effects on the mRNA. In the case where miRNA is complementary to MRE (bottom panel), the miRNA binds its complementary region within the mRNA and causes degradation of the mRNA target or prevents translation. **c.** Observing resource competition between miRNA expressing plasmid and miRNA sensor. Co-transfection of sensor containing 4xMREs with homologous miRNA or orthogonal miRNA. The two constructs compete for gene expression resources in the cell such that an increase in the expression of the red reporter by increasing the concentration of dox inducer (miRNA

expression cassette, red bars) leads to a decrease in the expression of the green reporter (sensor plasmid, purple bars).

Using miRNA expression plasmids introduces resource competition between the expressed reporters (Di Blasi et al., 2023). To differentiate between the effects of miRNA and burden in samples, we added controls where the sensor is co-expressed with an orthogonal miRNA expression plasmid that does not have complementarity in the MREs of the sensor, we called this the orthogonal control (**Figure 4.3.b**). The orthogonal control therefore has a similar burden background as conditions where the corresponding miRNA is expressed, but no miRNA effects allowing us to differentiate between miRNA effects and burden.

As expected, an increase in expression from pTet inducible miRNA expression cassette upon the addition of increasing concentrations of inducer dox (**Figure 4.3.c** right graph) was at the expense of the sensor expression (**Figure 4.3.c** left graph). This is such that expression of the sensor (GFP) decreases with increased expression of DsRed/miRNA cassette regardless of which miRNA was expressed. Moreover, there is no significant difference between complementary or orthogonal miRNA conditions, indicating we cannot resolve miRNA effects on sensor fluorescence in this set-up. Therefore, if the orthogonal control was not used, the effects of burden could have been misconstrued as miRNA mediated repression. It is also important to note that controls expressing orthogonal miRNAs should not respond due to basal expression of miRNAs in HEK cells from native/genomic regions as proven by (Gam, Babb & Weiss, 2018). Once burden was accounted for in experiments, sample geometric mean measurements of GFP fluorescence did not show a significant difference between samples containing miRNA of interest or the orthogonal miRNA.

To gain higher resolution insights into the optimal activity of the sensor in response to varying miRNA expression, two distinct approaches to the transfection of these plasmids—co-transfection and poly-transfection—were employed, as illustrated in **Figure 4.4**. The choice between these methods results in different plasmid transfection distributions for studying miRNA-mediated control over sensor function. While co-transfections mix plasmid DNA before liposome formation (**Figure 4.4.a**) for less variation in copy number of transfected components, poly-transfection (**Figure 4.4.b**) creates a pool of liposomes per plasmid which are then mixed for a wider variety of plasmid number combinations. Poly-transfection results in higher variation in plasmid copy number (Gam et al., 2019), as shown in a representative flow scatter plot of a sample poly-transfected with miRNA expression and miRNA sensor plasmid in **Figure 4.4.b** compared to the scatter plot of the same plasmids co-transfected in **Figure 4.4.a**. As similar sensor designs are commonly used in circuits which require miRNA response,

we believe varying plasmid ratios in transfections may shed light on sensitivity window of these systems.

To observe the effects of miRNA concentration on sensor abundance, cell fluorescence intensity events, measured by flow cytometry were grouped into bins according to ascending miRNA abundance using the proxy of DsRed expression **Figure 4.4.c**.

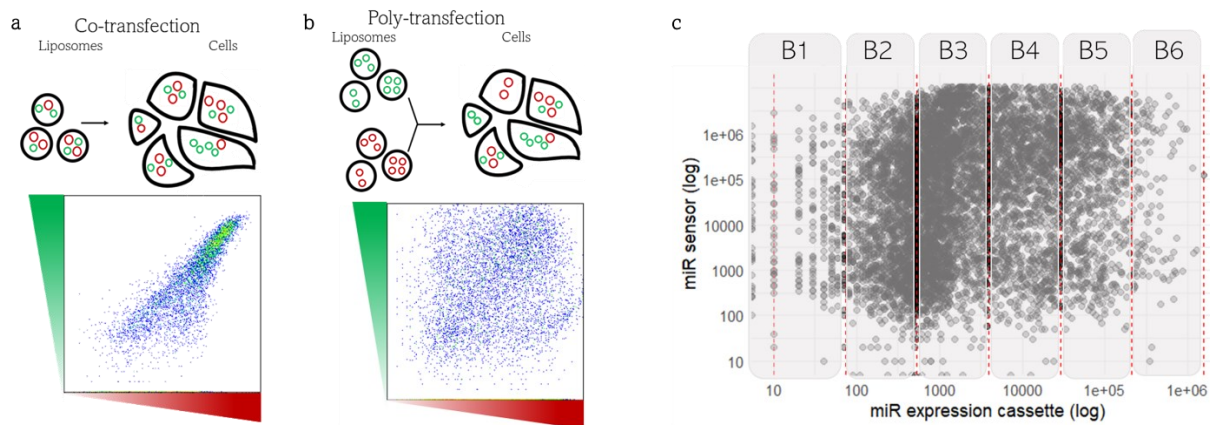


Figure 4.4: Experimental workflow followed to explore miRNA sensor sensitivity. To find the best conditions to observe various miRNA effects, plasmids can be co-transfected and analysed in the following way. **a.** DNA from both plasmids are mixed to certain ratios before the addition of the transfection reagent that forms liposomes. The resulting liposomes have a relatively narrow spread of plasmid mixtures. **b.** In another method, poly-transfection can be performed where pools of liposomes are made with each plasmid alone and then mixed at required ratios before addition to the cells. **c.** To further observe the effects of varied miRNA concentrations analysis using a custom-made R-script is run. This script uses obtained flow cytometry data and sorts events into bins depending on miRNA expression. This information can then be used for different statistical analyses and visualisation to observe miRNA effects.

Binning fluorescent events by reporter affiliated with miRNA expression is a relatively common approach for assessing sensor responses (Gam et al., 2019; Lemus-Diaz, Tamon & Gruber, 2018). Our method differed from other studies experimentally by not including a reporter on the sensor plasmid as a measure of sensor copy number in cells, to account for this we designed our own R code for analysis (Chapter 7, Section 1). Before implementation of the code events in singlets gate for each sample (see materials and methods) were extracted into new FCS files, using FlowJo software. The FCS file paths were then inputted into the code in pairs, each sample with its orthogonal control. For example, the file path of the sample co-transfected with a sensor with 4 MRE sites and intermediate expression of complementary miRNA followed by the file path of the sample co-transfected with the same sensor but with intermediate orthogonal miRNA expression cassette. The code reads flow files (.FCS), groups events into bins of miRNA

expression levels and outputs two data frames. The first data frame contained sample pair identifier (for example: pACTB-miR + MREx4 sensor) and within this for each sample, a table containing the events identifier, GFP value, RFP values and the assigned bin number. The second data frame summarised sample information by averaging the GFP expression of events in each bin of the sample. This information is then utilised to compare samples to respective controls, generate summary statistics and generate graphics. The exact code used is attached in Supplementary Information Chapter 7 Section 1 and a summary of the key findings are discussed below.

4.2.1.3 Scanning Effects of miRNA Expression on a Repression Sensor by Co-transfection

Co-transfection of miRNA expression cassettes and sensor cassettes allows the specific mixing of different plasmid ratios to explore circuit effects, as seen by the narrower distribution of GFP/RFP fluorescence combinations of co-transfected plasmids (**Figure 4.4.a**). We used co-transfection to study four different proposed miRNA sensors containing either 4 or 8 sites for miRNA recognition 100% complementary to miRNA1 or miRNA2 corresponding to target MRE1 or MRE2. Different sensors were transfected with their respective miRNA or an orthogonal miRNA. HEK-293T cells were analysed by flow cytometry 48 hours after transfection. We expected to see a shift in the number of measured events with less GFP fluorescence indicating events with repressed translation of sensor mRNA.

Experiments using ACTB intermediate expression are displayed in **Figure 4.5**, as this promoter was proven to have optimised resource usage per gene expression (Di Blasi et al., 2023). Moreover, reduction of interference from resource competition in these samples improved the resolution of sensitivity to miRNAs as displayed by work by MRes student Ruihan Bai, 2021. Events were analysed using code (Supplementary Information Chapter 7 Section 1) and visualised in **Figure 4.5** and **Figure 4.6**.

The violin plots in **Figure 4.5** display the spread of sensor expression in each miRNA expression bin when miRNA is expressed from the intermediate ACTB promoter, for sensors with 4 MRE sites. The sensors each contained 4 sites (MREx4) but specificity to 2 different miRNAs (miRNA1 in the top panel and miRNA2 in the bottom panel). Notably, the system led to the observation of different trends for the MRE1 sensor compared to the MRE2 sensor, with a general decrease in MRE1 sensor expression but an increase in MRE2 sensor expression in the presence of complementary miRNA compared to when the system is exposed to orthogonal

miRNAs. This is observed in the violin plots by a shift downwards in the density of cells in each bin in the case of the MRE 1 sensor whereas the MRE 2 sensor shows an upward shift in cell density showing a higher frequency of cells with higher sensor expression. The mean fluorescence of GFP fluorescence in each bin is also displayed in **Figure 4.5 c,d** for the MRE1 sensor or MRE2 sensor respectively.

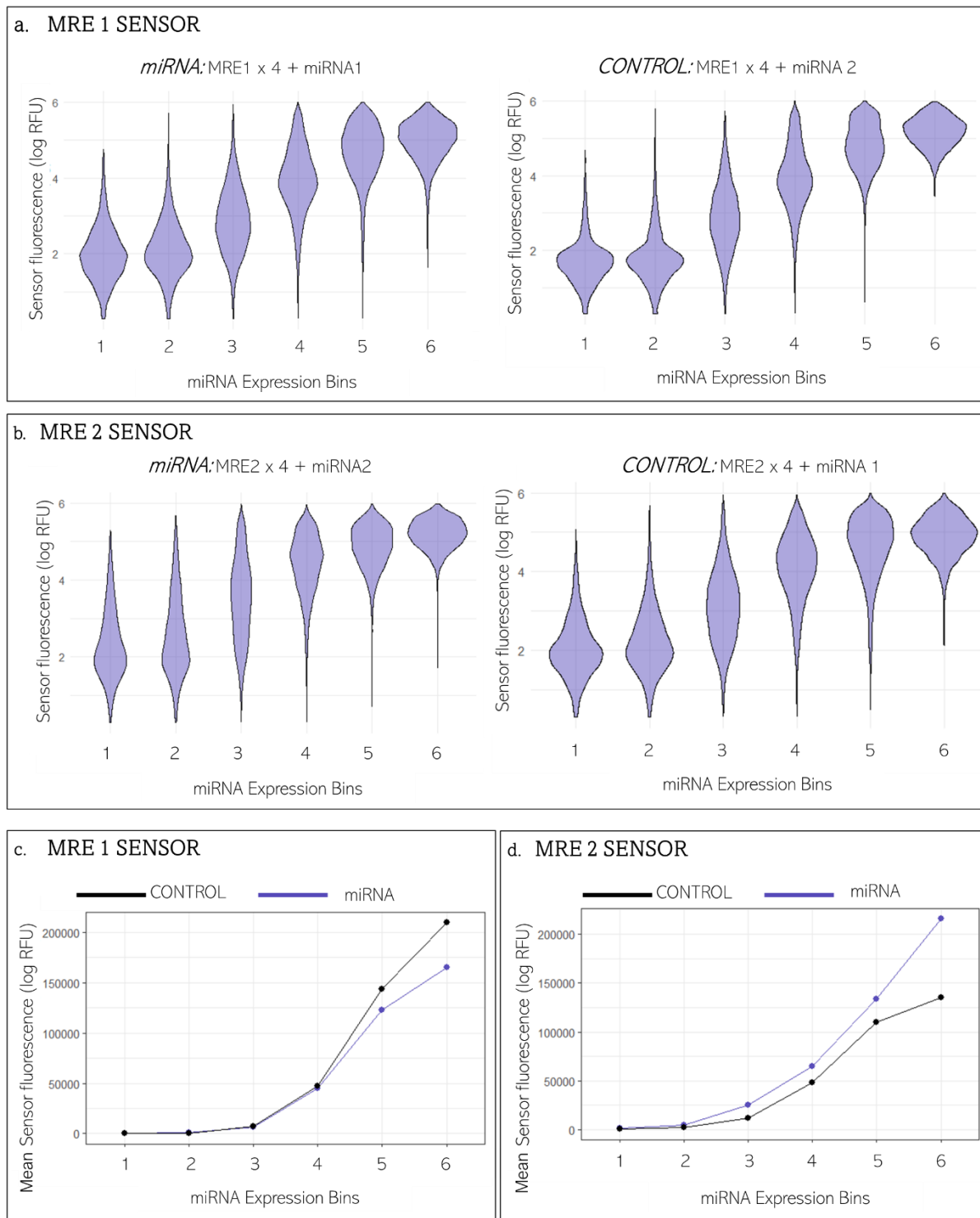


Figure 4.5: Results of a sensor with MREs in 3' UTR. Expression output of the experimental set-up where pACTB-miRNA and 4x MRE sensor were co-expressed for two sensor sensitivities. miRNA is expressed from the intermediate ACTB promoter and the sensor is expressed from the strong EF1a promoter **a.** Violin plot of sensor carrying miRNA1 complementary sites, showing the range of sensor fluorescence for different miRNA expression bins. This is such that the plot on the left expressed the complementary miRNA1 and the plot on the right expressed a miRNA with no sequence complementarity. **b.** Violin plot of sensor carrying miRNA2 complementary sites, showing the range of sensor fluorescence for different miRNA expression bins. This is such that the plot on the left expressed the complementary miRNA2 and the plot on the right expressed a miRNA with no sequence complementarity. **c.** Mean sensor fluorescence for each bin of samples expressing the control miRNA (black) or miRNA1 (purple) for the sensor with sites for miRNA1 binding. **d.** Mean sensor fluorescence for each bin of samples expressing the control miRNA (black) or miRNA2 (purple) for the sensor with sites for miRNA2 binding.

We, therefore, displayed how the orthogonal control samples and binning into groups of miRNA expression levels can be used for the detection of miRNA effects in the absence of a transfection marker, accounting for resource competition. Using this approach, we calculated fold changes between each sensor and control pair per bin (**Figure 4.6**) and observed trends in miRNA response across combinations. We observe that most increases in sensor expression compared to control occur at lower miRNA expression (bins 1-3), while most repression of sensor expression compared to control occurred at higher miRNA expression (bins 4-6). This does not support the mRNA queuing at degradation machinery model, where excessive miRNA targeting of transcripts to degradation machinery leads to a queue prolonging mRNA lifetime in cells. If this were occurring in samples, we would expect to observe most increases in sensor compared to expression plasmid to occur in the higher bins (bins 4-6), with higher miRNA expression. Therefore, the miRNA expression promoter's strength and the miRNA's sequence significantly influenced the sensor expression, leading to the observed variations in the fluorescence distribution across different miRNA expression bins highlighting the importance of considering the sensor, promoter pair.

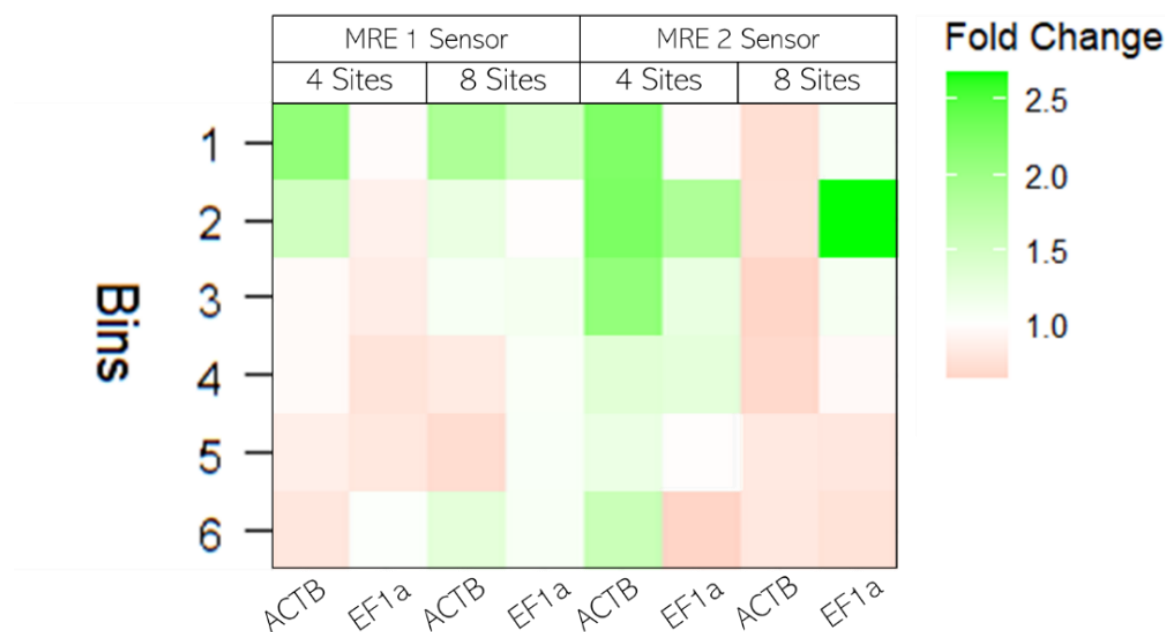


Figure 4.6: Heatmap of fold change between miRNA expression and control conditions of two miRNA sensors with two expression levels. Green locations indicate an increase in sensor compared to a control while red positions indicate conditions where sensor expression is less than the control. White positions indicate there was no change in the expression and so these conditions represent positions where the sensor was not responsive to miRNA.

The results obtained from co-transfection studies provide some insight into the response to miRNA expression on a fixed amount of sensor. But previous studies into the response of miRNAs on repression sensors, also describe effects by intracellular sensor copy number, indicating a global reduction in fluorescence may not be visualised at the population level as in our dataset (Mukherji et al., 2011; Gam, Babb & Weiss, 2018; Lemus-Diaz, Tamon & Gruber, 2018).

4.2.1.4 Scanning Effects of miRNA Expression on a Repression Sensor by Poly-transfection

Therefore we next used poly-transfection (Gam et al., 2019) in **Figure 4.8**, which uses single cell resolution of flow cytometry to detect varied stoichiometries of transfection components. We sought to maximize sensor, expression plasmid combinations, by poly-transfection **Figure 4.4.b** to understand more specifically the conditions in which the sensor performs effectively and predictably to expression of its miRNA.

In the original poly-transfection paper (Gam et al., 2019), coverage of most of the parameter space is displayed by plotting miRNA expression reporter vs transfection marker (on sensor plasmid). They then compare this scatter plot with miRNA expression reporter vs sensor reporter and observe a downward shift which they associate with miRNA effect. Similarly, we use our orthogonal control samples as an indicator of coverage of the parameter space and observe miRNA effects by comparing them to the scatter of complementary miRNA samples (**Figure 4.7**). A slight downshift can be detected when comparing the density plot of the complementary miRNA with 46% of samples occurring in the top half with high expression of sensor whereas 55% of events occur in the top half using the orthogonal control (density plots in **Figure 4.7**). This shift can be moreover observed in the histogram (**Figure 4.7**), where the density curve of complementary miRNA is shifted to the left. The detectable shift in the distribution of events, is similar to the trend observed in the poly-transfection paper with a transfection marker, indicating how our orthogonal controls can be used instead of a transfection marker to detect miRNA effects. The range of combinations achieved using poly-transfection can further be increased by transfecting larger amounts of cells and acquiring a higher number of events on the flow cytometer. Lastly, it is of note the fluorescence distribution in the red channel (YL2) was similar between samples with complementary miRNA expression or samples with orthogonal miRNA control indicating effects of miRNA or different levels of competition did not occur (**Figure 4.7**).

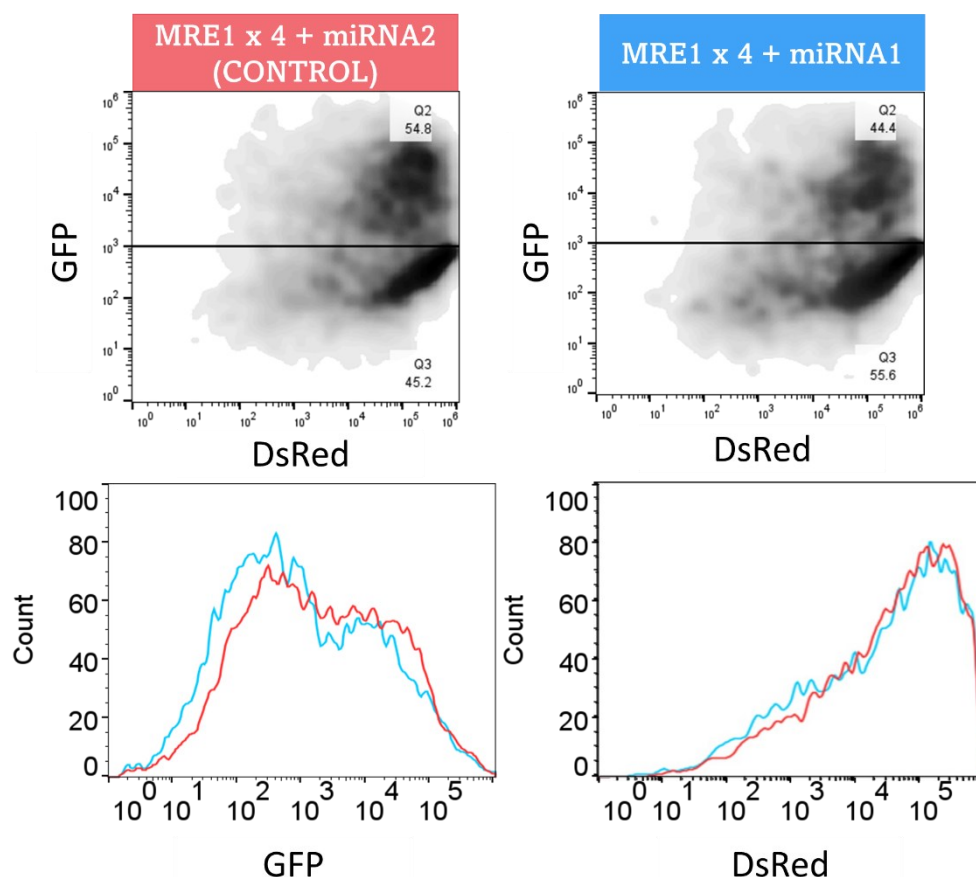


Figure 4.7: FlowJo-generated graphs illustrating the transfection landscape. BL1A represents the channel detecting eGFP expression from the miRNA sensor, while YL2A captures DsRed expression from the miRNA-expressing plasmid. A scatter plot depicts the coverage of transfection space for orthogonal miRNA samples poly-transfected with a miRNA sensor and a plasmid expressing orthogonal miRNA under pEF1 α (depicted in red) to be compared to miRNA-affected samples poly-transfected with a miRNA sensor and a second plasmid expressing complementary miRNA under pEF1 α (depicted in blue). Additionally, a histogram illustrates the distribution of sensor eGFP fluorescence (BL1-A channel, bottom left) in samples with complementary miRNA to the sensor (depicted by the blue line) and those without complementary miRNA (depicted by the red line). The histogram on the bottom right shows the distribution of DsRed fluorescence (YL2 channel) of control (red line) or complementary miRNA (blue line).

Similar to co-transfection results, no generalisable conclusions can be drawn as to which sensor designs (4 or 8 sites), or miRNA expression plasmid give the best sensor effects (**Figure 4.8.a**). However, it appears the majority of the resulting combinations cause repression of the sensor. For example, miRNA expression under pU6, an RNA polymerase III promoter, seems to increase miRNA1 sensor with 8 MREs expression, whereas it also seems to reduce miRNA2 sensor expression with the same number of MREs **Figure 4.8.a**.

One trend that reoccurs for both sensors with 4 MREs is when the miRNA is expressed from the strong pEF1 α promoter (**Figure 4.8.a**). A closer look at the distribution of sensor fluorescence for each bin, when miRNA is expressed from pEF1 α (**Figure 4.8.b,c**), shows there is a shift in the bulk of events from a more spread out distribution in control samples with orthogonal miRNA expression (right violin plots of **Figure 4.8.b,c**), to having the bulk of events with lower sensor output. This downward shift shows that while the mean of the population is not drastically altered the distribution of events within a bin does.

This study, unlike studies before it, tries to visualise repression effects on miRNA sensors while mitigating resource competition and testing with miRNAs expressed from a suite of promoters with different strengths. This is an important design improvement as resource competition interferes with the expression of sensor masking results. This study takes into account gene expression resource competition that occurs between plasmids by adding a control with a similar burden background for comparison. These findings suggest that the choice of promoter for miRNA expression and the number of MREs in the sensor design play a role in influencing sensor output and should be carefully considered for required designs and outputs. Very few trends can be generalised, one of which is that repression of sensors more often occurs in bins with higher miRNA expression. So, the function of this sensor layout may be difficult to achieve and predict. Further investigation and fine-tuning of parameters may help understand the underlying mechanisms governing the observed trends and optimize this sensor design for enhanced specificity and sensitivity in miRNA detection. Improving the sensor design to increase the circuit's output in response to detected reporters could enhance the design. This improvement may enable the detection of intracellular miRNAs at the population level.

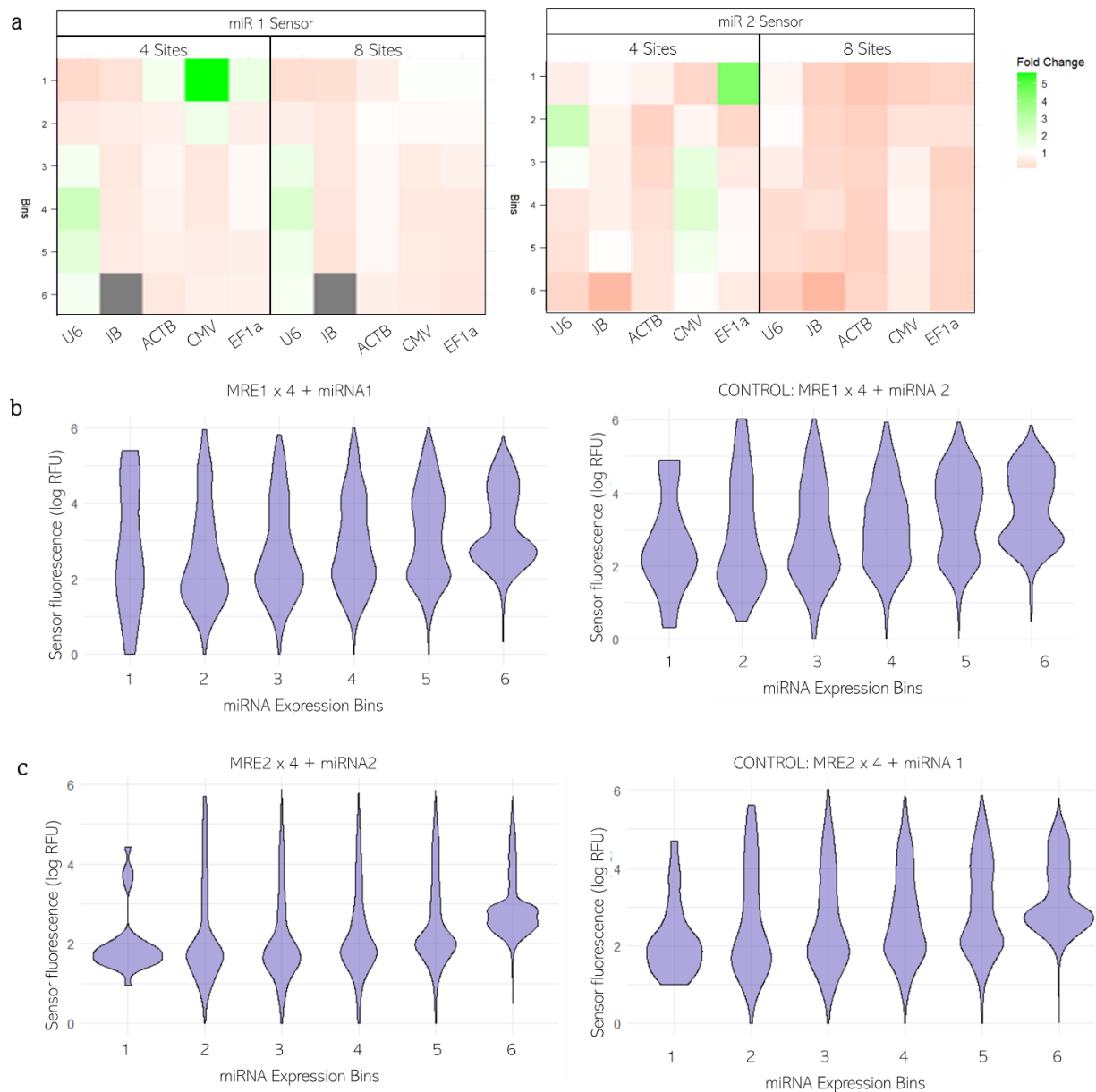


Figure 4.8: Assessing miRNA sensor sensitivity to changes in miRNA of interest using poly-transfection. **a.** Heatmaps displaying fold change of sensor vs control conditions for the MRE1 sensor (left) and MRE2 sensor (right). Positions in shades of green show an increase in sensor compared to control. Positions in white are conditions where there is no change in fluorescence, and positions in shades of red show a decrease in sensor fluorescence compared to the control, such that the stronger the pigment of the colour the more different the sensor and control values. **b.** Violin plot of a sensor with 4 MREs with 100% homology to miRNA1 in the presence of miRNA1 (left) or an orthogonal miRNA2 (right). **c.** Violin plot of a sensor with 4 MREs with 100% homology to miRNA2 in the presence of miRNA2 (left) or an orthogonal control miRNA1 (right).

4.2.2 INCREASING GENE EXPRESSION OF TARGET GENES. MIRNA ACTIVATION (MIRA)

Small RNAs, such as miRNAs, are traditionally known for their role in post-transcriptional gene regulation by inducing mRNA degradation or translational repression. However, many have reported small RNAs can also lead to transcriptional activation, although these interactions are less common and context specific. The mechanism that leads to transcriptional upregulation relies on some of the same proteins as miRNA post-transcriptional repression, but little else regarding the mode of action is understood. Also, these context specific effects seem to achieve relatively low activation. Extending the use of small RNA for predictable and reliable activation of transcription would provide a means to bypass the issues encountered by the repression sensor. The designs developed which increase output ensure that resource-depleting co-expression will not confound with miRNA effects on GOI. Additional advantages conferred by a miRNA gene activation include its reduced expression burden in comparison to expression of activator proteins, and a shorter half-life, enabling faster switch-off kinetics.

Activation by native miRNAs has been observed at specific locations in the genome throughout the core (Matsui et al., 2013; Dharap et al., 2013; Zhang et al., 2014), proximal (Huang et al., 2012; Wang et al., 2016) and distal promoter elements (Qu et al., 2015) (**Figure 4.1**). However, miRNAs/targets do not have to be naturally occurring. Synthetic double-stranded RNAs have been designed to target the genome for activation of targets (Li et al., 2006; Janowski et al., 2007), displaying the potential of this strategy for multiple control. This also removes the constraint of requiring native miRNA/promoter sequence homology and opens the potential to design synthetic promoters with RNA guided gene activation. Activation by miRNAs targeting proximal and distal regions of promoters act by altering genes epigenetic context yet less is known about the mechanism in which genes are activated within the core promoter other than that many of the same proteins as the repression pathway play a role.

The following study investigated the potential use of miRNAs in activating the transcription of synthetic promoters. Our approach adapts the use of native promoter/miRNA couples into plasmids and explores methods to extend the capability to induce transcriptional activation for a tuneable high dynamic range of activation. To achieve these objectives, we focused on the strategic engineering of miRNA binding sites within promoter elements conserved in synthetic promoters. While in cells small fold changes elicit significant effects on cells (Janowski et al., 2007) enhancing dynamic range improves usability in the mammalian synthetic biology community. We envisage providing these systems within plasmids for transfection of several

copies per cell may be one way to improve range. In pursuit of these goals, we selected two native promoters/miRNA pairs and designed two synthetic promoters/miRNA pairs to experiment with gene activation within plasmids.

The native promoters we chose were that of the RelA (v-rel avian reticuloendotheliosis viral oncogene homolog A, a subunit of NF- κ B) gene in rats, shown to be responsive to rno-miR-324, and the IL-2 (interleukin 2) promoter of humans, endogenously activated by the miRNA hsa-Let-7i. One thousand base pairs of genomic sequence ending at the start codon of these genes were amplified from genomic DNA of PC-12 (for pRelA) or HEK-293T (for IL2) and placed upstream of an eGFP reporter and SV40 polyA (**Figure 4.9.a**). The resulting pRelA-eGFP construct was co-transfected into HEK-293T cells along with a construct co-expressing a red reporter (DsRed) and intronic rno-miR-324 under the control of the pTet promoter. The miRNA expression cassette expresses reporter and miRNA upon the addition of dox. As the amount of dox was increased from 0.01ng/ μ L (low) to 0.05ng/ μ L (medium) or 0.1ng/ μ L (high), eGFP expression increased a maximum of five-fold compared to the “only” condition where rno-miR-324 is not induced (**Figure 4.9.b**). Similarly, we observed approximately a two-fold increase in eGFP expression downstream the pIL2 promoter, when the cell was exposed to hsa-Let-7i, in this instance expressed from the RNA polymerase III promoter U6. This system appeared to have a high background fluorescence in the “only” condition where no RNA is expressed. This is likely as this promoter was isolated from HEK-293T and therefore may be activated by its endogenous transcription factors.

To enhance the versatility of applications for miRNA-activated promoters, a larger range of activation is desirable. One such method we tested appended RNA with MS2 RNA loops. These loops serve to attract MCP proteins, which are fused to transcriptional activators or repressors. In this way when engineered RNAs bind their complementary sequences in DNA, effectors can be recruited to promote expression through MS2-MCP interactions.

To test whether this strategy could improve the dynamic range of RNA guided activation, we expressed engineered RNA to target either the 7 repeats of TetO in the pTet promoter or the 8 repeats of an artificial guide sequence in a psg promoter usually used with dCas9 activation (addgene plasmid #55197). These RNA plasmids were co-expressed with the activator MCP-VPR (VP64, p65, Rta). The co-transfection of these components resulted in distinct effects on the two promoters. Specifically, there was a two-fold increase in expression observed on the psg promoter in the presence of the engineered RNA/activator system compared to the presence of RNA alone, whereas there was no statistically significant difference in expression from the pTet promoter in the presence of the activator.

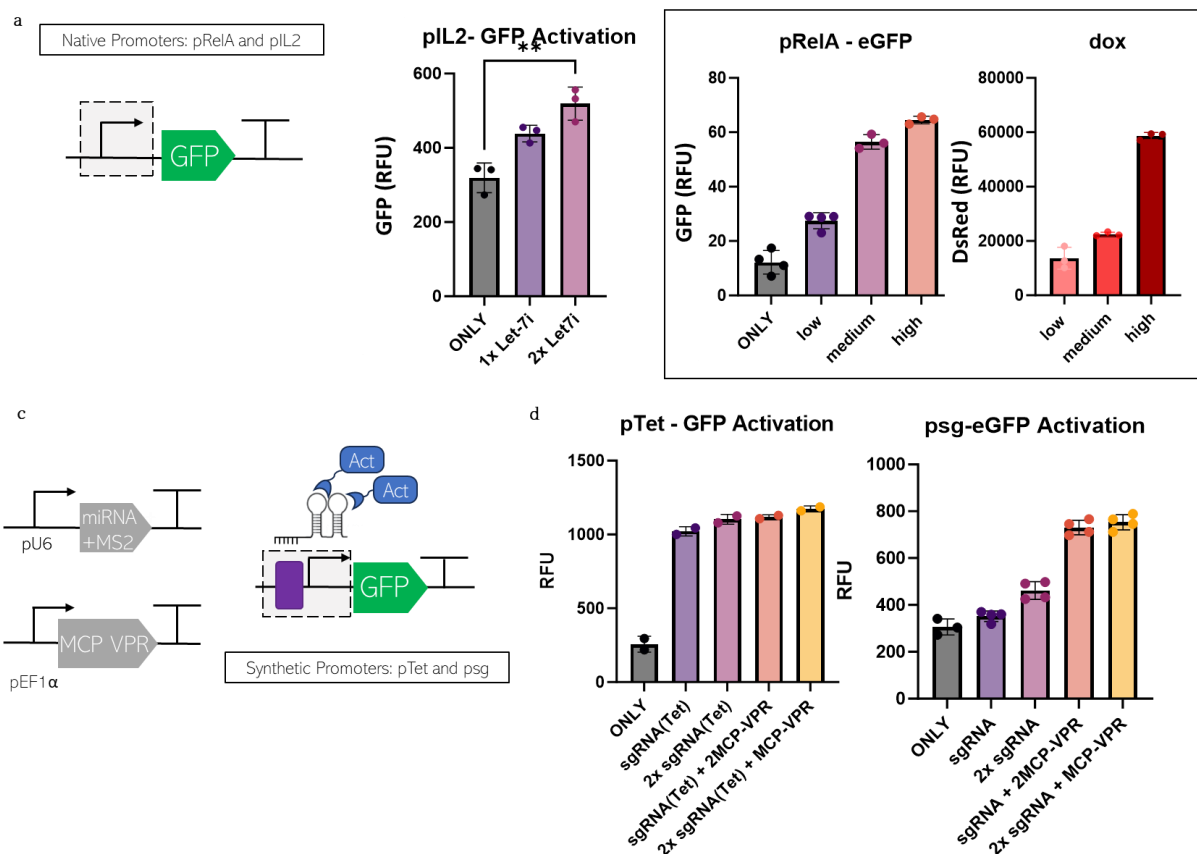


Figure 4.9: Engineering gene activation using miRNA inducers. **a.** Adapting natural promoters with miRNA sensitivity into plasmid systems increases the copy number in cells. **b.** IL2 promoter (pIL2) is isolated from HEK-293T cells and placed inframe and upstream an eGFP reporter. Activation of the IL2 promoter in the presence or absence of pU6-Let7i shows significant upregulation of eGFP expression. Another native promoter isolate from rat gDNA is the RelA promoter isolated from PC 12 rat cells is placed inframe and upstream of an eGFP reporter. As the amount of dox inducer of miRNA expression cassette (pTet-rno-miR-324) increases more DsRed reporter and miRNA were expressed. The increase in miRNA expression also leads to an increase in the expression of the pRelA controlled eGFP reporter. **c.** Design of an amplifier of miRNA activation on synthetic promoters which is based on the interaction between MS2 loops appended to targeting RNA sequence and MCP-activator. **d.** The application of a circuit on two synthetic promoters does not always lead to an increase in expression.

Further studies to characterise the optimal positions and sequence requirements for activating RNA for best activation are required. So far, we have shown native promoters can be moved into plasmids and trigger small increases in gene expression. The one activator system we attempted did not improve readout consistently and further work to improve the characteristics of this system would be advantageous. For applications which require small activation of GOIs, activating RNA may be a resource budget friendly alternative to other synthetic TF proteins. For sensing applications of small RNAs including miRNAs activation circuits improve on the first

sensor design described as an increase in expression is required which is easier to detect. This gets over the significant hurdles faced to untangle miRNA effects from resource competition for gene expression resources and degradation machinery. Therefore, there is potential for the use of miRNAs for the activation of targets for sensing and activation of GOIs.

4.3 CONCLUSION

In this chapter, we have constructed and examined two distinct types of miRNA responsive circuits: one where miRNA binding results in post-transcriptional repression of output expression, and another where miRNA binding triggers transcriptional activation of output. Work to develop these tools can provide valuable insights into the construction and optimization of miRNA responsive circuits, highlighting the potential of these circuits in various applications. MiRNA systems can be useful components of gene circuits, biosensors of diseases and control of noise in systems.

For the repression circuit, we investigated the impact of resource competition on results and scanned miRNA expression levels from co-transfection and poly-transfection experiments with code to bin miRNA reporter proxy. Our findings indicate that the choice of promoter for miRNA expression and the number of MREs in the sensor design significantly influence sensor output in different ways. This is consistent with other reports where “low” sensor (Gam, Babb & Weiss, 2018) and “high” sensor (Gam et al., 2019) designs are required to observe the effects of miRNAs. This is a significant challenge to use in this biosensor design. These factors should be meticulously considered for desired designs and outputs. However, the functionality of this sensor layout may be challenging to achieve and predict. Further investigation and parameter fine-tuning may aid in understanding the underlying mechanisms governing the observed trends to optimize future sensor designs for enhanced specificity and sensitivity in miRNA detection. A notable achievement of this work is the demonstration of how to establish a control to verify miRNA results in the absence of resource competition using orthogonal controls. The implementation of such controls may act as alternatives to transfection markers.

Therefore, miRNA repression sensors do not function similarly depending on the complementarity and multiplicity of MREs within the 3' UTR, although the orthogonal control provides a more accurate control for the dataset. Understanding what guides diverse responses to miRNAs would encourage more groups to use miRNAs more reliably as modules in their circuits.

On the other hand, in the second circuit layout, activating miRNAs provided a simple method for fast, tuneable, and potentially multiplexable control of genes. In this work, we show that these systems can be used in synthetic transcriptional units, instead of the native genomic activation often documented in literature. The complementary sites to the RNA of interest were positioned within the proximal and core promoter regions. While there was some activation of target

promoters, further studies are required to characterize the optimal positions and sequence requirements for activating RNA for high dynamic range. We have shown that native promoters can be transferred into plasmids and trigger small increases in expression in the presence of complementary miRNA. Attempts to improve dynamic range did not consistently improve readout, indicating that further or other systems need to be tested to enhance the characteristics of this system. For applications requiring minor activation of GOIs, activating RNA may serve as a resource budget-friendly alternative to other synthetic TF proteins. For sensing applications of small RNAs, including miRNAs, activation circuits improve on the first sensor design described, as an increase in expression is required, which makes it easier to detect the repression of mRNA. This overcomes the significant challenges faced in disentangling miRNA effects from resource competition for gene expression resources and degradation machinery. Therefore, there is potential for the use of miRNAs for the activation of targets for sensing and activation of GOIs.

One key way in which miRNAs achieve broad regulation is through their short sequences with homology in many transcripts. This may be an issue for specificity and off-target effects must be further studied.

Further work into the optimal designs would unlock the potential of miRNA for transcriptional and post-transcriptional regulation of expression.

5 DISCUSSION

5.1 SUMMARY

In this report, we have described the assembly and development of new tools for the tuneable engineering of gene expression in mammalian cells. Our approaches leverage small RNAs and TFBSs which offer multiplexable, scalable, and cooperative properties. Other advantages include TFBSs independence from expression machinery and miRNAs limited size and processing which lead to potent native gene regulation effects without large resource footprint in cell systems. We developed modular assembly methods for simplified design stages and tested the capacity of engineered TFBS and miRNA strategies to direct gene expression. In this section, we will compare these methods to each other and other approaches in the literature, while also discussing their current limitations.

Looking ahead, we will explore potential advancements to characterise these tools further and improve the capacity for gene expression regulation. We also address the wider applicability of these tools beyond gene expression engineering, considering their potential contributions to other research areas.

5.2 GENE EXPRESSION ENGINEERING

There is a growing interest in precise and versatile tools capable of manipulating gene expression in various biological contexts, a trend supported by the increasing number of peer-reviewed papers in this area (**Figure 5.1**). Gene regulation trends have evolved from uncontrolled drastic manipulations of gene expression (on to off or vice versa) to more graded methods of gene expression control (Eisenhut et al., 2024). While existing methods are valuable, they have limitations (Cox, Platt & Zhang, 2015; Laufer & Singh, 2015) including scalability, resource dependency, multiplexity, and challenges related to implementation in cells. These limitations highlight the timely need for new methods of gene expression modulation with varied characteristics.

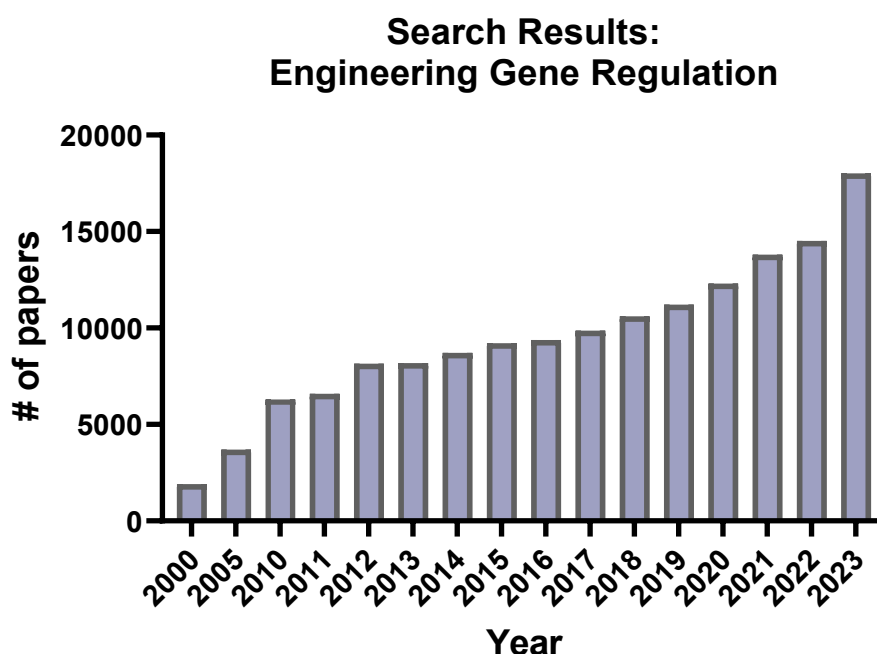


Figure 5.1: *There is an increasing interest in tools for engineering gene regulation as seen by the increase in the number of peer-reviewed papers published between the years 2000 and 2023. Graph generated using search results from Google Scholar with the prompt “Engineering gene regulation”.*

The development of new methods should aim to overcome these limitations by leveraging engineering biology principles thereby enhancing the reproducibility and implementation of these module types. MiRNAs and TFBSs emerge as promising starting points, given their favourable intrinsic properties that can be manipulated for various purposes. By TFBS regulation

of free TF levels and miRNA control over mRNA levels, broad regulation of native or orthogonal signals is achieved by cooperative, multiplexable characteristics in cells.

Some of the assessment criteria we assess our developed tools by are cellular resources required for activity, function across relevant mammalian cells, simplicity at the design stage and tuneability with high potential regulatory impact. In this research, we have made significant findings in the field of gene expression modulation, through the use of TF decoy plasmids and miRNAs.

5.2.1 TF DECOY PLASMIDS FOR TRANSCRIPTIONAL CONTROL

5.2.1.1 *Function:*

TF Decoy plasmids also known as DNA sponges can regulate gene expression of native and artificial TFs in mammalian cells by increasing TFBS concentration which sequesters TFs away from active sites of gene expression. TF decoy plasmids offer a faster approach to achieving control of gene expression compared to systems which require the expression of components. This regulation occurs rapidly because decoys do not require expression or processing for action. Instead, as soon as the DNA is added to the cell and reaches the nucleus, it is capable of binding TF targets. The addition of these tools transiently reduces issues related to the safety of permanent gene editing and limits exposure, where prolonged exposure may lead to increased mutation of these highly repetitive regions, which are prone to recombination. Further, these plasmids can be easily adapted to various targets by simply changing the sequence. Relatively limited knowledge from readily available databases such as JASPAR (Castro-Mondragon et al., 2022) is sufficient with no additional information about TF protein structure or interactions required for initial implementation. Further optimizations can be achieved to enhance TF decoy features, making them a versatile tool for gene expression modulation. In our work, we show that these tools can be used for a diversity of circuit types, for native and artificial TFs and also in human and hamster cell lines further highlighting their potential.

5.2.1.2 *Tuneability:*

The degree of regulation can be simply controlled using TF decoy plasmids, an important feature of engineered systems as cells rarely regulate expression in a binary way.

The tunability of TF decoy plasmids is demonstrated by their ability to control gene expression by increasing TFBS concentration in series (longer arrays of TFBSs per plasmid) or in parallel (larger intracellular copy number), offering a wide range of regulatory potential. The optimal dosage of TFBSs may differ depending on cellular context such as the abundance of TFs and the number of open genes containing TFBSs of interest. Consequently, achieving the desired level of tuning may necessitate testing of various numbers of TFBS repeats within TF decoys which is simplified by our use of CTRL. For example, our use of TetO decoys of increasing lengths reduced expression from pTet promoters yielding a maximum reduction of 81% in CHO-K1 or 92% in HEK-293T.

5.2.1.3 Resource Costs:

Other benefits to the implementation of TF decoys arise from their independence from gene expression resources and machinery. This is because these tools are provided into cells in their final form and do not require any further processing or expression of components. This independence from gene expression machinery offers several advantages. By providing TF decoy plasmids in their final active form, cells do not need to expend energy to transcribe and translate the TF decoys, not contributing to gene expression burden on cells. This could potentially lead to improved reliability of these systems due to less stress affecting cell health and viability, particularly important in experiments assessing cellular response. This is confirmed in our experiments by assessing the number of cell events in samples run on flow cytometry analysers and the lack of cell death in the conditions tested by Live/Dead cytotoxicity staining or counting throughout experiments where the cell death was not specifically being triggered. A more thorough assessment of cell stress can be used to confirm this further, for example by use of qPCR.

5.2.1.4 Comparison to other TFBS methods:

Methods which alter intracellular TFBS concentrations have been described before in mammalian cells. These methods include the use of oligonucleotides or unwinding epigenetically silenced genomic regions containing TFBS repeats. TF decoys in the form of oligonucleotides have been introduced into cells and shown to successfully compete for TF binding with active sites (Morishita et al., 1998; Crinelli et al., 2004). This diverges from our findings when we introduced oligonucleotides against a native TF of interest directly comparing this format with TFBS repeat plasmids for best responses. As we did not have access to the

chemical modifications some have recommended to prolong the lifetime for oligonucleotides in cells, the annealed oligonucleotides tested are likely highly degraded intracellularly. Moreover, mammalian TFs prefer binding multiple low affinity sites in DNA resulting in compounding effects. This is a feature captured in decoy plasmids by adding multiple repeats in plasmids but not oligonucleotide predecessors. To test this theory, TF decoys against a wider variety of TFs including ones previously established susceptible to oligonucleotide approaches should be compared. Another method used to reduce the expression of genomic GOIs has been the epigenetic regulation of TFBS repeats in the genome, displaying the potential of engineered versions which can be integrated and also controlled epigenetically (Janssen et al., 2000; Liu et al., 2007).

In summary, the implementation of TF decoy plasmids provides a rapid, safe, versatile, and resource-independent method for precise control of gene expression, demonstrating significant advantages over other TFBS modulation techniques.

5.2.2 MIRNAS FOR TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL CONTROL:

5.2.2.1 *Function:*

sRNA such as miRNAs through binding of complementary sequences control target gene expression dependent on the relative concentrations of miRNA to target transcripts present in cells. MiRNA function in the cytoplasm includes its binding to mRNA transcripts which influences mRNA stability and rate of translation. This kind of regulation levels out drastic fluctuations in transcription for stable protein expression (O'Brien et al., 2018). On the other hand, miRNA functions in the nucleus appear to induce the transcription of targets. The various effects of miRNAs in cells highlight their importance in broad cell regulation. Through their binding on nuclear DNA, miRNAs can affect the transcription rate by attracting transcription machinery or changing the epigenetic state of the DNA target (Fazi & Nervi, 2008).

Our research has demonstrated that miRNAs can be used to regulate gene transcription or post-transcriptional processing of mRNAs by positioning MREs within promoters or 3' UTRs respectively. Activation of genes containing promoter MREs is generally low and not significantly improved in synthetic promoters with multiplicity upstream compared to native versions with one

MRE. Positioning of MREs within the 3' UTR of mRNA sequences, did not always lead to repression of targets as expected and resulted in low fold changes up to 5 fold difference than controls. Therefore, while miRNAs have the potential to regulate gene expression, the extent of this regulation is not always drastic and can vary depending on the context and specific conditions.

5.2.2.2 Tuneability:

In literature, miRNA concentration with respect to the concentration of target determines the level of activity on its targets. In our study, we varied concentrations of miRNA by increasing the strength of the promoter expressing miRNA from the intron of a reporter, so that reporter expression can act as a proxy for miRNA expression.

Dynamic control of mRNA levels post-transcriptionally, in the process often referred to as RNAi facilitates quicker response to extrinsic signals by binding readily available transcripts. The concentration-based regulatory potential of miRNAs is a key area for further exploration, offering a wide range of applications for gene expression modulation. Unfortunately, miRNA responses were complicated to predict and did not provide an easy global readout, even after further processing of results and careful comparison to controls. This may be due to many confounding causes such as: Resource competition between miRNA expressing cassettes and sensors (Di Blasi et al., 2023), backlog of miRNA targeted transcripts at degradation machinery (Cella et al., 2023) and miRNA induced repression of mRNAs (Mukherji et al., 2011).

On the other hand, transcriptional control is slower requiring both transcription and translation or protein degradation to observe effects. Control of transcription using miRNA did generally lead to an increase in transcription of reporters, although the fold change was a maximum of 5.3-fold when pRelA was activated by rno-miR324 in HEK-293T. The smallest fold change was achieved when pIL2 was induced with hsa-Let-7i (the MRE is within the TATA box of the promoter) for a 1.7-fold change compared to the miRNA's absence.

Further attempts to improve activation at these sites through co-expression with MCP fused to transcriptional activator VPR did not consistently improve expression from miRNA binding sites. This design relies on protein expression increasing the resources required, making this solution less than ideal.

The study investigates miRNA concentration's impact on gene expression tuneability, revealing challenges in predicting miRNA responses on post-transcriptional regulation of targets while transcriptional control via miRNA showed a limited range of activation.

5.2.2.3 Resource Costs:

Endogenous expression of miRNAs in cells and those expressed from synthetic constructs both have resource costs, which rely on the expenditure of energy and cellular machinery, but these costs can differ significantly. Our studies of the resource requirements of the miRNA expression cassettes show that indeed resources are circumvented to miRNA expression at the expense of sensor expression. In order to prevent misanalysing our results we accounted for this with a control expressing miRNA with no affinity for the sensor. In this way, any decrease observed from the co-expression of the orthogonal miRNA is attributed to resource competition. The addition of engineered constructs that code for the overexpression of genes can lead to stress induced unpredictability in the system (Bolognesi & Lehner, 2018; Hetz & Papa, 2018), which may be observed by a lack of repression on mRNA targets. We also briefly explored the use of miRNA mimics, which can be directly transfected into cells and do not require transcription. This alternative is resource aware but compromises the visibility of miRNA load in cells.

5.2.2.4 Comparison to other methods:

MiRNAs are common features in engineered circuits with many previously described sensors which detect a change in output upon the addition of complementary miRNAs. Moreover, miRNA mimics and siRNAs are commonly used in many cell biology studies (Wang, Zhou & Richards, 2021).

MiRNAs are often also used in engineered circuits showing a variety of different responses such as increased expression due to degradation backlog of miRNA targeted transcripts (Cella et al., 2023), methods of mitigating against burden (Lillacci, Benenson & Khammash, 2018), or classifier circuits (Sayeg et al., 2015) al., 2011).

Our work differs from others in a few ways. Our work assesses the effects of different promoters for the expression of miRNAs on sensors. Moreover, in studies where the sensing of miRNA is required, resource competition effects are often not identified or discussed. Despite these differences, our dataset also displays the need to vary circuit characteristics in order to respond to different miRNA levels (Gam, Babb & Weiss, 2018; Gam et al., 2019). This is a significant

limitation of this post-transcriptional approach. The ideal design would respond to a larger variety of miRNA concentration inputs with a corresponding change in gene expression.

5.2.3 COMPARISON BETWEEN GENE EXPRESSION CONTROL METHODS

Our next objective is to understand how the gene expression regulators we developed fit into the landscape of readily available transcriptional modulators currently available to the scientific community (**Figure 5.2**). We compared gene expression modulators we tested with published methods such as overexpression or CRISPR editing. The criteria for comparison included the simplicity of design, the size of the machinery involved, the capacity for broad regulation, and the predicted immunogenicity and toxicity.

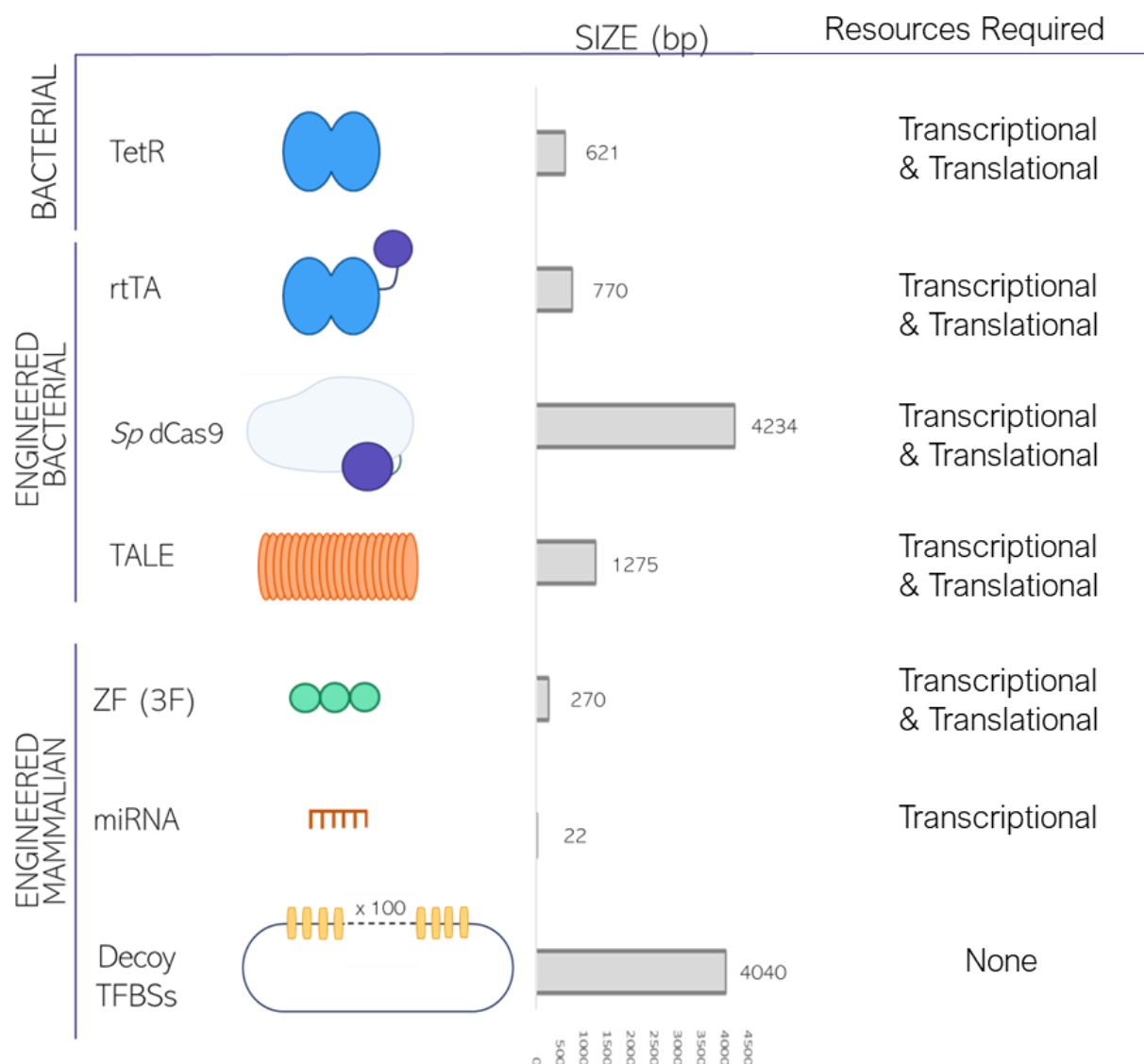


Figure 5.2: Origins, size and schematic representation of transcriptional regulators described or tested in this thesis (right). The methods of transcriptional regulation either bind to or compete with binding sites within regulatory elements of GOI expression cassettes (left). Adapted from a similar figure in Bhatt, García-Díaz & Foight, 2023.

The initial stage of any gene expression modulator implementation involves the design and assembly of regulatory tools. This process must be streamlined to speed up reaching the stage for testing the ability to control gene expression.

Our approach simplifies this stage by facilitating the modular assembly of tools. For miRNA tools, this includes adding short complementary sites to the miRNA of interest into defined regions of the expression cassette. For TF decoys, this entails leveraging publicly available biological context information flanked by defined spacers for domestication into CTRL iterative assembly workflow (details in Chapter 2). For TFs such as TetR with specific TFBS recognition, engineering requires careful addition of a pTet promoter upstream of GOIs, which is more

complicated for endogenous genes. A more adaptable TF (likely at the expense of specificity) is dCas9, whose design includes the assembly of complementary guide sequences into available backbones for sgRNA expression. Many readily available tools are accessible to help predict capacity for engineering using guide sequences (ex: Synthego).

While all these methods require some degree of cloning to achieve optimal engineering, the amount of cloning varies. For instance, creating longer TF decoys involves many iterative cycles, but it also generates smaller decoys for testing along the way. For dCas9 methods, several guides may need to be designed and cloned, which balances out the time difference between developing these tools for testing. So, all transcriptional engineering strategies require some cloning, the amount of which depends on the number of iterations required for initial testing and optimisation for context.

Secondly, the size of tools (**Figure 5.2**) and the multiplicity required to be delivered may affect the transfection method used and the eventual success of engineering in mammalian cells (Fus-Kujawa et al., 2021). For example, methods that rely on expression within cells can transfect a small amount of DNA. This DNA can then be transcribed multiple times, thereby amplifying the protein signals derived from a DNA sequence to levels that are physiologically detectable. TF decoys, on the other hand, must be introduced at the multiplicity required to achieve required responses as no further processing or signal amplification occurs upon introduction into cells. This can be overcome in some cases with the use of episomal vectors or viral origins of replication in cell lines containing viral replication machinery which facilitate replication and maintenance of plasmids in mammalian cells (Roig-Merino et al., 2022). DNA compaction must also be considered for TF decoys, which may affect sites available for TF binding (Shen & Hu, 1986). While in the short timescale of our experiments, we saw potent regulatory effects without variation in TF decoy response across biological replicates, a more detailed study of the effects of coiling of plasmids in experiments can be assessed. Therefore, it is important to consider the balance between the amplification of modulators and the resource demand they inflict on cells. In modulators that do not require expression, DNA coiling may improve the transfection of large DNA constructs, but ensuring this does not interfere with function is important.

Thirdly, upon introduction into cells, the capacity for broad regulation on the target of transcriptional modifiers should be considered. The engineered TFs in **Figure 5.2**, have some level of off-target responses and broad regulation often involves the engineering of more components per target complicating further the previous two steps of assembly and transfection (Waryah et al., 2018). For example, multiplexed gene engineering using dCas9 is often

described by expression multiple guides (Jiang et al., 2023). On the other hand, both miRNAs and TFBSs, described in our research, exert broad regulation on cells without the requirement of further engineering. So, if specific regulation of a site is required TFs can be considered, whereas if broad but specific regulation is required TFBSs and miRNAs might be the easier to engineer alternative.

When considering cell engineering, especially for therapeutic applications, an important factor to take into account is immunogenicity. Immunogenicity refers to the ability to provoke an immune response, impacting the effectiveness of engineered components or provoking significant adverse reactions leading to issues with safety. Tet based TFs and CRISPR based TFs have been associated with immunogenicity likely due to their microbial origins (Bhatt, García-Díaz & Foight, 2023). This association has led to a reduction in their use across various applications. While immunogenicity may not pose a problem for certain applications, it is a critical factor to consider before employing these TFs within gene or cell therapies. This may be another benefit to implementing TF decoys or miRNAs which are based on natural versions in mammalian cells instead of non-mammalian engineered TFs (**Figure 5.2**).

Lastly, both strategies we describe are amendable to and have been tested on natural or synthetic TFs or miRNAs for proof of wide applicability. Our research has shown that TFBS plasmids specifically and consistently regulated gene expression as anticipated without the overexpression of genes, an important consideration of gene expression controllers. Moreover, decoy plasmids only require one component which can be scaled by increasing the repeat number, whereas additional parts may be required to increase the dynamic range from miRNA activated promoters.

In summary, the key findings of this research demonstrate how TF decoy plasmids and miRNAs complement available tools for versatile gene expression modulation, offering unique features and tuneable methods for controlling gene expression at the transcriptional and post-transcriptional levels. Between the two approaches tested, TFBS repeat plasmids seem to be the most promising avenue.

Further comparisons to how these methods stack up with other published gene expression engineering methods for specific purposes such as the bioproduction of recombinant proteins would further improve understanding of these systems and whether the hypothesized benefits lead to improved results.

5.3 CURRENT LIMITATIONS

While our study has identified novel routes for modulating mammalian gene expression and performed studies to outline their development and few uses, it is crucial to acknowledge and address certain limitations that may impact their implementation. These limitations, range from the lack of switch-off our methods exhibit or the potential off-target effects and interference. For example, in miRNA experiments, interference from cellular processes may have contributed to issues with the interpretation of results and their applicability in various contexts. In this section, we will delve into these limitations in some detail, discussing their implications and suggesting potential avenues for further research and improvement.

One limitation that we considered is related to the inducibility of our gene expression modulators. The gene expression modulators explored in this study are "always ON", whereas the ability to turn on or turn off the regulation is of benefit to many engineering biology applications. The lack of inducibility may limit the generalizability of use with some applications. This is mostly an issue in stably integrated versions as in transient versions plasmid number is halved with each cell cycle until they are eventually lost from the mammalian cells. This could impact the interpretation of the results, where a real-time TFBS copy number is required. We initiated a collaboration to address this need to control regulators using the integration of TF decoys into the iPSC genome at safe harbour loci and control access via epigenetic modifiers.

Another limitation, which we have not explored in our work is these tool's potential off-target effects which could lead to the unintentional regulation of other non-target genes. Implementation of these gene expression modulators without careful consideration before use could therefore raise major safety concerns. For example, a 10 bp binding site of a TF could occur naturally several times within the 3 billion size genome of humans. Models describing how TFs in cells search for and identify recognition sites have been developed, generally pointing to the benefits of having multiple lower affinity highly cooperative sites (Normanno et al., 2015; Bragdon et al., 2023), similarly low affinity highly cooperative TF decoy is therefore conceivable. We mitigated against potential off-target effects in our study of native TF sequestration by using native binding site sequences, likely with low affinity to the TF. Further, specific testing would be useful to understand the potential effects of these decoys on off-target genomic locations. The right balance could be essential for certain outcomes, like when synthetic TFBS repeats with

stronger affinity for TF compared to native active sites was added to yeast cells, bimodal populations were obtained (Lee & Maheshri, 2012).

This is similar to miRNAs which are known for their broad regulation in cells. When delving into the biological context, the background expression of the miRNA of interest and the expression of any competing endogenous RNAs can potentially hinder our understanding of intracellular miRNA levels. These endogenous RNAs may neutralize miRNA by possessing enough complementary sequences to bind and deactivate them. However, the degree of competition among endogenous RNAs under physiological conditions continues to be a subject of active research and discussion (Bosson, Zamudio & Sharp, 2014). By understanding the biological context of possible miRNA targets in cells or using mixes of native and synthetic miRNA designs some applications can mitigate against some of these potential issues.

An additional constraint in this investigation is the inadequate knowledge of the cellular processes implicated in transcriptional activation mediated by miRNAs. In our research, we try to pinpoint the best positioning for the highest activation. In this aim, we varied multiplicity upstream a promoter, as well as miRNA binding site positioning at the TATA site, or 41 bp upstream start codon which naturally causes upregulation of the IL2 promoter in human cells or activation of the natural RelA promoter of rats respectively. In the conditions tested a maximum of 5.3-fold change was achieved, and amplification of this signal was not achieved. Understanding further whether specific constraints are preventing further amplification of the system may be important to further developing miRNA activation of promoters. This system may still be used in the regulation of specific situations where there is a positive feedback loop which can encourage further activation from the modest initial activation achieved by the activating miRNAs (**Figure 5.3**).

Lastly, one of the largest shortcomings encountered in this study was the interference of the miRNA expression cassette with sensor expression reducing the capacity to infer conclusions of miRNA effects on sensors. The experimental set-up of the miRNA repression sensor in the presence of a miRNA expression plasmid interfered with potential RNAi effects making it unclear what was caused by resource competition and what was response to miRNA concentration. The study's miRNA repression sensor setup may therefore not clearly demonstrate the effects of RNAi, particularly in cases where the visibility of miRNA amount causes a burden. This limitation could impact the interpretation of the results, especially in the absence of good controls.

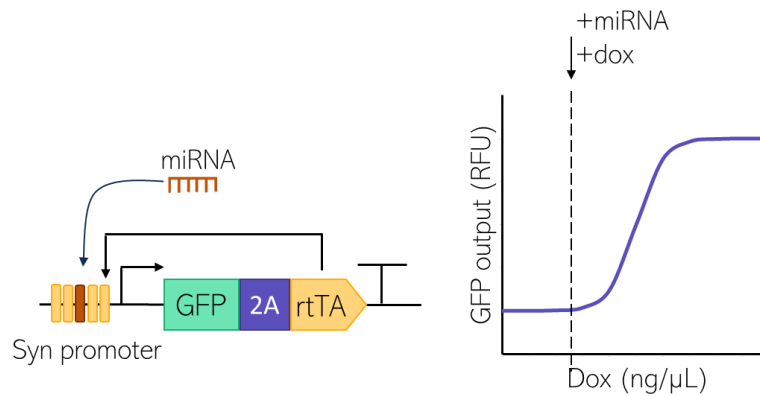


Figure 5.3: *Depiction and expected response curve of a potential method of increasing signal from miRNA activation circuits using PFLs. In the absence of doxycycline and miRNA, there is no expression from the cassette. Upon addition of complementary miRNA and doxycycline low expression of GFP and rtTA (expressed from the same transcriptional unit by a self-cleaving peptide such as P2A) is achieved. The rtTA expression can then further activate expression from the circuit.*

Despite the limitations discussed, it is important to note that many of these may not pose significant issues in some practical implementations of our gene expression modulators. The concerns raised are largely based on our current understanding of the system. With further research and experimentation, we can not only validate these assumptions but also develop strategies to mitigate any potential impact. Therefore, while we acknowledge these limitations, we remain optimistic about the potential of our methods. We believe that with careful design and testing, we can overcome these challenges and unlock the full potential of our gene expression regulators for a wide range of applications.

5.4 FUTURE DIRECTIONS

Building upon the current limitations identified in this research, there are several promising avenues for future exploration. These potential directions not only address the existing gaps but also aim to explore the scope of potential uses of the developed methods. The following suggestions for future research are subdivided into methods for understanding and improving the current tools, further applications extending from those we tried in gene expression engineering and beyond.

5.4.1 IMPROVING CURRENT METHODOLOGIES AND TOOLS

The first area of focus for future research we discuss is the enhancement of the methodologies and tools currently in use, to address the identified limitations and improve their efficiency and effectiveness.

Our findings, while promising, give rise to many avenues of research which could provide improvements, and a deeper understanding of the functionality, and limitations of these tools. One key area of improvements to gene expression from RNA activation systems is the development of amplifiers that enhance circuit properties such as dynamic range and detection range without severe impacts on resource footprint. We suspect positive feedback loops might be a promising place to start. This is through the PFL characteristics where small signals can be amplified to yield ultrasensitive responses. In this case, the miRNA initial activation of a promoter would induce the expression of a TF and reporter, the TF would further amplify the expression of the PFL (**Figure 5.3**). Further understanding limitations reducing dynamic range using miRNAs may further assist us in future engineering efforts of these systems, this is likely not due to multiplicity or positioning within promoters as these aspects have been assessed.

Concerning the work on miRNA post-transcriptional repression it may be our current workflow does not show sensitivity in the miRNA ranges tested. Future engineering can take this into account by using a similar sensor layout in cell lines known to express a miRNA of interest this will show if transfected cassettes were overexpressed in our system. In our optimisation experiments, we attempted low concentrations of expression plasmids in transfections, but this led to low resolution detection and high variability between experimental replicates. We also attempted to visualise the sensitive window by binning miRNA expression.

TF decoy plasmids or DNA sponges can also benefit from additional experiments. Firstly, a comparison between TF decoy plasmids and TFBS double-stranded oligonucleotides for a range of TFs with varying affinity cooperativity profiles would be useful. Showing under which conditions using TF decoy is preferable to duplexed oligonucleotides.

In our native TF application, we tuned the oxidation response using TF NRF2, a basic leucine zipper with a Cap “n” Collar structure. This structure facilitates heterodimerisation of the TF and DNA recognition with higher affinity (Sengoku et al., 2022). Further expanding our study to other native TF groups, with different characteristics can further prove broad utility. We are collaborating with teams to achieve this goal. Additional coupling of results with other protocols such as coprecipitation, as used in (Lee & Maheshri, 2012) can assist in systematically characterising system preferences and can help guide future designs. Moreover, our research on the impact of synthetic transcription factor (TF) sequestration within self-regulating feedback loops revealed the influence of decoys across various circuits. Further assessing whether the same trends are observed in natural feedback loops which are common in cells, for example in the differentiation of muscle cells (Vicente-García, Hernández-Camacho & Carvajal, 2022), could be informative of how cells regulate the expression of native TFs to encourage or discourage differentiation for example.

Longer-term goals of being able to directly compare TFs and their characteristics based on curve features can also improve the useability of this tool. TF binding dynamics are actively being studied across many groups. Investigating TF binding dynamics typically explores binding affinities to target sequences, where many databases log the probability of a nucleotide at each position of the TFBS in position weight matrices (PWMs). But TF binding probability is not necessarily the biggest determining factor in expression and instead exploring the cooperative effects of adjacent TFBSs when multiple TFs are present, is more important in many systems (Bragdon et al., 2023). Contributing a tool to understand a TFs preference for affinity, cooperativity or specificity will likely also require modelling to understand the principal components that lead to curve (response vs inducer) characteristics.

Other improvements that would expand the versatility and applicability of these tools would be engineering TF decoys to further be switched ON and OFF. Work is ongoing to integrate Tet decoys into human safe-harbour genomic loci and control ON/OFF switching using epigenetics. This collaborative work will also shed light on tool stability in mammalian cells as repetitive sites tend not to be too stable in cells (Paulson, 2018). Moreover, other than repetitive sequences themselves, long integrations several TUs long have been known to be silenced over time

(Zimak et al., 2021). This brings up questions regarding the natural silencing of longer stretches of TF decoys and whether this is affected if a selective advantage to keeping the TF decoy unsilenced is used.

Lastly, it would be interesting to explore the extent to which coiling and supercoiling of TF decoy plasmids affect responses in cells. While DNA coiling improves the transfection efficiency of large DNA constructs, ensuring this does not interfere with TF decoy capacity for regulation is important. Papers suggest that in cells plasmids find an equilibrium between coiled and uncoiled conformations (Shen & Hu, 1986), this often relates to plasmids which express recombinant proteins, so equilibrium might at least partially be dictated by binding of cellular machinery for expression functions, so checking the degree of accessible plasmid may further help decide the amount of plasmid required. A similar approach as utilised in this paper, i.e. isolating transfected TF decoy DNA from mammalian cells and observing fraction supercoiled may be an initial starting point.

Artificial Intelligence (AI) and Machine Learning (ML) are revolutionizing engineering biology by enhancing predictive design and optimizing engineered biosystems, driving progress in various fields. Specifically, the use of ML and AI in gene expression engineering is likely to extract and highlight information from patterns present in old and new data sets. As these capabilities advance, they are expected to contribute to transformative developments in engineering biology and can be used to predict and improve the functionality of gene expression modulators.

5.4.2 APPLICATIONS IN GENE EXPRESSION ENGINEERING

In this study, we primarily tested tools in applications for engineering gene expression regulation testing the effect of these tools on specific gene expression circuits and native regulation. This provides a foundation in which more gene expression engineering studies can be performed. For example, for cell and gene therapies, a graded response instead of a drastic change in gene expression is often required (Kitada et al., 2018). In these applications modulating the level of active tool in cells can provide the level of gene expression required. In the field of bioproduction, cell engineering is often performed to improve the yield of protein products per cell. Assessment of which TFs reduce the output of functional proteins can be considered as targets for engineering, potentially removing cellular blocks for expressing processing machinery or delaying cell death.

Furthermore, these methods can be applied to enhance our understanding of fundamental biological processes. For example, the use of TF decoy plasmids in studying TF binding dynamics and quantitative biology efforts, as well as the exploration of the bottlenecks in degradation machinery caused by miRNAs, could contribute to significant advancements in our knowledge of gene regulation and cellular processes. Already, TF decoys have been used to study TF regulation from timespan to interaction strength and free TF concentration in bacterial and yeast strains (Lee & Maheshri, 2012; Bokes & Singh, 2015; Ali & Brewster, 2022). An interesting future direction would be assessing whether similar rules govern gene regulation in mammalian cells or uncover mammalian specific attributes.

Ultimately molecular tools that can regulate more than one regulatory step at a time, will provide specific benefits to gene circuits, with more multi-layer circuits being designed and tested (Bartoli et al., 2020; Greco et al., 2021b; Gao, Wang & Wang, 2023; Eisenhut et al., 2024). Integration of both TF decoy tools, as well as miRNA gene regulation into gene circuits that implement responses to both miRNAs and TFs, may achieve improved features and regulation of the output. This integrated approach holds promise for the development of more sophisticated and precise gene expression modulation systems with enhanced functionality and regulatory control.

In conclusion, this work shows a variety of implementations in the regulation of gene circuits or native expression aided by graded responses to varying tool concentrations.

5.4.3 APPLICATIONS BEYOND GENE EXPRESSION ENGINEERING

The research avenues offer potential applications to a wide range of real-world problems and challenges, extending beyond gene expression engineering.

One promising direction is the use of the developed gene expression modulators as disease models. For instance, the TF decoys could be utilized to create disease models for studying repeat expansion diseases (Paulson, 2018; Depienne & Mandel, 2021), providing valuable insights into the underlying mechanisms of these conditions. Additionally, the dysregulation of miRNAs is also linked to various diseases such as cancer, autoimmune disorders, and neurodegenerative conditions (Qu et al., 2015; Huang et al., 2012), presenting an opportunity to leverage the developed gene expression modulators for disease modelling by mimicking dysregulation of miRNAs observed in various diseases.

Understanding biological processes outside of gene regulation can also be considered. An interesting application of this may be using synthetic TF decoy plasmids to generate synthetic nucleoli (Grob, Colleran & McStay, 2014), with interesting properties such as increased density whose functions are still being studied (Kim & Guck, 2020).

Lastly, it has previously been discussed that the properties of TF decoy sites through their binding to TFs can change response times in cells. Response time is an important regulatory feature of many cell operations which rely on specific timings such as circadian rhythms or cell division (Zhang, Dey & Singh, 2022). Testing whether TF decoy sites to specific targets can affect response times may be an interesting way of studying these cell functions.

The findings of this thesis project present some of the numerous potential future research directions for using TF decoys and miRNA tools for advancing gene expression engineering and beyond.

5.5 CONCLUSIONS

New methods of gene engineering have the potential to revolutionize many biotechnology processes, including insights into native gene expression, function and regulation. Genetically modified organisms have a growing role in our daily lives although the need for careful implementation is very important.

The insights and methods developed in this thesis leverage the varied nature of genetic processing in cells for robust control of gene regulation that can be manipulated for new investigations. In this direction, we identify two new potentially successful transcriptional regulators of mammalian cells. The work pays attention to the effects of resource footprint on experimental setups and takes into account stakeholder opinions to develop new artificial, transient methods for gene expression regulation and modulation in mammalian cells.

The study focuses on the emerging field of gene expression manipulation, highlighting the shift towards more controlled methods. It explores the use of TF decoy plasmids, also known as DNA sponges, and miRNAs as promising tools for gene expression control. These tools offer a rapid, safe, and versatile method for precise gene expression control, showing significant advantages over other techniques.

The role of miRNAs in gene regulation is examined, noting that their regulatory potential needs further exploration. The study discusses the resource costs of miRNA expression in cells and the effects of different promoters for miRNA expression on sensors. It underscores the challenges and potential of miRNA in gene regulation and highlights the complexities in gene expression modulation using post-transcriptional engineering strategies.

The research acknowledges limitations, including lack of inducibility, potential off-target effects, and inadequate knowledge of cellular processes. Despite these challenges, the study remains optimistic about the potential of these tools, with further research and improvement needed to address these limitations.

Future investigations to understand constraints to implementation of these tools, can improve applicability and make significant contributions to our understanding of gene expression, disease mechanisms, and the development of advanced therapeutic strategies.

In conclusion, this thesis outlines the development and uses of gene expression modulators, highlighting the need for further research and current limitations. Despite challenges, the practical implementation of gene expression modulators remains promising.

6 MATERIALS AND METHODS

6.1 CELL CULTURE

6.1.1 MAMMALIAN CELL CULTURE

6.1.1.1 Cell Viability & Count

Cells were collected and centrifuged at 1500 rpm for 5 minutes before removal of supernatant. Cells were then resuspended in suitable complete media making sure to separate large clumps of cells with adequate trypsinisation and resuspension. Once the cells have been successfully resuspended and any large clumps have been separated, a 50 μ L aliquot of the cell suspension was transferred to a new tube. To this tube, 2.5 μ L of Solution 18 (AO•DAPI Staining Reagent, ChemoMetec) was added. Following the addition of Solution 18, 10 microlitres of the resulting solution was added to well of an NC-Slide A8™. The NC-Slide A8™ is a specialized slide designed for use with the NucleoCounter® NC-250™ Automated Cell Analyzer (ChemoMetec). The NucleoCounter® NC-250™ Automated Cell Analyzer measures cellular events and percent cell viability and concentration of cells per mL.

6.1.1.2 Mycoplasma Testing

Our cultures did not use antibiotics to safeguard against bacterial contamination. This decision necessitated the use of meticulous handling procedures and strict aseptic techniques during all aspects of cell culture to prevent contamination. One less visible but common contaminant of mammalian cell cultures is mycoplasma, a type of bacteria that is small enough to pass through filters designed to remove bacteria. Mycoplasma contamination can have a variety of detrimental effects on cells, including altering their growth and metabolism, and can also interfere with experiments. To ensure that our cell cultures remained free of mycoplasma, we tested them biannually using the MycoAlert Mycoplasma Detection Kit from Lonza (product number LT07-318). Briefly, to spent cell culture medium MycoAlert Reagent was added and luminescence measured. MycoAlert Substrate was then added, and luminescence measured

again. The ratio of these measurements (After/Before) indicates the presence or absence of mycoplasma such that a ratio of around 1 is not contaminated with mycoplasma.

6.1.1.3 HEK-293T

HEK-293T (ATCC-CRL-3216), or Lineariser integrated HEK-293T (gift from Prof. Balázs Lab, Stony Brook University), were cultured in DMEM Glutamax (Invitrogen), supplemented with 10% foetal bovine serum (FBS) containing necessary growth factors. The cells were sub-cultured when they reached 80% confluence on tissue culture plastic, or when the cell count reached 2×10^6 cells/mL, as determined by counting. This was in accordance with the protocol provided by ATCC, ensuring optimal growth and health of the cells. Cells were kept at 37 °C and 5% CO₂.

6.1.1.4 CHO-K1

CHO-K1 (gift from Prof. Kontoravdi Lab, Department of Chemical Engineering Imperial College London) were cultured in MEM α (Sigma Aldrich) supplemented with 10% FBS, 1% non-essential amino acids and 1% L-glutamine. Cells were cultured at 37 °C and 5% CO₂. Cells were sub-cultured according to ATCC protocol and kept in culture under 30 passages.

6.1.1.5 PC-12

PC-12 cells (gift from Prof. Green Lab, Department of Bioengineering Imperial College London) were grown in suspension in RPMI ATCC blend supplemented with 5% fetal bovine serum and 10% horse serum. Cells were sub-cultured according to ATCC protocol and kept for under 10 passages. Cells are cultured at 37 °C and 5% CO₂.

6.1.2 E. COLI DH5A

DH5 α (F[−] ϕ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17(rK[−], mK⁺) phoA supE44 λ -thi-1 gyrA96 relA1) was cultured in LB (Luria Broth) or on LB agar and grown in shaking or static incubators respectively at 37°C overnight. Antibiotics were used for selection such that the final concentration of Kanamycin was 50 μ g/mL and the final concentration of Ampicillin was 100 μ g/mL. To prepare competent cells for efficient chemical transformations, bacterial cultures (OD 0.3-0.4) are harvested and resuspended in a cold solution containing 10% PEG 8000, 30mM MgCl₂, 5% DMSO before aliquoting and storing at -80°C. *E. coli* were transformed by adding DNA on ice prior to heat shock at 42°C and recovery in LB.

6.1.3 *S. CEREVISIAE* BY4741 AND VL6-48N

BY4741 (MAT α his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0) and VL6-48N (MAT α , his3- Δ 200, trp1- Δ 1, ura3- Δ 1, lys2, ade2-101, met14) were the *S. cerevisiae* yeast strains used in this work. These strains were cultured in a Synthetic Defined Medium, which is a type of growth medium that contains all the nutrients required for yeast growth. The cultures were incubated at a temperature of 30°C for a period of 2 days.

For the auxotrophic strains, which are unable to synthesize certain essential nutrients, supplements were added to the medium. These included various amino acids and nucleotides, such as uracil, which were added as needed to support growth. To prepare the *S. cerevisiae* strains for transformation, a process that allows them to take up foreign DNA, they were made competent using the Frozen-EZ Yeast Transformation II kit (product number T2001, manufactured by Zymo Research). After preparation, the competent cells were stored at a temperature of -80°C in aliquots of 100 μ L each until they were needed for use. This storage method helped maintain the viability and competency of the cells, allowing for efficient transformation when needed in later used in experiments.

6.2 MOLECULAR BIOLOGY (CLONING)

Parts not generated by CTRL (described in Chapter 2) were constructed using the combinations of the following cloning techniques, described below.

6.2.1 DNA EXTRACTION

DNA was assessed for purity and concentration in nuclease-free water using Biodrop (Biochrom Ltd.) prior to use elsewhere. Purity was assessed by measuring absorbance ratios A260/A280 and A260/A230 ratio around 2.0. A260/A280 values outside the acceptable range may indicate protein contaminations. A260/A230 values outside the acceptable range indicate sample contamination with salts. Contaminants of DNA preparation were discouraged as they interfere with downstream reactions.

6.2.1.1 DNA extraction from bacterial cultures:

Overnight bacterial cultures containing plasmids of interest were used in DNA extraction using QIAprep Spin miniprep for around 500 ng/μL yield from 5 mL of culture or midiprep for 1.5 μg/μL yield from 50 mL cultures. Culture was spun down to discard spent media and cell pellets were resuspended with resuspension buffer containing RNase A for RNA degradation. Cells were then lysed using a lysis solution and the reaction neutralised to prevent extended denaturation of DNA of interest. The reactions were then spun down at a high speed to separate the cell debris which accumulates into a pellet containing proteins and genomic DNA and plasmid DNA containing supernatant. The supernatant was then applied to a QIAprep spin column containing a silica membrane for selective binding to plasmid DNA. The rest of the solution was eluted through the column, and DNA washed to get rid of any further contaminants. The membrane was then dried, and the plasmid DNA eluted in nuclease-free water.

6.2.1.2 DNA extraction of mammalian genomic DNA:

The isolation of genomic DNA from mammalian cells was performed using the NucleoSpin Tissue kit (Macherey-Nagel), following the manufacturer's recommended protocol and utilizing the provided buffers. The process begins with the collection of approximately 10^7 cultured mammalian cells, which were then centrifuged at 1500 rpm for a duration of 5 minutes. The supernatant containing the spent medium was discarded, leaving behind a pellet that was resuspended in a lysis buffer. Buffer contain Proteinase K and SDS, which aid in the breakdown of proteins and the disruption of cell membranes, respectively. Following resuspension, the solution was incubated at a temperature of 70 °C for a period of 15 minutes. After this incubation period, 210 μL of absolute ethanol was added to the solution and thoroughly mixed. This step encouraged DNA precipitation. The resulting solution was then loaded onto a silica membrane column, provided as part of the NucleoSpin Tissue kit. Centrifugation at high speed was then used to pass the solution through the column, allowing DNA to bind to the silica membrane while the rest of the solution to flowthrough. Subsequently, the membrane was washed to remove any remaining impurities. After discarding the flowthrough from this wash step, the membrane was left to dry. This drying step ensured the removal of any residual ethanol, which could interfere with downstream applications. In the final step, the bound genomic DNA was eluted from the membrane using nuclease-free water.

6.2.1.3 DNA extraction from Agarose Gels:

DNA fragments ranging from 200 bp to 10 kb were separated using a 1% agarose gel, which was prepared by dissolving agarose in 1x TAE buffer. To enable DNA visualization, 1x SYBR Safe (Invitrogen) was incorporated into the molten agarose before it is poured into a casting tray to form a gel with wells of the desired size. Set agarose gels were loaded with DNA samples alongside a 1 kb Plus Ladder (NEB), which served as a reference for determining the size of the DNA fragments. The gel was then subjected to electrophoresis in 1X TAE buffer at 100 V for approximately one hour. Following electrophoresis, the agarose gel was imaged to visualize the separated DNA fragments. For the extraction of specific DNA fragments from the gel, the QIAQuick Gel Extraction Kit (QIAGEN) was used. The desired DNA fragment was carefully excised from the agarose gel using a clean scalpel, ensuring minimal excess gel is included. The gel slice was then mixed with QG resuspension buffer and incubated at 65°C for 10-20 minutes with intermittent vortexing to promote dissolution of the agarose and release of the DNA. Isopropanol was added to the solution to facilitate the binding of the DNA to the silica membrane of a spin column. The solution was then applied to the column and centrifuged, allowing the DNA to bind to the membrane while the flowthrough was discarded. The column was subsequently washed to remove any remaining contaminants. After drying the column to remove residual ethanol, the purified DNA was eluted using nuclease-free water.

6.2.2 RESTRICTION DIGESTION

This technique was employed for two main purposes - to validate the success of DNA assemblies and to fragment longer DNA molecules, thereby enabling the isolation of specific DNA fragments of interest. The process involved the use of restriction enzymes sourced from NEB, which are proteins that can cut DNA at specific sequences. The reaction setup included at least 1 µg of the DNA of interest, up to 1 unit of enzyme 1x buffer and water for a total reaction volume of 50 µL. The reactions took place in 1x CutSmart buffer or an equivalent buffer (NEB), which provided the optimal conditions for the restriction enzymes to function. The reaction mixture was mixed and incubated at a temperature of 37°C. The use of restriction digestion allowed precise cuts in the DNA, generating predictable and reproducible fragments. The success of the digestion was often confirmed by assessment of fragments using agarose gel electrophoresis. These fragments were compared to expected sizes alongside a ladder and used in downstream applications. Incomplete reactions due to inefficient incubation time or partially inactivated enzymes sometimes led to additional DNA bands at other sizes. This included supercoiled, nicked, or linear DNA.

6.2.3 PCR

Primers, synthesized by IDT, were designed to flank DNA and reconstituted in nuclease-free water to a working concentration of 10 μM . These primers were between 20 and 40 nucleotides long, had similar melting temperatures below 72 $^{\circ}\text{C}$, and a GC nucleotides made up between 40-60% of their sequence. The annealing temperatures of the primers were determined using an online tool <https://tmcalculator.neb.com/#!/main>.

PCR reactions were prepared using either the high-fidelity Phusion DNA polymerase (refer to **Table 6.1**) or the lower fidelity Taq DNA polymerase for colony PCRs, where a colony was diluted in 20 μL of nuclease-free water and used as the DNA template (refer to **Table 6.2**).

Table 6.1: *PCR Conditions set up using NEB Phusion polymerase.*

Component	50 μl Reaction	Final Concentration
Nuclease-free water	to 50 μl	
5X Phusion HF or GC Buffer	10 μl	1X
10 mM dNTPs	1 μl	200 μM
10 μM Forward Primer	2.5 μl	0.5 μM
10 μM Reverse Primer	2.5 μl	0.5 μM
Template DNA		< 250 ng
DMSO	1.5 μl	3%
Phusion DNA Polymerase	0.5 μl	1.0 units/50 μl PCR

Table 6.2: *PCR conditions for colony PCR using Taq polymerase.*

Component	25 μl Reaction	Final Concentration
Nuclease-free water	to 25 μl	
10X Standard Taq Reaction Buffer	2.5 μl	1X
10 mM dNTPs	0.5 μl	200 μM
10 μM Forward Primer	0.5 μl	0.2 μM
10 μM Reverse Primer	0.5 μl	0.2 μM
Template DNA	1 μL	variable
Taq DNA Polymerase	0.125 μl	0.625 units

Reactions are cycled with temperatures and timings, for 30 cycles:

Table 6.3: *Temperature Cycler Conditions for PCR. Temperature cycler set up for PCR reactions using Phusion DNA polymerase or Taq DNA polymerase.*

Step	Phusion Polymerase	Taq Polymerase
Initial Denaturation	98°C, 30 seconds	95°C, 30 seconds
30 Cycles	98°C, 10 seconds calculated °C, 30 seconds 72°C, 1 minute/kb	95°C, 30 seconds calculated °C, 30 seconds 68°C, 1 minute/kb
Final Extension	68°C, 5 minutes	68°C, 5 minutes
Hold	10°C	10°C

The success of PCR products was confirmed by running them on an agarose gel. A band should appear at the correct size relative to a DNA ladder. Following this, PCR products were treated with 1 unit of DpnI for 3 hours at 37°C to remove methylated DNA templates. This was done before cleanup with a PCR-clean up kit (QIAquick PCR purification kit, QIAGEN). In brief, a 5x PB solution was added to the PCR volume and applied to a silica membrane for the flow-through and cleanup of contaminants, followed by the drying of membranes and elution of the DNA of interest in nuclease-free water.

6.2.4 RCA

Rolling circle amplification reactions consisted of a combination of approximately 10µg/µL of short circular DNA template which was combined with 5 µL of TempliPhi™ sample buffer (Cytiva). This mixture was then subjected to an incubation period at a temperature of 95°C for a duration of 3 minutes. Following this initial incubation for denaturation, 5 µL of TempliPhi™ DNA Amplification Kit Reaction Buffer and 0.2 µL of enzyme mix were added. This addition was done on ice to maintain the integrity of the components. Subsequently, the mixture was incubated at a lower temperature of 30°C. This incubation period lasted for 5 hours. This allowed ample time for the reaction to generate branched amplifications of the original template. Finally, to deactivate the reaction, the mixture was incubated at a temperature of 65°C for 10 minutes. This step ensures that the reaction halted, and the resulting product was ready for further

analysis or use. The products were quantified and run on an agarose gel to prove the success of the reaction.

6.2.5 GOLDEN GATE REACTIONS

Thirteen fmol of DNA parts including the receiver vector were combined along with 10 units of Type IIS enzyme, 200 units of T4 DNA ligase and 2 μ L Ligase buffer topping up the reaction to reach a final volume of 20 μ L in nuclease-free water. The reaction was then placed in a thermocycler programmed to incubate the reaction mixture at:

37 °C for 5 mins, (37 °C for 5 mins, 16 °C for 10 mins) x25, 16 °C for 20 mins, 37 °C for 30 mins, 75 °C for 5 mins, 10 °C; hold.

The resulting reactions were transformed into *E. coli* DH5 α competent cells and plated on selective agar plates.

6.2.6 GIBSON ASSEMBLY

Where seamless cloning was required Gibson assembly was performed. DNA regions were amplified with overhangs of 20-30 bp of homology. Equimolar amounts of purified PCR fragments were combined with NEB enzymes (T5 5' exonuclease, Phusion DNA polymerase and Taq DNA ligase) and incubated at 50°C for 1-2 hours. Assembled products were transformed into DH5 α competent bacteria and plated onto selective agar plates.

6.3 DNA TRANSFECTIONS

For the transfection of CHO-K1, 4×10^4 cells/well were seeded in 24 well plates 24 hours prior to transfection. On the day of transfection, 500 ng total DNA was combined with 1.5 μ L of TRANSIT-X2 (MIRUS) to a total of 50 μ L OptiMEM (Invitrogen) per sample. Transfection mixes were incubated for 15 minutes at room temperature before adding dropwise to cell containing wells and mixing by shaking the plate.

For the transfection of HEK-293T, 1×10^5 cells/well were seeded in 24 well plates 24 hours prior to transfection. On the day of transfection, transfection mixes were made up containing 400 ng total DNA, 3 times μ g DNA in μ L of Xtreme-HP reagent (Merck) diluted up to the final volume of 50 μ L in OptiMEM. Transfection mixes were incubated at room temperature for 30 minutes

before adding dropwise onto cells. The plates were then shaken and replaced into humidified incubators for the required time prior to the endpoint assay.

For the co-transfection of miRNA mimics (Thermo Fisher) and plasmid DNA into HEK-293T cells, Attractene (QIAGEN) was used. 1×10^5 cells were seeded in 500 μL of fully supplemented DMEM 24 hours before transfection. On the day of transfection, 100 ng/well of plasmid DNA and 10 nM miRNA mimic are mixed in 60 μL of OptiMEM before the addition of 1.5 μL Attractene Transfection Reagent. The mixture was incubated at room temperature for 15 minutes before dropwise addition to wells. The plates were then shaken and replaced into humidified incubators for the required time prior to the endpoint assay.

6.4 DATA COLLECTION AND ANALYSIS

6.4.1 FLUORESCENT MICROSCOPY

Cells in 24 well plates, that were seeded to the same concentration, transfected according to the appropriate method and incubated for the required time were loaded onto the Ti2 fluorescent microscope (Nikon). Adjustments were made to exposure time, gain, and focal plane to optimize image quality and were maintained across biological replicates and plates. Images were captured and processed by R scripts for qualities. GFP fluorescence was detected by bandpass filter cube GFP-B with excitation filter 470/40. mTagBFP2 fluorescence was detected using excitation DAPI filter cube with excitation range 375/28 and lastly red fluorescence including mCherry and DsRed was detected using Texas Red filter cube with excitation filter 560/40.

6.4.2 FLOW CYTOMETRY

6.4.2.1 *Sample Preparation*

Spent media was removed from cells and discarded. Cells were gently washed with 1x PBS (Phosphate Buffered Saline) before the addition of 200 μL 1x Trypsin-EDTA. After incubation for 2-5 mins at 37°C 800 μL of 1xPBS+10% FBS mix was added to cells to neutralise Trypsin. Cells were then collected into tubes and centrifuged at 1500 rpm for 5 mins in conditions where dead

cells were to be removed or at 3000 rpm for 5 mins to pellet live and dead cells. The supernatant is removed and resuspended in 0.5 mL of 1x PBS before further processing.

6.4.2.2 *Live/Dead Staining*

Harvested cell pellets were resuspended into a solution containing 0.2 μ M calcein AM in DMSO and 16 μ M ethidium homodimer 1. The samples were then incubated in the dark for 20 minutes before acquisition.

6.4.2.3 *Acquisition*

Samples were transferred into flow tubes by passing through mesh to increase single cell events. Events are recorded with an Attune NxT flow cytometer (Thermo Fisher). The voltages used are reported in **Table 6.4**.

Unless for Live/Dead stained samples, 20,000 viable singlet events, gated from samples of non-transfected cells were recorded. Singlet gates were set in the standard practice from FSC/SSC and FSC-H/FSC-A plots. For Live/Dead staining all events from 100 μ L of sample are recorded.

Table 6.4: Attune flow cytometry settings. Settings used in the acquisition of CHO-K1 and HEK-293T samples reported in this work.

Parameter	Voltage (CHO)	Voltage (HEK)
FSC-A	100	80
SSC-A	370	350
BL1-A	180	150
YL2-A	250	250
YL1-A	400	400
VL1-A	270	270
FSC-H	100	80
SSC-H	370	350
BL1-H	180	150
YL2-H	250	250
YL1-H	400	400

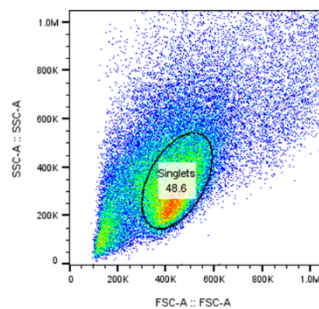
VL1-H	270	270
FSC-W	100	80
SSC-W	370	350
BL1-W	180	150
YL2-W	250	250
YL1-W	400	400
VL1-W	270	270

6.4.3 DATA ANALYSIS

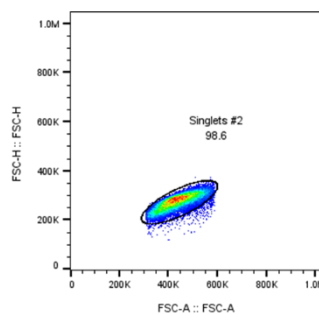
Data collected from the analyser were input into FlowJo for gating and analysis (

STEP 1: Load Acquired Samples into FlowJo

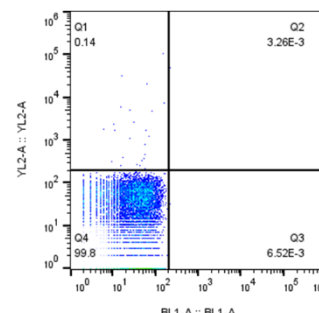
STEP 2: To NTC
Select Singlet population
SSC-A vs FSC-A



STEP 3:
Narrow Singlet population
FSC-H vs FSC-A



STEP 4:
Apply Fluorescence Gate
YL2-A vs BL1-A



STEP 5: Apply to All Samples & Export

Figure 6.1). Single cells were gated on FSC vs SSC, and the BL1 channel (488 nm laser with 530/30 bandpass filter) was used to detect GFP content, YL2 channel (561 nm laser with 620/15 bandpass filter) was used to detect RFP content of samples and VL1 channel (405 nm laser with 450/40 bandpass filter) was used to detect BFP content. Unless otherwise specified geometric mean of fluorescence was used for statistics. The mean of samples from biological repeats are accumulated into the figures and summary statistics performed by GraphPad Prism. In instances where binning was required, singlet gates were extracted as FCS files and imported into R for analysis and figure generation.

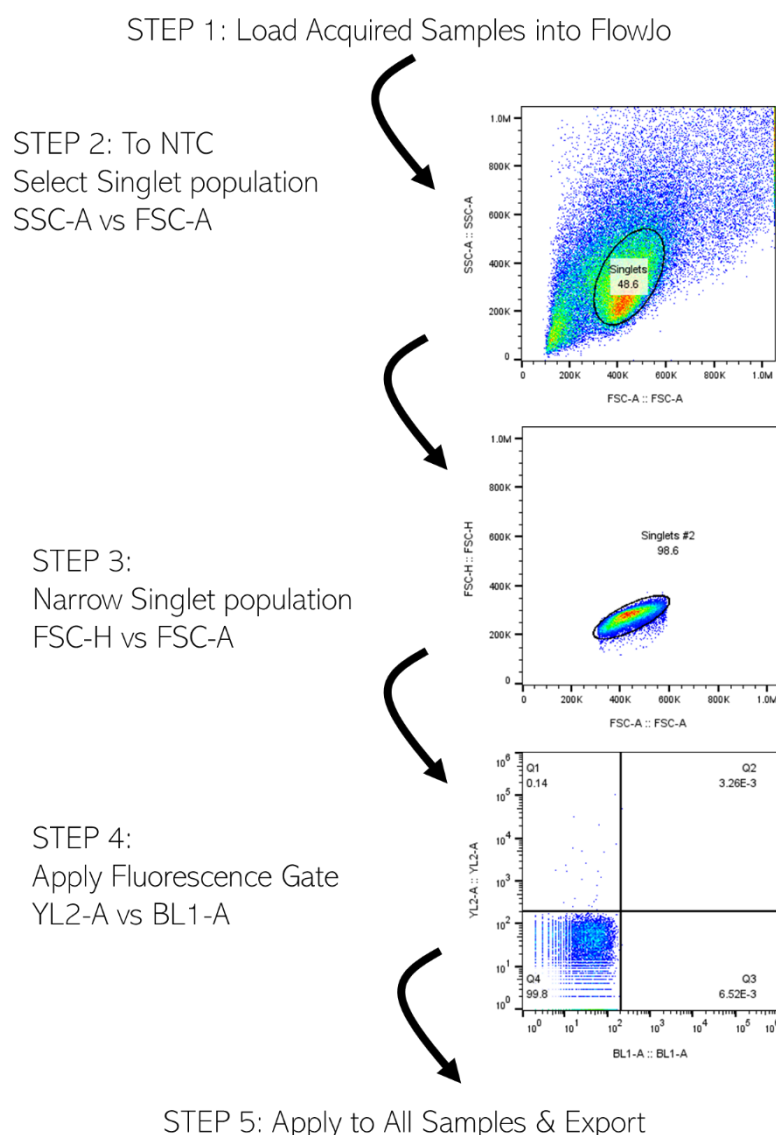


Figure 6.1: Flow analysis using FlowJo. Firstly, the samples are loaded into the workspace of FlowJo. The non-transfected sample (NTC) is then used for sequential steps. The Singlet population is identified and gated. Events that fall within this gate

are then further narrowed down when observed in FSC-H vs FSC-A. A Singlets gate #2 is drawn. The population within this gate is then used to identify fluorescence and relevant gate(s) and histograms are drawn to differentiate non-transfected cells from fluorescent cells. These are then applied to all other samples for comparison. Statistics of the gate(s) can be calculated (count, proportion of parent or geometric mean) and exported into Excel or GraphPad Prism for further visualisation and analysis.

6.5 LIST OF MATERIALS USED

6.5.1 TABLE OF PRIMERS:

Name	Function	Sequence (5' – 3')	Purpose
AZ_Pr_14	PCR	/5Phos/CGCCTTAATTAATTATTAGCCTG ACTTGTACAGCTCGTC	Amplifying backbone mNg in position 7 and adding 3' Pac1 and Asc1 site
AZ_Pr_15	PCR	/5Phos/CGCCGCCCCGAGACGAAGACGA ATTCTCTA	Amplifying backbone mNg in position 7 and adding 3' Pac1 and Asc1 site
AZ_Pr_16	Anneal	TAACTGCGGGAAGGGGGGGAAGAAGT TAGGTGCGGGAAGGGGGGGAAGAAGT	For annealing, introduces 5' Pac1 site and has miR1 complementary region
AZ_Pr_17	Anneal	GGTAACTTCTTCCCCCCTTCCCGCAC CTAACTTCTTCCCCCCTTCCCGCAGTT AAT	For annealing, introduces 5' Pac1 site and has miR1 complementary region
AZ_Pr_48	Anneal	TACCTGCGGGAAGGGGGGGAAGAAGT ATGGTGCGGGAAGGGGGGGAAGAAGT	Creating MREx8 with AscI and PacI overhangs along with AZ_Pr_16 and AZ_Pr_17
AZ_Pr_49	Anneal	CCATACTTCTTCCCCCCTTCCCGCAC CATACTTCTTCCCCCCTTCCCGCA	Creating MREx8 with AscI and PacI overhangs along with AZ_Pr_16 and AZ_Pr_17
AZ_Pr_50	Anneal	CCTAACTTCTTCCCCCCTTCCCGCAC CATACTTCTTCCCCCCTTCCCGCA	Creating MREx8 with AscI and PacI overhangs along with AZ_Pr_16 and AZ_Pr_17

AZ_Pr_51	Anneal	ATGGTGCGGGAAGGGGGGGAAGAAGT ATGGTGCGGGAAGGGGGGGAAGAAGT	Creating MREx8 with Ascl and PacI overhangs along with AZ_Pr_16 and AZ_Pr_17
AZ_Pr_52	Anneal	CGCGCCACTTCTTCCCCCCTTCCCGC AGGTAACCTTCTTCCCCCCTTCCCGCA	Creating MREx8 with Ascl and PacI overhangs along with AZ_Pr_16 and AZ_Pr_17
AZ_Pr_53	Anneal	TAGGTGCGGGAAGGGGGGGAAGAAGT TACCTGCGGGAAGGGGGGGAAGAAGT GG	Creating MREx8 with Ascl and PacI overhangs along with AZ_Pr_16 and AZ_Pr_17
AZ_Pr_65	PCR	aaaagcgatcgcgccaccatggcctcctccga	Forming loop for the expression of DsRed with intronic miR1
AZ_Pr_66	PCR	aaaagttaaacgcatgtagtaagcttgagctcgagat ctc	Forming loop for the expression of DsRed with intronic miR1
AZ_Pr_82	Annealing	caccgagtcgctgtagcgaagca	Fw sgRNA1 with SAM sgRNA overhang (BbsI)
AZ_Pr_83	Annealing	aaactgcttcgctacacgcgactc	Rv sgRNA1 with SAM sgRNA overhang (BbsI)
AZ_Pr_95_fw	PCR	CCTTATTTTTTCAGTAAGTACACTCATG	To amplify RelA promoter from PC12 cells
AZ_Pr_96_rv	PCR	CCATTCGCCAGAGGC	To amplify RelA promoter from PC12 cells
AZ_Pr_97	Annealing	gtatgctgtgacagtgagcgccagcagcacctgggg cagtggtagtgaagcc	Creating TetOff_DsRed_rno- 324-3p
AZ_Pr_98	Annealing	tcactaccactgccccaggtgctgctggcgctcactgtc aacagcatac	Creating TetOff_DsRed_rno- 324-3p
AZ_Pr_99	Annealing	acagatgtaccactgccccaggtgctgctggtgcctac tgctcggag	Creating TetOff_DsRed_rno- 324-3p
AZ_Pr_100	Annealing	aattctccgaggcagtaggcaccagcagcacctgggg cagtgtacatctgtggct	Creating TetOff_DsRed_rno- 324-3p
AZ_Pr_101_R elA FWD	PCR	ggataaccgtattaccgccaCCTTATTTTTTCAGT AAGTACACTCATGTT	For Gibson cloning pRela into pRD04 in place of ACTB
AZ_Pr_102_	PCR	cagctcttcgcccttagacaCCATTCGCCAGAG GCGGAAA	For Gibson cloning pRela into pRD04 in

RelA RV			place of ACTB
AZ_Pr_ 169	Annealin g	caccTACGTTCTCTATCACTGATA	Makes sgRNA2 protospacer (for TetO targetting)
AZ_Pr_ 170	Annealin g	aaacTATCAGTGATAGAGAACGTA	
AZ_Pr_ 177	Anneal	ctagcTGCAAAATCGCAGTCACAGTGACT CAGCAGAATCTGAGCCTAGGgctcatactg gac	Prefix Sequence (NRF- 2)
AZ_Pr_ 178	Anneal	agtatgagcCCTAGGCTCAGATTCTGCTGA GTCAGTGTGACTGCGATTTTGCAg	Prefix Sequence (NRF- 2)
AZ_Pr_ 179	Anneal	cagagtctttTGCAAAATCGCAGTCACAGTG ACTCAGCAGAATCTGAGCCTAGGtaggct	Redundant overhang (NRF-2)
AZ_Pr_ 180	Anneal	taCCTAGGCTCAGATTCTGCTGAGTCAC TGTGACTGCGATTTTGCAaaagactctggtcc	Redundant overhang (NRF-2)
AZ_Pr_ 181	Anneal	cctttgtcaaaagtgatcgctTGCAAAATCGCAG TCACAGTGACTCAGCAGAATCTGAGCC TAGGggac	Redundant overhang (NRF-2)
AZ_Pr_ 182	Anneal	CCTAGGCTCAGATTCTGCTGAGTCACT GTGACTGCGATTTTGCAagcgatcacttttgac aaaggagcc	Redundant overhang (NRF-2)
AZ_Pr_ 183	Anneal	gcttaagtcaatccaatgcgttcaggtgg	Suffix Sequence (NRF- 2)
AZ_Pr_ 184	Anneal	aattccacctgaacgcattggattgacttaagcgctcc	Suffix Sequence (NRF- 2)
AZ_Pr_ 185	PCR	aaGGCGCGCCgcattcattttatgtttcaggttcagg gggag	Creates ARE promoter upstream eGFP
AZ_Pr_ 186	PCR	aaGCGATCGCTttaccaacagtaccggattgccaa gct	Adds AsiSI after ARE/NRF2 promoter
AZ_Pr_ 187	Non- Templat e PCR	TTTTCTAGAGTGACTCAGCAGCTCATA TGGACCAGAGTCTTTGTGACTCAGCATA GGCTCCTTTGTCAAAAGTG	XbaI+AREx2
AZ_Pr_ 188	Non- Templat e PCR	ATCCGAGGAAACAGTTTTTCACGTGACTC AGCAGGACGCTTAAGTCGTTTCAGGTGG TGACTCAGCAATCCGAGGAAACAGTTTT CAC	Elongation AREx2
AZ_Pr_ 189	Non- Templat e PCR	ATCCGAGGAAACAGTTTTTCACGTGACTC AGCATCGGAGGCTTTGACTTTCTGCGT GACTCAGCAATCCGAGGAACAGTTCAC TGATCATT	_AREx2_ SpeI
AZ_Pr_ 269	PCR	aacgtctcgtaGgaaggtacgggaggtattggacag g	Isolating pNFkB from AZ_PL_120 with Bsbml site and overhang C from EMMA kit
AZ_Pr_ 270	PCR	accgtctcaGCTGCCAACAGTACCGGATTG CCAAG	Isolating pNFkB from AZ_PL_120 with Bsbml site and overhang G from EMMA kit

6.5.2 TABLE OF G-BLOCKS:

ID	DNA Sequence	Use
AZ_gblock_1	GTTTAAACGCTAGCAGTGAGTTGATTGCAGTCCAGTTACGCT GGAGTCTGAGGCTCGTCCTGAATGATATCAAGCTTGAATTCG TACGTCTCGGACTTTATACGAAGTTATCTCGGCGCGCCCAGTT CGAGTTTATCATTATCAATACTGCCATTTCAAAGAATACGTA AATAATTAATAGTAGTGATTTTCCTAACTTTATTTAGTCAAAA AATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAAT AGGGGGCGGGTTACACAGAATATATAACATCGTAGGTGTCTG GGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAGC CCGCTTTTTAAGCTGGCATCCAGAAAAAAAAGAATCCCAGC ACCAAAATATTGTTTTCTTCACCAACCATCAGTTCATAGGTCC ATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGG CAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCC TGGAGTAAATGATGACACAAGGCAATTGACCCACGCATGTAT CTATCTCATTTTCTTACACCTTCTATTACCTTCTGCTCTCTG ATTTGAAAAAGCTGAAAAAAAAGGTTGAAACCAGTTCCCT GAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGT AGGTATTGATTGTAATTCTGTAAATCTATTTCTTAAACTTCTT AAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTAAACAC CAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAA GATCTATGAGTGTGATCGCTAAACAAATGACTTATAAGGTCT ACATGTCAGGAACAGTGAATGGACACTATTTTCGAAGTAGAG GGTGACGGTAAAGGTAACCGTATGAGGGTGAACAGACGGT CAAACCTACCGTAACGAAGGGTGGACCCCTGCCTTTTGCCTG GGATATTCTTAGCCACAGTGCCAGTATGGTTCTATTCCCTTC ACCAAAATATCCCGAAGATATTCCAGACTATGTTAAACAAAGT TTTCCTGAGGGCTACACATGGGAGAGAATTATGAATTTTGAG GATGGAGCAGTATGTACAGTATCCAACGACTCCAGCATACAA GGAAATTGCTTTATTTATCATGTGAAGTTTAGTGGTCTGAATT TCCCGCCGAACGGCCCTGTAATGCAGAAAAAACGCAAGGC TGGGAGCCAAACACGGAGAGACTTTTCGCCCGTGATGGAATG CTGTTGGGGAACAACCTTCATGGCCCTGAAGTTAGAAGGAGGA GGCCATTACTTATGTGAGTTTAAGACTACGTATAAAGCTAAA AAACCTGTAAAGATGCCTGGATACCACTATGTCGATAGAAAA CTTGACGTAACCAATCACAATAAAGACTATACGTCCGTCGAG CAGTGTGAGATTAGTATTGCAAGGAAACCCGTTGTTGCGGGA TCCTAACTCGAGATAAAGCAATCTTGATGAGGATAATGATTT TTTTTTGAATATACATAAAATACTACCGTTTTTCTGCTAGATTT TGTGATGACGTAAATAAGTACATATTACTTTTTAAGCCAAGA CAAGATTAAGCATTAACCTTTACCCCTTTTCTTCTAAGTTTCAA TATTAGTTATCACTGTTTAAAGTTATGGCGAGAACGTCGGC GGTAAAAATATATTACCCGTAACGGGTACCTCCATAGTTGCC TGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGC TCACGCTAGCTTATAA	pHook cloning to add rca hooks
AZ_gblock_2	ACGTCTCAAGCAGCAAGCAGTGACTCAGCAGGCTCATACTTG CCCAGAGTCTTTGTGACTCAGCATAGCAACCTTTGTCAAAAG TGATCGCTGTGACTCAGCAACTAGCTTAAGTCAATCCAATGC GTTCAGGTGGTGACTCAGCACTGATTACAGCTGAAGAGACGA	Creating AREx3 for domestication into CTRL

6.5.3 TABLE OF PLASMIDS:

ID	Plasmid	Antibiotic resistance bacteria
AZ_PL_1	pVEGFR1_TEV(C)_TCS(Q'G)_NLS-HA-dCas9(C)-VP64_T2A_MCP-P65-HSF1	Ampicillin
AZ_PL_2	pVEGFR2_TEV(N)_NES_TCS(Q'L)_HA-dCas9(N)_P2A_Puro-WPRE	Ampicillin
AZ_PL_3	pBDKRB2_TCS(Q'G)_NLS-HA-dCas9(C)-VP64_T2A_MCP-P65-HSF1	Ampicillin
AZ_PL_4	pBDKRB2_TCS(Q'L)_HA-dCas9(N)_P2A-Puro-WPRE	Ampicillin
AZ_PL_5	pCDNA3.1(+)-CMV-bArrestin2-TEV	Ampicillin
AZ_PL_6	LmRA constitutive (Voigt Lab)	Ampicillin
AZ_PL_7	pLmRA-YFP (Voigt)	Ampicillin
AZ_PL_8	A3_EMMA kit_RFP_p3	Kanamycin
AZ_PL_9	EMMA A->C Connector	Kanamycin
AZ_PL_10	EMMA d->F Connector	Kanamycin
AZ_PL_11	EMMA ATG	Kanamycin
AZ_PL_12	EMMA Terminator	Kanamycin
AZ_PL_13	EMMA Connector H->K	Kanamycin
AZ_PL_14	EMMA Connector L->Y	Kanamycin
AZ_PL_15	EMMA Connector Y->Z	Kanamycin
AZ_PL_16	EMMA reciever plasmid	Ampicillin
AZ_PL_17	pEF1a in EMMA p3	Kanamycin
AZ_PL_18	mNeoGreen with PacI miR1x4-Ascl in EMMA p7 colony 1	Kanamycin
AZ_PL_19	mNeoGreen with PacI miR1x4-Ascl in EMMA p7 colony 5	Kanamycin
AZ_PL_20	mNeoGreen with PacI miR1x4-Ascl in EMMA p7 colony 5	Kanamycin
AZ_PL_24	Addgene sgRNA1 responsive promoter (8 binding sites) Addgene #55197	Ampicillin
AZ_PL_25	pTRETightBI-RY-0 Addgene #31463	Ampicillin
AZ_PL_26	pTRETightBI-RY-1 Addgene #31464	Ampicillin
AZ_PL_27	pTRETightBI-RY-4 Addgene #31465	Ampicillin
AZ_PL_28	pTRETightBI-RY-7 Addgene #31466	Ampicillin
AZ_PL_29	Control Sensor (pEF1a-mNG-PEST_pA)	Ampicillin
AZ_PL_30	Control Sensor (pEF1a-mNG-PEST_pA)	Ampicillin
AZ_PL_31	2x MRE (pEF1a-mNG-PEST-PacI-MREx2-Ascl)	Ampicillin
AZ_PL_32	4x MRE (pEF1a-mNG-PEST-PacI-MREx4-Ascl)	Ampicillin
AZ_PL_33	sgRNA1 in EMMA p19	Kanamycin
AZ_PL_34	pBI-CMV-EGFP	Ampicillin
AZ_PL_35	pEF1a-UbVr-mNG-PEST_pA)	Ampicillin
AZ_PL_36	pEF1a-UbVr-mNG-PEST-PacI-MREx2-Ascl)	Ampicillin
AZ_PL_37	pEF1a-UbVr-mNG-PEST-PacI-MREx4-Ascl	Ampicillin
AZ_PL_38	psgRNA_MLP in p14 (sgRNA1 responsive promoter in EMMA p14)	Kanamycin
AZ_PL_39	BiDirectional Promoter CMV in EMMA position 3	Kanamycin
AZ_PL_40	sgRNA1 responsive promoter (Addgene #55197)_BFP_Terminator in EMMA rec	Ampicillin
AZ_PL_45	8xMRE UBVR (pEF1a-mNG-PEST-PacI-MREx8-Ascl-UbVr)	Ampicillin
AZ_PL_46	pGLM92-TetOff FF4 expressing and FF5 targetting plasmid	Ampicillin

	(ETH)	
AZ_PL_47	8xMRE (pEF1a-mNG-PEST-Pacl-MREx8-Ascl)	Ampicillin
AZ_PL_48	pEF1a-mNG-PEST-Pacl-MREx4-Ascl)	Ampicillin
AZ_PL_49	pEf1a_mNG_Pacl_MREx0_AscI_PEST_pA	Ampicillin
AZ_PL_50	pEf1a_mNG_UbVr_Pacl_MREx0_AscI_PEST_pA	Ampicillin
AZ_PL_51	pMLP(sgx8)_eGFP + pEf1a_sgRNA_pA	Ampicillin
AZ_PL_52	pMLP(sgx8)_eGFP (EMMA backbone)	Ampicillin
AZ_PL_53	pEF1a-mNG-UbVr_Pacl-MREx4-Ascl_PEST_pA)	Ampicillin
AZ_PL_54	pGLM92 backbone with intronic FF4 replaced with miR1249 (TetOff_DsRed_miR1249)	Ampicillin
AZ_PL_55	Tre3Gp-DsRed express (intron:miR1249-3p)-SV40pA; pEF1a-TetON3G (TetON_DsRed_miR1249)	Ampicillin
AZ_PL_56	pMLP(sgx8)_eGFP (Promoter backbone)	Ampicillin
AZ_PL_57	pEF1a_AscI_AsiSI + psgRNA_mTagBFP	Ampicillin
AZ_PL_60	pDEST-tetOx224_PhiC31attB_FRT-EF1aGFP-FRT (addgene: 119904)	Ampicillin
AZ_PL_61	sgRNA(MS2) cloning backbone (addgene: 61424)	Ampicillin
AZ_PL_62	sgRNA1-SAM scaffold in addgene #61424 backbone	Ampicillin
AZ_PL_65	PB_UniSAM Addgene 99866 (dCas9_SAM +sgRNA_SAM)	Ampicillin
AZ_PL_66	TetOx320 Bao Jun sponge paper (https://www.nature.com/articles/s41467-020-19552-9#Abs1)	Kanamycin
AZ_PL_68	AZ_PL_66 pU6_sgRNA1_pT_KpnI	Ampicillin
AZ_PL_69	TetO(EMMA)	Kanamycin
AZ_PL_70	TetOff_rno_miR324 (pGLM92 backbone	Ampicillin
AZ_PL_71	pRelA_eGFP-SV40	Ampicillin
AZ_PL_72	TetON_DsRed_324	Ampicillin
AZ_PL_73	pJB_DsRed_324	Ampicillin
AZ_PL_74	pEF1a_DsRed_324	Ampicillin
AZ_PL_75	pACTB_DsRed-miR324(intronic)	Ampicillin
AZ_PL_76	TetOx21 in EMMA backbone	Kanamycin
AZ_PL_77	pCMV_DsRed_rno-miR-324	Ampicillin
AZ_PL_78	pHook (from Fabio Liberante)	Kanamycin
AZ_PL_79	Yeast empty backbone (Addgene: pRSII413 Plasmid #35450)	Ampicillin
AZ_PL_80	TetOx320 in Yeast backbone (TetOx320+AZ_PL_79)	Ampicillin
AZ_PL_81	Short Hook rca template vector NRF2	Kanamycin
AZ_PL_82	Long hook TAR Vector	Kanamycin
AZ_PL_83	LongHook rca template vector NRF2	Kanamycin
AZ_PL_84	sgx8	Kanamycin
AZ_PL_85	sgRNA2 (TetO targetting) in AZ_PL_66 backbone	Ampicillin
AZ_PL_86	pRelA(fully complementary miR-21-5p)-eGFP	Ampicillin
AZ_PL_87	Addgene #177775-pGL4.27_ARE/NRF2-SPE Luciferase	Ampicillin
AZ_PL_88	AREx3	Ampicillin
AZ_PL_89	Tetscrx3	Kanamycin
AZ_PL_90	pNRF2/ARE-eGFP	Ampicillin
AZ_PL_91	Tetscrx12	Kanamycin
AZ_PL_92	Tetscrx24	Kanamycin
AZ_PL_93	pLoop: pUAP Entry Vector	Chloramphenicol

AZ_PL_94	pLoop: pCK1 (High Copy-Chromoproteins)	Kanamycin
AZ_PL_95	pLoop: pCK2 (High Copy-Chromoproteins)	Kanamycin
AZ_PL_96	pLoop: pCK3 (High Copy-Chromoproteins)	Kanamycin
AZ_PL_97	pLoop: pCK4 (High Copy-Chromoproteins)	Kanamycin
AZ_PL_98	pLoop: pCSA (High Copy - LacZ)	Spectinomycin
AZ_PL_99	pLoop: pCSB (High Copy - LacZ)	Spectinomycin
AZ_PL_100	pLoop: pCSC (High Copy - LacZ)	Spectinomycin
AZ_PL_101	pLoop: pCSD (High Copy - LacZ)	Spectinomycin
AZ_PL_102	pLoop: pOdd1 (Low Copy - LacZ)	Kanamycin
AZ_PL_103	pLoop: pOdd2 (Low Copy - LacZ)	Kanamycin
AZ_PL_104	pLoop: pOdd3 (Low Copy - LacZ)	Kanamycin
AZ_PL_105	pLoop: pOdd4 (Low Copy - LacZ)	Kanamycin
AZ_PL_106	pLoop: pEven1 (Low Copy - LacZ)	Spectinomycin
AZ_PL_107	pLoop: pEven2 (Low Copy - LacZ)	Spectinomycin
AZ_PL_108	pLoop: pEven3 (Low Copy - LacZ)	Spectinomycin
AZ_PL_109	pLoop: pEven4 (Low Copy - LacZ)	Spectinomycin
AZ_PL_110	NRFx12 in pLoop Vector PcS1	Spectinomycin
AZ_PL_111	Tetscr x48	Kanamycin
AZ_PL_112	Mammalian Sponge Assembly: Entry Vector 1 Amp position 1	Ampicillin
AZ_PL_113	Mammalian Sponge Assembly: Entry Vector 2 Amp position 2	Ampicillin
AZ_PL_114	Mammalian Sponge Assembly: Entry Vector 3 Amp position 3	Ampicillin
AZ_PL_115	Mammalian Sponge Assembly: Entry Vector 4 Amp position 4	Ampicillin
AZ_PL_116	Mammalian Sponge Assembly: Entry Vector 5 Kan position 1	Kanamycin
AZ_PL_117	Mammalian Sponge Assembly: Entry Vector 6 Kan position 2	Kanamycin
AZ_PL_118	Mammalian Sponge Assembly: Entry Vector 7 Kan position 3	Kanamycin
AZ_PL_119	Mammalian Sponge Assembly: Entry Vector 8 Kan position 4	Kanamycin
AZ_PL_120	pNFkB-CFP	Ampicillin

7 SUPPLEMENTARY INFORMATION

7.1 SUPPLEMENTARY DATA

7.1.1.1 *Code for Data Analysis by Binning.*

The following R code was used to analyse the flow cytometry data collected from the coexpression of intronic miRNA expression plasmid and miRNA repression sensor plasmids. Prior to introducing .FCS files into the code, the raw files must be processed to export the singlet population.

```
# Load required libraries

library(flowCore)
library(ggplot2)
library(dplyr)
library(gridExtra)
library(gplots)
library(gridExtra)
library(viridis)
library(tidyr)
library(lattice)

# Function to process a pair of FCS files and calculate fold change

process_fcs_pair <- function(file_path_sensor, file_path_control,
pair_index) {
  flow_data_sensor <- read.FCS(file_path_sensor, transformation = TRUE,
truncate_max_range = FALSE)
  flow_data_control <- read.FCS(file_path_control, transformation = TRUE,
truncate_max_range = FALSE)

# Extract YL2 and BL1 values into dataframes and convert them to type
numeric

yl2_values_sensor <- as.numeric(flow_data_sensor$"YL2-A"@exprs)
bl1_values_sensor <- as.numeric(flow_data_sensor$"BL1-A"@exprs)
```

```

yl2_values_control <- as.numeric(flow_data_control$"YL2-A"@exprs)
bl1_values_control <- as.numeric(flow_data_control$"BL1-A"@exprs)

# Create data frames for YL2 and BL1 values and filter out rows with
negative values

table_sensor <- data.frame(yl2_values = yl2_values_sensor, bl1_values =
bl1_values_sensor)

table_sensor <- table_sensor[!is.na(table_sensor$yl2_values) &
table_sensor$yl2_values >= 0 &
                                !is.na(table_sensor$bl1_values) &
table_sensor$bl1_values >= 0, ]
table_control <- data.frame(yl2_values = yl2_values_control, bl1_values =
bl1_values_control)
table_control <- table_control[!is.na(table_control$yl2_values) &
table_control$yl2_values >= 0 &
                                !is.na(table_control$bl1_values) &
table_control$bl1_values >= 0, ]
# Define the number of bins and bin width & create bins
max_log <- log10(max(1000000))
bin_width <- 10
yl2_bins_sensor <- cut(log10(table_sensor$yl2_values), breaks = seq(0,
max_log, by = log10(bin_width)),
                        include.lowest = TRUE, ordered_result = TRUE)
yl2_bins_control <- cut(log10(table_control$yl2_values), breaks = seq(0,
max_log, by = log10(bin_width)),
                        include.lowest = TRUE, ordered_result = TRUE)

# Combine YL2 bins with corresponding BL1 values

df2s <- data.frame(table_sensor, yl2_bins = yl2_bins_sensor)
df2c <- data.frame(table_control, yl2_bins = yl2_bins_control)

# Group up bins and calculate the mean of GFP values
agg_tbls <- df2s %>% group_by(yl2_bins) %>%
  summarise(mean_GFP_sensor = mean(bl1_values), .groups = 'drop')
agg_tblc <- df2c %>% group_by(yl2_bins) %>%
  summarise(mean_GFP_control = mean(bl1_values), .groups = 'drop')

# Merge agg_tbls and agg_tblc based on the bin identifier

merged_data <- merge(agg_tbls, agg_tblc, by = "yl2_bins", all = TRUE)

# Calculate the fold difference

```

```

merged_data$fold_difference <- ifelse(is.na(merged_data$mean_GFP_sensor)
| is.na(merged_data$mean_GFP_control),
                                     NA, merged_data$mean_GFP_sensor /
merged_data$mean_GFP_control)

# Replace Inf or -Inf with NA if needed
merged_data$fold_difference[is.infinite(merged_data$fold_difference)] <-
NA

return(list(df2 = list(df2s, df2c), merged_data = merged_data))
}

# Path to file pairs ( where sensor filepath, control filepath )
# Make sure to use singlets gated export. Also the same number of events
between control and sensor pairs for comparison of violin plots

file_pairs <- list(
  c("C:/Users/user/Desktop/test/Sensor_Singlets_P3.fcs",
"C:/Users/user/Desktop/test/Control_Singlets_P3.fcs")
)

#Run function and store in results data frame
results <- lapply(file_pairs, function(pair) {
  process_fcs_pair(pair[1], pair[2])
})

```

7.2 STATISTICAL TEST SUMMARY

In this section, we present a comprehensive summary of our statistical analysis. We detail the number of biological replicates for each sample and employ the two-sided Mann–Whitney test to determine statistical significance. It's important to note that the Mann–Whitney test assumes that the samples are independent and that the responses are ordinal. The summary P-values are denoted as follows: **** for $P < 0.0001$, *** for $P < 0.0005$, ** for $P < 0.005$, and * for $P < 0.05$.

In addition to the test results, this section also includes tables that provide descriptive statistics for the graphs. These tables offer a more detailed view of the data distribution and central tendencies.

All statistical analyses and graphical representations were generated using GraphPad Prism.

7.2.1 FIGURE 3.3

Table 7.1: Mann Whitney U Test non parametric test between graph components of Figure 3.3.b

	circuit_i	circuit_j	n_i	n_j	p_val_mannwhitney		SIGNIFICANCE SUMMARY
					Green	Red	
HEK293T	pEmpty	TetO scr	3	3	0.7	-	ns
	pEmpty	pEF1a	3	3	0.4	-	ns
	pEmpty	pTet	3	3	0.0043	-	**
	pEmpty	TetO x7	3	3	0.008	-	**

7.2.2 FIGURE 3.4

Table 7.2: Mann Whitney U nonparametric statistical test between graph components of Figure 3.4

	circuit_i	circuit_j	n_i	n_j	p_val_mannwhitney		SIGNIFICANCE SUMMARY
					Green	Red	
Fold Change (Decoy/No Decoy)	TetOx7	TetOx320	12	12	<0.0001	-	****
	TetO x21	TetOx320	12	12	<0.0001	-	****
	1x TetO7	2x TetO7	4	4	0.2	-	ns

	2x TetO7	4x TetO7	4	4	0.2322	-	ns
	1x TetO21	2x TetO21	4	4	0.1143	-	ns
	2x TetO21	4x TetO21	4	4	0.6857	-	ns
	1x TetO320	2x TetO320	4	4	0.0286	-	*
	2x TetO320	4x TetO320	4	4	0.0286	-	*

Table 7.3: Descriptive Statistics for Figure 3.4.a for the Equimolar transfection of GOI and TetO sponges with 7,21 or 320 sites. Graph title is 1x Tet Decoy

A. 1X				
Number of values	5	5	5	5
Minimum	1767	1878	1941	2018
Maximum	27570	26232	19793	15793
Range	25804	24354	17852	13775
Mean	19018	17264	14656	11300
Std. Deviation	10481	9114	7306	5560
Std. Error of Mean	4687	4076	3267	2486
Lower 95% CI of mean	6004	5948	5585	4397
Upper 95% CI of mean	32032	28580	23727	18204
Geometric mean	13736	12915	11404	9214
Geometric SD factor	3.196	2.963	2.703	2.369

Table 7.4: Descriptive Statistics for Figure 3.4.a for the transfection of double TetO sponges with 7,21 or 320 sites compared to GOI. Graph title is 2x Tet Decoy

A. 2X				
Number of values	5	6	6	6
Minimum	1767	1890	1881	1916
Maximum	27570	21150	16309	7861
Range	25804	19260	14428	5945
Mean	19018	15429	12865	5988
Std. Deviation	10481	7213	5551	2269
Std. Error of Mean	4687	2945	2266	926.5
Lower 95% CI of mean	6004	7859	7040	3606
Upper 95% CI of mean	32032	22999	18690	8370
Geometric mean	13736	12309	10610	5445

Geometric SD factor 3.196 2.54 2.346 1.714

Table 7.5: Descriptive Statistics for Figure 3.4.a for the transfection of quadruple TetO sponges with 7,21 or 320 sites compared to GOI. Graph title is 4x Tet Decoy

A. 4X				
Number of values	5	6	6	6
Minimum	1767	1768	2033	1773
Maximum	27570	17136	17781	5427
Range	25804	15369	15748	3654
Mean	19018	13119	13219	4397
Std. Deviation	10481	5718	5670	1427
Std. Error of Mean	4687	2335	2315	582.6
Lower 95% CI of mean	6004	7117	7268	2899
Upper 95% CI of mean	32032	19120	19170	5894
Geometric mean	13736	10696	10983	4121
Geometric SD factor	3.196	2.426	2.297	1.544

7.2.3 FIGURE 3.5

Table 7.6: Mann Whitney U nonparametric statistical test between graph components of Figure 3.5

	circuit_i	circuit_j	n_i	n_j	p_val_mannwhitney		SIGNIFICANCE SUMMARY
					Green	Red	
HEK	TetOx7	TetOx320	12	12	0.0332	-	*
	TetO x21	TetOx320	12	12	0.0387	-	*
	1x TetO7	2x TetO7	4	4	0.4857	-	ns
	2x TetO7	4x TetO7	4	4	0.3429	-	ns
	1x TetO21	2x TetO21	4	4	0.2	-	ns
	2x TetO21	4x TetO21	4	4	0.3429	-	ns
	1x TetO320	2x TetO320	4	4	0.3492	-	ns
	2x TetO320	4x TetO320	4	4	0.2	-	ns

Table 7.7: Descriptive Statistics for Figure 3.5.a for the Equimolar transfection of GOI and TetO sponges with 7,21 or 320 sites. Graph title is 1x Tet Decoy

A. 1X				
	No Sponge	TetO x7	TetO x21	TetO x320
Number of values	4	4	4	4
Minimum	1069	1074	991.5	1299
Maximum	16952	13562	12486	2603
Range	15882	12488	11494	1304
Mean	10328	9046	7770	1832
Std. Deviation	6891	5550	4943	577.4
Std. Error of Mean	3446	2775	2472	288.7
Lower 95% CI of mean	-637.7	214.3	-95.94	912.8
Upper 95% CI of mean	21293	17877	15635	2750
Geometric mean	6979	6395	5538	1767
Geometric SD factor	3.574	3.312	3.205	1.357
Lower 95% CI of geo. mean	919.6	951.1	867.7	1087
Upper 95% CI of geo. mean	52965	42998	35340	2873

Table 7.8: Descriptive Statistics for Figure 3.5.a for the transfection of double TetO sponges with 7,21 or 320 sites compared to GOI. Graph title is 2x Tet Decoy.

A. 2X				
	No Sponge	TetO x7	TetO x21	TetO x320
Number of values	4	4	4	4
Minimum	965.1	692.5	862.4	1209
Maximum	16034	13924	10980	1796
Range	15069	13231	10118	587.9
Mean	10123	8980	7033	1516
Std. Deviation	6506	6063	4363	261.7
Std. Error of Mean	3253	3032	2181	130.8
Lower 95% CI of mean	-229.8	-667.4	91.33	1099
Upper 95% CI of mean	20476	18628	13975	1932
Geometric mean	6779	5677	4997	1498
Geometric SD factor	3.709	4.144	3.261	1.193
Lower 95% CI of geo. mean	842	591	761.9	1132

Upper 95% CI of geo. mean	54588	54528	32777	1984
------------------------------	-------	-------	-------	------

Table 7.9: Descriptive Statistics for Figure 3.5.a for the transfection of quadruple TetO sponges with 7,21 or 320 sites compared to GOI. Graph title is 4x Tet Decoy

A. 4X				
	No Sponge	TetO x7	TetO x21	TetO x320
Number of values	3	3	3	3
Minimum	1767	9564	9247	1144
Maximum	22536	13338	11414	2794
Range	20769	3774	2167	1651
Mean	13760	11633	10180	1858
Std. Deviation	10752	1913	1114	847.4
Std. Error of Mean	6208	1105	643.4	489.3
Lower 95% CI of mean	-12949	6881	7411	-247.2
Upper 95% CI of mean	40469	16386	12948	3963
Geometric mean	8776	11524	10140	1736
Geometric SD factor	4.037	1.185	1.114	1.568
Lower 95% CI of geo. mean	274.1	7556	7760	568.1
Upper 95% CI of geo. mean	281014	17576	13250	5302

7.2.4 FIGURE 3.8

Table 7.10: Mann Whitney U nonparametric statistical test between graph components of Figure 3.8.c for studies in HEK and 3.8.d for studies in CHO or HEK293T cross talk test 3.8.e

	circuit_i	circuit_j	n_i	n_j	p_val_mannwhitney		SIGNIFICANCE SUMMARY
					Green	Red	
	3.8.c						
HEK	No Decoy	ARE Oligos	6	3	0.381	-	ns
	No Decoy	scr Oligos	6	3	0.381	-	ns
	ARE Oligos	scr Oligos	3	3	0.2683	-	ns
	3.8.e						
	AREx0	AREx16	6	6	0.0346	-	*
	AREx0	AREx32	6	6	0.0087	-	**

	AREx64	AREx128	6	6	0.0043	-	**
	3.8.e						
	No Decoy	AREx16	5	3	-	0.6905	ns
	No Decoy	AREx32	5	3	-	0.3095	ns
	No Decoy	AREx64	5	3	-	0.0079	**
	No Decoy	AREx128	5	3	-	0.0357	*
	No Decoy	TNFa	5	3	-	0.0357	*
3.8.d							
CHO	No Decoy	ARE Oligos	3	3	0.2	-	ns
	No Decoy	scr Oligos	3	3	0.1	-	ns
	ARE Oligos	scr Oligos	3	3	>0.9999	-	ns
	AREx0	AREx16	3	3		-	*
	AREx0	AREx32	3	3		-	**
	AREx64	AREx128	3	3		-	**

Table 7.11: Descriptive Statistics for Figure 3.8.c for the transfection of ARE TF decoys with 16,32, 64 or 128 sites into HEK..

3.8.c					
	x 16	x 32	x 64	x 128	0
Number of values	4	4	4	4	4
Minimum	0.4048	0.3519	0.2355	0.3217	0.2505
Maximum	0.7758	0.761	0.6204	0.6911	0.9298
Range	0.371	0.4091	0.3849	0.3693	0.6793
Mean	0.654	0.5885	0.508	0.4524	0.5572
Std. Deviation	0.1708	0.1712	0.1824	0.1718	0.3018
Std. Error of Mean	0.08541	0.0856	0.09119	0.08589	0.1509
Lower 95% CI of mean	0.3822	0.3161	0.2178	0.179	0.0769
Upper 95% CI of mean	0.9258	0.8609	0.7982	0.7257	1.037
Geometric mean	0.6338	0.5667	0.4741	0.4304	0.494
Geometric SD factor	1.355	1.394	1.596	1.428	1.785
Lower 95% CI of geo. mean	0.3908	0.3341	0.2254	0.2441	0.1964
Upper 95% CI of geo. mean	1.028	0.9612	0.9972	0.7589	1.242

Table 7.12: Descriptive Statistics for Figure 3.8.c for the transfection of ARE TF decoys with 16,32, 64 or 128 sites into CHO.

3.8.d				
	x 16	x 32	x 64	x 128
Number of values	3	3	3	3
Minimum	0.8518	0.5174	0.02382	0.01682
Maximum	1.072	0.68	0.5955	0.5289
Range	0.22	0.1626	0.5717	0.5121
Mean	0.9807	0.6107	0.3439	0.2758
Std. Deviation	0.1148	0.08389	0.2919	0.2561
Std. Error of Mean	0.06628	0.04843	0.1685	0.1479
Lower 95% CI of mean	0.6956	0.4023	-0.3813	-0.3604
Upper 95% CI of mean	1.266	0.8191	1.069	0.9119
Geometric mean	0.9761	0.6067	0.1802	0.1358
Geometric SD factor	1.128	1.153	5.824	6.272
Lower 95% CI of geo. mean	0.7233	0.4263	0.002264	0.00142
Upper 95% CI of geo. mean	1.317	0.8635	14.34	12.99

8 BIBLIOGRAPHY

Adhikary, R., Roy, A., Jolly, M.K. & Das, D. (2023) Effects of microRNA-mediated negative feedback on gene expression noise. *Biophysical Journal*. doi:10.1016/j.bpj.2023.09.019.

Afroz, T. & Beisel, C.L. (2013) Understanding and exploiting feedback in synthetic biology. *Chemical Engineering Science*. 103, 79–90. doi:10.1016/j.ces.2013.02.017.

Alejandro Lavopa and Michele Delera (2021) *What is the Fourth Industrial Revolution?* 31 January 2021. Industrial Analytical Platform. <https://iap.unido.org/articles/what-fourth-industrial-revolution> [Accessed: 27 November 2023].

Ali, M.M., Li, F., Zhang, Z., Zhang, K., Kang, D.-K., Ankrum, J.A., Le, X.C. & Zhao, W. (2014) Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chemical Society Reviews*. 43 (10), 3324. doi:10.1039/c3cs60439j.

Ali, M.Z. & Brewster, R.C. (2022) Controlling gene expression timing through gene regulatory architecture. *PLOS Computational Biology*. 18 (1), e1009745. doi:10.1371/journal.pcbi.1009745.

Amer, B. & Baidoo, E.E.K. (2021) Omics-Driven Biotechnology for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*. 9. doi:10.3389/fbioe.2021.613307.

Anon (2017) The commercial tipping point. 35 (3), 181–181. doi:10.1038/nbt.3829.

Appleton, E., Madsen, C., Roehner, N. & Densmore, D. (2017) Design Automation in Synthetic Biology. *Cold Spring Harbor Perspectives in Biology*. 9 (4), a023978. doi:10.1101/cshperspect.a023978.

Arkin, A.P. (2013) A wise consistency: engineering biology for conformity, reliability, predictability. *Current Opinion in Chemical Biology*. 17 (6), 893–901. doi:10.1016/j.cbpa.2013.09.012.

Arora, S., Rana, R., Chhabra, A., Jaiswal, A. & Rani, V. (2013) miRNA–transcription factor interactions: a combinatorial regulation of gene expression. *Molecular Genetics and Genomics*. 288 (3–4), 77–87. doi:10.1007/s00438-013-0734-z.

Bae, S., Park, J. & Kim, J.-S. (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*. 30 (10), 1473–1475. doi:10.1093/bioinformatics/btu048.

Baeumler, T.A., Ahmed, A.A. & Fulga, T.A. (2017) Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors. *Cell Reports*. 20 (11), 2639–2653. doi:10.1016/j.celrep.2017.08.044.

- Bartoli, V., Meaker, G.A., di Bernardo, M. & Goroehowski, T.E. (2020) Tunable genetic devices through simultaneous control of transcription and translation. *Nature Communications*. 11 (1), 2095. doi:10.1038/s41467-020-15653-7.
- Bashor, C.J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J.J. & Khalil, A.S. (2019) Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science*. 364 (6440), 593–597. doi:10.1126/science.aau8287.
- Beitz, A.M., Oakes, C.G. & Galloway, K.E. (2022) Synthetic gene circuits as tools for drug discovery. *Trends in Biotechnology*. 40 (2), 210–225. doi:10.1016/j.tibtech.2021.06.007.
- Bhatt, B., García-Díaz, P. & Foight, G.W. (2023) Synthetic transcription factor engineering for cell and gene therapy. *Trends in Biotechnology*. doi:10.1016/j.tibtech.2023.09.010.
- Black, J.B., Perez-Pinera, P. & Gersbach, C.A. (2017) Mammalian Synthetic Biology: Engineering Biological Systems. *Annual Review of Biomedical Engineering*. 19 (1), 249–277. doi:10.1146/annurev-bioeng-071516-044649.
- Di Blasi, R., Marbiah, M.M., Siciliano, V., Polizzi, K. & Ceroni, F. (2021) A call for caution in analysing mammalian co-transfection experiments and implications of resource competition in data misinterpretation. *Nature Communications*. 12 (1), 2545. doi:10.1038/s41467-021-22795-9.
- Di Blasi, R., Pisani, M., Tedeschi, F., Marbiah, M.M., Polizzi, K., Furini, S., Siciliano, V. & Ceroni, F. (2023) Resource-aware construct design in mammalian cells. *Nature Communications*. 14 (1), 3576. doi:10.1038/s41467-023-39252-4.
- Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., et al. (2016) The Genome Project-Write. *Science*. 353 (6295), 126–127. doi:10.1126/science.aaf6850.
- Bojar, D. & Fussenegger, M. (2020) The Role of Protein Engineering in Biomedical Applications of Mammalian Synthetic Biology. *Small*. 16 (27). doi:10.1002/sml.201903093.
- Bokes, P. & Singh, A. (2015) Protein Copy Number Distributions for a Self-Regulating Gene in the Presence of Decoy Binding Sites. *PLOS ONE*. 10 (3), e0120555. doi:10.1371/journal.pone.0120555.
- Bolognesi, B. & Lehner, B. (2018) Protein Overexpression: Reaching the limit. *eLife*. 7. doi:10.7554/eLife.39804.
- Borkowski, O., Ceroni, F., Stan, G.-B. & Ellis, T. (2016) Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology*. 33, 123–130. doi:10.1016/j.mib.2016.07.009.
- Bosson, A.D., Zamudio, J.R. & Sharp, P.A. (2014) Endogenous miRNA and Target Concentrations Determine Susceptibility to Potential ceRNA Competition. *Molecular Cell*. 56 (3), 347–359. doi:10.1016/j.molcel.2014.09.018.

- Bowman, E.K., Deaner, M., Cheng, J.-F., Evans, R., Oberortner, E., Yoshikuni, Y. & Alper, H.S. (2020) Bidirectional titration of yeast gene expression using a pooled CRISPR guide RNA approach. *Proceedings of the National Academy of Sciences*. 117 (31), 18424–18430. doi:10.1073/pnas.2007413117.
- Bracken, C.P., Scott, H.S. & Goodall, G.J. (2016) A network-biology perspective of microRNA function and dysfunction in cancer. *Nature Reviews Genetics*. 17 (12), 719–732. doi:10.1038/nrg.2016.134.
- Bragdon, M.D.J., Patel, N., Chuang, J., Levien, E., Bashor, C.J. & Khalil, A.S. (2023) Cooperative assembly confers regulatory specificity and long-term genetic circuit stability. *Cell*. 186 (18), 3810–3825.e18. doi:10.1016/j.cell.2023.07.012.
- Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P. & Church, G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Research*. 40 (15), e117–e117. doi:10.1093/nar/gks624.
- Brophy, J.A.N. & Voigt, C.A. (2014) Principles of genetic circuit design. *Nature Methods*. 11 (5), 508–520. doi:10.1038/nmeth.2926.
- Brown, A.J., Sweeney, B., Mainwaring, D.O. & James, D.C. (2014) Synthetic promoters for CHO cell engineering. *Biotechnology and Bioengineering*. 111 (8), 1638–1647. doi:10.1002/bit.25227.
- Bruetschy, C. (2019) The EU regulatory framework on genetically modified organisms (GMOs). *Transgenic Research*. 28 (S2), 169–174. doi:10.1007/s11248-019-00149-y.
- Buchler, N.E. & Louis, M. (2008) Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology*. 384 (5), 1106–1119. doi:10.1016/j.jmb.2008.09.079.
- Buddingh', B.C. & van Hest, J.C.M. (2017) Artificial Cells: Synthetic Compartments with Life-like Functionality and Adaptivity. *Accounts of Chemical Research*. 50 (4), 769–777. doi:10.1021/acs.accounts.6b00512.
- Calkhoven, C.F. & AB, G. (1996) Multiple steps in the regulation of transcription-factor level and activity. *Biochemical Journal*. 317 (2), 329–342. doi:10.1042/bj3170329.
- Cappell, K.M. & Kochenderfer, J.N. (2023) Long-term outcomes following CAR T cell therapy: what we know so far. *Nature Reviews Clinical Oncology*. 20 (6), 359–371. doi:10.1038/s41571-023-00754-1.
- Cardinale, S., Joachimiak, M.P. & Arkin, A.P. (2013) Effects of Genetic Variation on the E. coli Host-Circuit Interface. *Cell Reports*. 4 (2), 231–237. doi:10.1016/j.celrep.2013.06.023.

- Casamassimi, A. & Ciccodicola, A. (2019) Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation. *International Journal of Molecular Sciences*. 20 (6), 1281. doi:10.3390/ijms20061281.
- Casini, A., Storch, M., Baldwin, G.S. & Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology*. 16 (9), 568–576. doi:10.1038/nrm4014.
- Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., et al. (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. 50 (D1), D165–D173. doi:10.1093/nar/gkab1113.
- Cella, F., Perrino, G., Tedeschi, F., Viero, G., Bosia, C., Stan, G.-B. & Siciliano, V. (2023) MIRELLA: a mathematical model explains the effect of microRNA-mediated synthetic genes regulation on intracellular resource allocation. *Nucleic Acids Research*. 51 (7), 3452–3464. doi:10.1093/nar/gkad151.
- Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods*. 12 (5), 415–418. doi:10.1038/nmeth.3339.
- Chassin, H., Müller, M., Tigges, M., Scheller, L., Lang, M. & Fussenegger, M. (2019) A modular degron library for synthetic circuits in mammalian cells. *Nature Communications*. 10 (1), 2013. doi:10.1038/s41467-019-09974-5.
- Chaturvedi, S., Pablo, M., Wolf, M., Rosas-Rivera, D., Calia, G., Kumar, A.J., Vardi, N., Du, K., Glazier, J., Ke, R., Chan, M.F., Perelson, A.S. & Weinberger, L.S. (2022) Disrupting autorepression circuitry generates “open-loop lethality” to yield escape-resistant antiviral agents. *Cell*. 185 (12), 2086–2102.e22. doi:10.1016/j.cell.2022.04.022.
- Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., Housden, B.E., Perrimon, N., Collins, J.J. & Church, G. (2016) Comparison of Cas9 activators in multiple species. *Nature Methods*. 13 (7), 563–567. doi:10.1038/nmeth.3871.
- Chen, W.C.W., Gaidukov, L., Lai, Y., Wu, M.-R., Cao, J., Gutbrod, M.J., Choi, G.C.G., Utomo, R.P., Chen, Y.-C., Wroblewska, L., Kellis, M., Zhang, L., Weiss, R. & Lu, T.K. (2022) A synthetic transcription platform for programmable gene expression in mammalian cells. *Nature Communications*. 13 (1), 6167. doi:10.1038/s41467-022-33287-9.
- Chen, X., Xiong, W. & Li, H. (2016) Comparison of microRNA expression profiles in K562-cells-derived microvesicles and parental cells, and analysis of their roles in leukemia. *Oncology Letters*. 12 (6), 4937–4948. doi:10.3892/ol.2016.5308.

- Cho, H.-Y. (2013) Genomic Structure and Variation of Nuclear Factor (Erythroid-Derived 2)-Like 2. *Oxidative Medicine and Cellular Longevity*. 2013, 1–24. doi:10.1155/2013/286524.
- Chong, S., Graham, T.G.W., Dugast-Darzacq, C., Dailey, G.M., Darzacq, X. & Tjian, R. (2022) Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription. *Molecular Cell*. 82 (11), 2084–2097.e5. doi:10.1016/j.molcel.2022.04.007.
- Chu, H.S., Ryum, J., Park, S.Y., Kim, B.G., Kim, D.M. & Won, J.I. (2011) A new cloning strategy for generating multiple repeats of a repetitive polypeptide based on non-template PCR. *Biotechnology Letters*. 33 (5), 977–983. doi:10.1007/s10529-010-0510-7.
- Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M. & Zaret, K.S. (2002) Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell*. 9 (2), 279–289. doi:10.1016/S1097-2765(02)00459-8.
- Clark, A.J. (1973) RECOMBINATION DEFICIENT MUTANTS OF *E. COLI* AND OTHER BACTERIA. *Annual Review of Genetics*. 7 (1), 67–86. doi:10.1146/annurev.ge.07.120173.000435.
- Cox, D.B.T., Platt, R.J. & Zhang, F. (2015) Therapeutic genome editing: prospects and challenges. *Nature Medicine*. 21 (2), 121–131. doi:10.1038/nm.3793.
- Crinelli, R., Bianchi, M., Gentilini, L., Palma, L., Sørensen, MD., Bryld, T., Babu, R.B., Arar, K., Wengel, J. & Magnani, M. (2004) Transcription factor decoy oligonucleotides modified with locked nucleic acids: an in vitro study to reconcile biostability with binding affinity. *Nucleic Acids Research*. 32 (6), 1874–1885. doi:10.1093/nar/gkh503.
- Dagar, G., Gupta, A., Masoodi, T., Nisar, S., Merhi, M., Hashem, S., Chauhan, R., Dagar, M., Mirza, S., Bagga, P., Kumar, R., Akil, A.S.A.-S., Macha, M.A., Haris, M., Uddin, S., Singh, M. & Bhat, A.A. (2023) Harnessing the potential of CAR-T cell therapy: progress, challenges, and future directions in hematological and solid tumor treatments. *Journal of Translational Medicine*. 21 (1), 449. doi:10.1186/s12967-023-04292-3.
- Dangi, A.K., Sinha, R., Dwivedi, S., Gupta, S.K. & Shukla, P. (2018) Cell Line Techniques and Gene Editing Tools for Antibody Production: A Review. *Frontiers in Pharmacology*. 9. doi:10.3389/fphar.2018.00630.
- Depienne, C. & Mandel, J.-L. (2021) 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *The American Journal of Human Genetics*. 108 (5), 764–785. doi:10.1016/j.ajhg.2021.03.011.
- Dharap, A., Pokrzywa, C., Murali, S., Pandi, G. & Vemuganti, R. (2013) MicroRNA miR-324-3p Induces Promoter-Mediated Expression of RelA Gene. *PLoS ONE*. 8 (11), e79467. doi:10.1371/journal.pone.0079467.

Ding, Y., Perez-Ortiz, G., Peate, J. & Barry, S.M. (2022) Redesigning Enzymes for Biocatalysis: Exploiting Structural Understanding for Improved Selectivity. *Frontiers in Molecular Biosciences*. 9. doi:10.3389/fmolb.2022.908285.

Donahue, P.S., Draut, J.W., Muldoon, J.J., Edelstein, H.I., Bagheri, N. & Leonard, J.N. (2020) The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nature Communications*. 11 (1), 779. doi:10.1038/s41467-019-14147-5.

Donati, S., Ciuffi, S. & Brandi, M.L. (2019) Human Circulating miRNAs Real-time qRT-PCR-based Analysis: An Overview of Endogenous Reference Genes Used for Data Normalization. *International Journal of Molecular Sciences*. 20 (18), 4353. doi:10.3390/ijms20184353.

Doshi, J., Willis, K., Madurga, A., Stelzer, C. & Benenson, Y. (2020) Multiple Alternative Promoters and Alternative Splicing Enable Universal Transcription-Based Logic Computation in Mammalian Cells. *Cell Reports*. 33 (9), 108437. doi:10.1016/j.celrep.2020.108437.

Durrant, M.G., Fanton, A., Tycko, J., Hinks, M., Chandrasekaran, S.S., Perry, N.T., Schaepe, J., Du, P.P., Lotfy, P., Bassik, M.C., Bintu, L., Bhatt, A.S. & Hsu, P.D. (2023) Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nature Biotechnology*. 41 (4), 488–499. doi:10.1038/s41587-022-01494-w.

Ebert, M.S., Neilson, J.R. & Sharp, P.A. (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods*. 4 (9), 721–726. doi:10.1038/nmeth1079.

Ebrahimi, S.B., Samanta, D., Kusmierz, C.D. & Mirkin, C.A. (2022) Protein transfection via spherical nucleic acids. *Nature Protocols*. 17 (2), 327–357. doi:10.1038/s41596-021-00642-x.

Eisenhut, P., Marx, N., Borsi, G., Papež, M., Ruggeri, C., Baumann, M. & Borth, N. (2024) Manipulating gene expression levels in mammalian cell factories: An outline of synthetic molecular toolboxes to achieve multiplexed control. *New Biotechnology*. 79, 1–19. doi:10.1016/j.nbt.2023.11.003.

Elani, Y. (2021) Interfacing Living and Synthetic Cells as an Emerging Frontier in Synthetic Biology. *Angewandte Chemie International Edition*. 60 (11), 5602–5611. doi:10.1002/anie.202006941.

Fang, Z. & Rajewsky, N. (2011) The Impact of miRNA Target Sites in Coding Sequences and in 3'UTRs. *PLoS ONE*. 6 (3), e18067. doi:10.1371/journal.pone.0018067.

Fazi, F. & Nervi, C. (2008) MicroRNA: basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular Research*. 79 (4), 553–561. doi:10.1093/cvr/cvn151.

Feng, Y.-L., Liu, S.-C., Chen, R.-D., Sun, X.-N., Xiao, J.-J., Xiang, J.-F. & Xie, A.-Y. (2023) Proximal binding of dCas9 at a DNA double strand break stimulates homology-directed repair

as a local inhibitor of classical non-homologous end joining. *Nucleic Acids Research*. 51 (6), 2740–2758. doi:10.1093/nar/gkad116.

Fischer, S., Marquart, K.F., Pieper, L.A., Fieder, J., Gamer, M., Gorr, I., Schulz, P. & Bradl, H. (2017) miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. *Biotechnology and Bioengineering*. 114 (7), 1495–1510. doi:10.1002/bit.26280.

Fonseca, J., Bonny, A., Town, J. & El-Samad, H. (2020) Assembly of Genetic Circuits with the Mammalian ToolKit. *BIO-PROTOCOL*. 10 (5). doi:10.21769/BioProtoc.3547.

Fonseca, J.P., Bonny, A.R., Kumar, G.R., Ng, A.H., Town, J., Wu, Q.C., Aslankoohi, E., Chen, S.Y., Dods, G., Harrigan, P., Osimiri, L.C., Kistler, A.L. & El-Samad, H. (2019) A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synthetic Biology*. 8 (11), 2593–2606. doi:10.1021/acssynbio.9b00322.

Frei, T., Cella, F., Tedeschi, F., Gutiérrez, J., Stan, G.-B., Khammash, M. & Siciliano, V. (2020) Characterization and mitigation of gene expression burden in mammalian cells. *Nature Communications*. 11 (1), 4641. doi:10.1038/s41467-020-18392-x.

Friedman, R.C., Farh, K.K.-H., Burge, C.B. & Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*. 19 (1), 92–105. doi:10.1101/gr.082701.108.

Fritz, B.R., Timmerman, L.E., Daringer, N.M., Leonard, J.N. & Jewett, M.C. (2010) Biology by Design: From Top to Bottom and Back. *Journal of Biomedicine and Biotechnology*. 2010, 1–11. doi:10.1155/2010/232016.

Fukunaga, K., Dhamodharan, V., Miyahira, N., Nomura, Y., Mustafina, K., Oosumi, Y., Takayama, K., Kanai, A. & Yokobayashi, Y. (2023) Small-Molecule Aptamer for Regulating RNA Functions in Mammalian Cells and Animals. *Journal of the American Chemical Society*. 145 (14), 7820–7828. doi:10.1021/jacs.2c12332.

Fus-Kujawa, A., Prus, P., Bajdak-Rusinek, K., Teper, P., Gawron, K., Kowalczyk, A. & Sieron, A.L. (2021) An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Frontiers in Bioengineering and Biotechnology*. 9. doi:10.3389/fbioe.2021.701031.

Gam, J.J., Babb, J. & Weiss, R. (2018) A mixed antagonistic/synergistic miRNA repression model enables accurate predictions of multi-input miRNA sensor activity. *Nature Communications*. 9 (1), 2430. doi:10.1038/s41467-018-04575-0.

Gam, J.J., DiAndreth, B., Jones, R.D., Huh, J. & Weiss, R. (2019) A 'poly-transfection' method for rapid, one-pot characterization and optimization of genetic systems. *Nucleic Acids Research*. 47 (18), e106–e106. doi:10.1093/nar/gkz623.

- Ganner, A., Pfeiffer, Z.-C., Wingendorf, L., Kreis, S., Klein, M., Walz, G. & Neumann-Haefelin, E. (2020) The acetyltransferase p300 regulates NRF2 stability and localization. *Biochemical and Biophysical Research Communications*. 524 (4), 895–902. doi:10.1016/j.bbrc.2020.02.006.
- Ganoza, M.C., Kiel, M.C. & Aoki, H. (2002) Evolutionary Conservation of Reactions in Translation. *Microbiology and Molecular Biology Reviews*. 66 (3), 460–485. doi:10.1128/MMBR.66.3.460-485.2002.
- Gao, W., Guo, L., Yang, Y., Wang, Y., Xia, S., Gong, H., Zhang, B.-K. & Yan, M. (2022) Dissecting the Crosstalk Between Nrf2 and NF- κ B Response Pathways in Drug-Induced Toxicity. *Frontiers in Cell and Developmental Biology*. 9. doi:10.3389/fcell.2021.809952.
- Gao, X., Tsang, J.C.H., Gaba, F., Wu, D., Lu, L. & Liu, P. (2014) Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Research*. 42 (20), e155–e155. doi:10.1093/nar/gku836.
- Gao, Y., Wang, L. & Wang, B. (2023) Customizing cellular signal processing by synthetic multi-level regulatory circuits. *Nature Communications*. 14 (1), 8415. doi:10.1038/s41467-023-44256-1.
- Giannoukos, G., Ciulla, D.M., Marco, E., Abdulkarim, H.S., Barrera, L.A., Bothmer, A., Dhanapal, V., Gloskowski, S.W., Jayaram, H., Maeder, M.L., Skor, M.N., Wang, T., Myer, V.E. & Wilson, C.J. (2018) UDiTaSM, a genome editing detection method for indels and genome rearrangements. *BMC Genomics*. 19 (1), 212. doi:10.1186/s12864-018-4561-9.
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. & Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 6 (5), 343–345. doi:10.1038/nmeth.1318.
- Del Giudice, M., Bo, S., Grigolon, S. & Bosia, C. (2018) On the role of extrinsic noise in microRNA-mediated bimodal gene expression. *PLOS Computational Biology*. 14 (4), e1006063. doi:10.1371/journal.pcbi.1006063.
- Glykofrydis, F. & Elfick, A. (2022) Exploring standards for multicellular mammalian synthetic biology. *Trends in Biotechnology*. 40 (11), 1299–1312. doi:10.1016/j.tibtech.2022.06.001.
- Gómez-Schiavon, M., Dods, G., El-Samad, H. & Ng, A.H. (2020) Multidimensional Characterization of Parts Enhances Modeling Accuracy in Genetic Circuits. *ACS Synthetic Biology*. 9 (11), 2917–2926. doi:10.1021/acssynbio.0c00288.
- Gossen, M. & Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*. 89 (12), 5547–5551. doi:10.1073/pnas.89.12.5547.

- Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W. & Bujard, H. (1995) Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science*. 268 (5218), 1766–1769. doi:10.1126/science.7792603.
- Greco, F.V., Pandi, A., Erb, T.J., Grierson, C.S. & Gorochoowski, T.E. (2021a) Harnessing the central dogma for stringent multi-level control of gene expression. *Nature Communications*. 12 (1), 1738. doi:10.1038/s41467-021-21995-7.
- Greco, F.V., Pandi, A., Erb, T.J., Grierson, C.S. & Gorochoowski, T.E. (2021b) Harnessing the central dogma for stringent multi-level control of gene expression. *Nature Communications*. 12 (1), 1738. doi:10.1038/s41467-021-21995-7.
- Grob, A., Di Blasi, R. & Ceroni, F. (2021) Experimental tools to reduce the burden of bacterial synthetic biology. *Current Opinion in Systems Biology*. 28, 100393. doi:10.1016/j.coisb.2021.100393.
- Grob, A., Colleran, C. & McStay, B. (2014) Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. *Genes & Development*. 28 (3), 220–230. doi:10.1101/gad.234591.113.
- Gross, F. (2019) Occam's Razor in Molecular and Systems Biology. *Philosophy of Science*. 86 (5), 1134–1145. doi:10.1086/705474.
- Guerriaud, M. & Kohli, E. (2022) RNA-based drugs and regulation: Toward a necessary evolution of the definitions issued from the European union legislation. *Frontiers in Medicine*. 9. doi:10.3389/fmed.2022.1012497.
- Gyorgy, A. (2023) Competition and evolutionary selection among core regulatory motifs in gene expression control. *Nature Communications*. 14 (1), 8266. doi:10.1038/s41467-023-43327-7.
- He, C. (2019) Special Issue on Regulating the Central Dogma. *Biochemistry*. 58 (5), 295–296. doi:10.1021/acs.biochem.9b00059.
- He, F., Ru, X. & Wen, T. (2020) NRF2, a Transcription Factor for Stress Response and Beyond. *International Journal of Molecular Sciences*. 21 (13), 4777. doi:10.3390/ijms21134777.
- Health and Safety Executive (2019) *Genetically Modified Organisms (Contained Use)*. 2019. <https://www.hse.gov.uk/biosafety/gmo/index.htm> [Accessed: 24 November 2023].
- Heffner, K.M., Wang, Q., Hizal, D.B., Can, Ö. & Betenbaugh, M.J. (2018) *Glycoengineering of Mammalian Expression Systems on a Cellular Level*. In: pp. 37–69. doi:10.1007/10_2017_57.
- Hendel, S.J. & Shoulders, M.D. (2021) Directed evolution in mammalian cells. *Nature Methods*. 18 (4), 346–357. doi:10.1038/s41592-021-01090-x.

- Henley, M.J. & Koehler, A.N. (2021) Advances in targeting ‘undruggable’ transcription factors with small molecules. *Nature Reviews Drug Discovery*. 20 (9), 669–688. doi:10.1038/s41573-021-00199-0.
- Hermesen, R., Erickson, D.W. & Hwa, T. (2011) Speed, Sensitivity, and Bistability in Auto-activating Signaling Circuits. *PLoS Computational Biology*. 7 (11), e1002265. doi:10.1371/journal.pcbi.1002265.
- Hetz, C. & Papa, F.R. (2018) The Unfolded Protein Response and Cell Fate Control. *Molecular Cell*. 69 (2), 169–181. doi:10.1016/j.molcel.2017.06.017.
- Hobert, O. (2004) Common logic of transcription factor and microRNA action. *Trends in Biochemical Sciences*. 29 (9), 462–468. doi:10.1016/j.tibs.2004.07.001.
- Hobert, O. (2008) Gene Regulation by Transcription Factors and MicroRNAs. *Science*. 319 (5871), 1785–1786. doi:10.1126/science.1151651.
- Van Hoecke, L. & Roose, K. (2019) How mRNA therapeutics are entering the monoclonal antibody field. *Journal of Translational Medicine*. 17 (1), 54. doi:10.1186/s12967-019-1804-8.
- Holland, A.J., Fachinetti, D., Han, J.S. & Cleveland, D.W. (2012) Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. *Proceedings of the National Academy of Sciences*. 109 (49). doi:10.1073/pnas.1216880109.
- Hoose, A., Vellacott, R., Storch, M., Freemont, P.S. & Ryadnov, M.G. (2023) DNA synthesis technologies to close the gene writing gap. *Nature Reviews Chemistry*. 7 (3), 144–161. doi:10.1038/s41570-022-00456-9.
- Horton, C.A., Alexandari, A.M., Hayes, M.G.B., Marklund, E., Schaepe, J.M., Aditham, A.K., Shah, N., Suzuki, P.H., Shrikumar, A., Afek, A., Greenleaf, W.J., Gordân, R., Zeitlinger, J., Kundaje, A. & Fordyce, P.M. (2023) Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*. 381 (6664). doi:10.1126/science.add1250.
- Hou, Q. & Jaffrey, S.R. (2023) Synthetic biology tools to promote the folding and function of RNA aptamers in mammalian cells. *RNA Biology*. 20 (1), 198–206. doi:10.1080/15476286.2023.2206248.
- Howell, E.L., Scheufele, D.A., Brossard, D., Xenos, M.A., Kwon, S., Youtie, J. & Shapira, P. (2020) *Scientists’ and the Publics’ Views of Synthetic Biology*. In: pp. 371–387. doi:10.1007/978-3-030-27264-7_16.
- Huang, V., Place, R.F., Portnoy, V., Wang, J., Qi, Z., Jia, Z., Yu, A., Shuman, M., Yu, J. & Li, L.-C. (2012) Upregulation of Cyclin B1 by miRNA and its implications in cancer. *Nucleic Acids Research*. 40 (4), 1695–1707. doi:10.1093/nar/gkr934.
- Huliák, I., Sike, Á., Zencir, S. & Boros, I.M. (2012) The Objectivity of Reporters: Interference Between Physically Unlinked Promoters Affects Reporter Gene Expression in Transient

Transfection Experiments. *DNA and Cell Biology*. 31 (11), 1580–1584.
doi:10.1089/dna.2012.1711.

Illumina (n.d.) *Revealing the most challenging regions of the genome*.
<https://www.illumina.com/products/by-brand/complete-long-reads-portfolio.html> [Accessed: 18 March 2024].

Jahnke, K., Ritzmann, N., Fichtler, J., Nitschke, A., Dreher, Y., Abele, T., Hofhaus, G., Platzman, I., Schröder, R.R., Müller, D.J., Spatz, J.P. & Göpfrich, K. (2021) Proton gradients from light-harvesting *E. coli* control DNA assemblies for synthetic cells. *Nature Communications*. 12 (1), 3967. doi:10.1038/s41467-021-24103-x.

Jahnke, K., Weiss, M., Weber, C., Platzman, I., Göpfrich, K. & Spatz, J.P. (2020) Engineering Light-Responsive Contractile Actomyosin Networks with DNA Nanotechnology. *Advanced Biosystems*. 4 (9). doi:10.1002/adbi.202000102.

Janowski, B.A., Younger, S.T., Hardy, D.B., Ram, R., Huffman, K.E. & Corey, D.R. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nature Chemical Biology*. 3 (3), 166–173. doi:10.1038/nchembio860.

Janssen, S., Cuvier, O., Müller, M. & Laemmli, U.K. (2000) Specific Gain- and Loss-of-Function Phenotypes Induced by Satellite-Specific DNA-Binding Drugs Fed to *Drosophila melanogaster*. *Molecular Cell*. 6 (5), 1013–1024. doi:10.1016/S1097-2765(00)00100-3.

Javdan, S.B. & Deans, T.L. (2021) Design and development of engineered receptors for cell and tissue engineering. *Current Opinion in Systems Biology*. 28, 100363.
doi:10.1016/j.coisb.2021.100363.

Johansson, K., Cebula, M., Rengby, O., Dreij, K., Carlström, K.E., Sigmundsson, K., Piehl, F. & Arnér, E.S.J. (2017) Cross Talk in HEK293 Cells Between Nrf2, HIF, and NF- κ B Activities upon Challenges with Redox Therapeutics Characterized with Single-Cell Resolution. *Antioxidants & Redox Signaling*. 26 (6), 229–246. doi:10.1089/ars.2015.6419.

Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., Kuksenko, O., Abudayyeh, O.O., Gootenberg, J.S., Fu, Z., Macrae, R.K., Buenrostro, J.D., Regev, A. & Zhang, F. (2023) A transcription factor atlas of directed differentiation. *Cell*. 186 (1), 209–229.e26. doi:10.1016/j.cell.2022.11.026.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2.

Jusiak, B., Cleto, S., Perez-Piñera, P. & Lu, T.K. (2016) Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends in Biotechnology*. 34 (7), 535–547.
doi:10.1016/j.tibtech.2015.12.014.

- Kabir, M.H., Patrick, R., Ho, J.W.K. & O'Connor, M.D. (2018) Identification of active signaling pathways by integrating gene expression and protein interaction data. *BMC Systems Biology*. 12 (S9), 120. doi:10.1186/s12918-018-0655-x.
- Kallunki, T., Barisic, M., Jäättelä, M. & Liu, B. (2019) How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells*. 8 (8). doi:10.3390/cells8080796.
- Kang, K.-N. & Lee, Y.-S. (2012) *RNA Aptamers: A Review of Recent Trends and Applications*. In: pp. 153–169. doi:10.1007/10_2012_136.
- Kang, R., Zhang, Y., Huang, Q., Meng, J., Ding, R., Chang, Y., Xiong, L. & Guo, Z. (2019) EnhancerDB: a resource of transcriptional regulation in the context of enhancers. *Database*. 2019. doi:10.1093/database/bay141.
- Kastberg, L.L.B., Ard, R., Jensen, M.K. & Workman, C.T. (2022) Burden Imposed by Heterologous Protein Production in Two Major Industrial Yeast Cell Factories: Identifying Sources and Mitigation Strategies. *Frontiers in Fungal Biology*. 3. doi:10.3389/ffunb.2022.827704.
- Khalil, A.S. & Collins, J.J. (2010) Synthetic biology: applications come of age. *Nature Reviews Genetics*. 11 (5), 367–379. doi:10.1038/nrg2775.
- Kim, K. & Guck, J. (2020) The Relative Densities of Cytoplasm and Nuclear Compartments Are Robust against Strong Perturbation. *Biophysical Journal*. 119 (10), 1946–1957. doi:10.1016/j.bpj.2020.08.044.
- Kim, M.-S. & Kini, A.G. (2017) Engineering and Application of Zinc Finger Proteins and TALEs for Biomedical Research. *Molecules and Cells*. 40 (8), 533–541. doi:10.14348/molcells.2017.0139.
- Kingwell, K. (2021) Small activating RNAs lead the charge to turn up gene expression. *Nature Reviews Drug Discovery*. 20 (8), 573–574. doi:10.1038/d41573-021-00127-2.
- Kitada, T., DiAndreth, B., Teague, B. & Weiss, R. (2018) Programming gene and engineered-cell therapies with synthetic biology. *Science*. 359 (6376). doi:10.1126/science.aad1067.
- Kojima, R. & Fussenegger, M. (2019) Synthetic Biology: Engineering Mammalian Cells To Control Cell-to-Cell Communication at Will. *ChemBioChem*. 20 (8), 994–1002. doi:10.1002/cbic.201800682.
- Krawczyk, K., Scheller, L., Kim, H. & Fussenegger, M. (2020) Rewiring of endogenous signaling pathways to genomic targets for therapeutic cell reprogramming. *Nature Communications*. 11 (1), 608. doi:10.1038/s41467-020-14397-8.
- Krinsky, N., Kaduri, M., Zinger, A., Shainsky-Roitman, J., Goldfeder, M., Benhar, I., HersHKovitz, D. & Schroeder, A. (2018) Synthetic Cells Synthesize Therapeutic Proteins inside Tumors. *Advanced Healthcare Materials*. 7 (9). doi:10.1002/adhm.201701163.

- Kuosmanen, S.M., Viitala, S., Laitinen, T., Peräkylä, M., Pölönen, P., Kansanen, E., Leinonen, H., Raju, S., Wienecke-Baldacchino, A., Närvänen, A., Poso, A., Heinäniemi, M., Heikkinen, S. & Levonen, A.-L. (2016) The Effects of Sequence Variation on Genome-wide NRF2 Binding—New Target Genes and Regulatory SNPs. *Nucleic Acids Research*. 44 (4), 1760–1775. doi:10.1093/nar/gkw052.
- Lai, E.C. (2002) Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genetics*. 30 (4), 363–364. doi:10.1038/ng865.
- Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. & Weirauch, M.T. (2018) The Human Transcription Factors. *Cell*. 172 (4), 650–665. doi:10.1016/j.cell.2018.01.029.
- Lancrey, A., Joubert, A. & Boulé, J.-B. (2018) Locus specific engineering of tandem DNA repeats in the genome of *Saccharomyces cerevisiae* using CRISPR/Cas9 and overlapping oligonucleotides. *Scientific Reports*. 8 (1), 7127. doi:10.1038/s41598-018-25508-3.
- Larsen, J.S., Karlsson, R.T.G., Tian, W., Schulz, M.A., Matthes, A., Clausen, H., Petersen, B.L. & Yang, Z. (2020) Engineering mammalian cells to produce plant-specific N-glycosylation on proteins. *Glycobiology*. 30 (8), 528–538. doi:10.1093/glycob/cwaa009.
- Laufer, B.I. & Singh, S.M. (2015) Strategies for precision modulation of gene expression by epigenome editing: an overview. *Epigenetics & Chromatin*. 8 (1), 34. doi:10.1186/s13072-015-0023-7.
- Lee, T.H. & Maheshri, N. (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Molecular Systems Biology*. 8. doi:10.1038/msb.2012.7.
- Lee, T.I. & Young, R.A. (2013) Transcriptional Regulation and Its Misregulation in Disease. *Cell*. 152 (6), 1237–1251. doi:10.1016/j.cell.2013.02.014.
- Lehal, R., Zaric, J., Vigolo, M., Urech, C., Frismantas, V., et al. (2020) Pharmacological disruption of the Notch transcription factor complex. *Proceedings of the National Academy of Sciences*. 117 (28), 16292–16301. doi:10.1073/pnas.1922606117.
- Lemus-Diaz, N., Tamon, L. & Gruber, J. (2018) Dual Fluorescence Reporter Based Analytical Flow Cytometry for miRNA Induced Regulation in Mammalian Cells. *BIO-PROTOCOL*. 8 (17). doi:10.21769/BioProtoc.3000.
- Li, L.-C., Okino, S.T., Zhao, H., Pookot, D., Place, R.F., Urakami, S., Enokida, H. & Dahiya, R. (2006) Small dsRNAs induce transcriptional activation in human cells. *Proceedings of the National Academy of Sciences*. 103 (46), 17337–17342. doi:10.1073/pnas.0607015103.
- Li, X. & Heyer, W.-D. (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research*. 18 (1), 99–113. doi:10.1038/cr.2008.1.

- Li, Y. & Zhou, L. (2021) dCas9 techniques for transcriptional repression in mammalian cells: Progress, applications and challenges. *BioEssays*. 43 (9). doi:10.1002/bies.202100086.
- Li, Y.F. & Altman, R.B. (2018) Systematic target function annotation of human transcription factors. *BMC Biology*. 16 (1), 4. doi:10.1186/s12915-017-0469-0.
- Lienert, F., Lohmueller, J.J., Garg, A. & Silver, P.A. (2014) Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature Reviews Molecular Cell Biology*. 15 (2), 95–107. doi:10.1038/nrm3738.
- Lillacci, G., Benenson, Y. & Khammash, M. (2018) Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic Acids Research*. 46 (18), 9855–9863. doi:10.1093/nar/gky795.
- Lim, F., Solvason, J.J., Ryan, G.E., Le, S.H., Jindal, G.A., Steffen, P., Jandu, S.K. & Farley, E.K. (2024) Affinity-optimizing enhancer variants disrupt development. *Nature*. doi:10.1038/s41586-023-06922-8.
- Lister, A., Nedjadi, T., Kitteringham, N.R., Campbell, F., Costello, E., Lloyd, B., Copple, I.M., Williams, S., Owen, A., Neoptolemos, J.P., Goldring, C.E. & Park, B.K. (2011) Nrf2 is overexpressed in pancreatic cancer: implications for cell proliferation and therapy. *Molecular Cancer*. 10 (1), 37. doi:10.1186/1476-4598-10-37.
- Liu, Q., Schumacher, J., Wan, X., Lou, C. & Wang, B. (2018) Orthogonality and Burdens of Heterologous AND Gate Gene Circuits in *E. coli*. *ACS Synthetic Biology*. 7 (2), 553–564. doi:10.1021/acssynbio.7b00328.
- Liu, X., Wu, B., Szary, J., Kofoed, E.M. & Schaufele, F. (2007) Functional Sequestration of Transcription Factor Activity by Repetitive DNA. *Journal of Biological Chemistry*. 282 (29), 20868–20876. doi:10.1074/jbc.M702547200.
- Lo, T.W., Choi, H.K.J., Huang, D. & Wiggins, P.A. (2023) Noise robustness and metabolic load determine the principles of central dogma regulation. *bioRxiv*. 2023.10.20.563172. doi:10.1101/2023.10.20.563172.
- Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. (2010) Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnology*. 10 (1), 81. doi:10.1186/1472-6750-10-81.
- MacCarthy, C.M., Wu, G., Malik, V., Menuchin-Lasowski, Y., Velychko, T., Keshet, G., Fan, R., Bedzhov, I., Church, G.M., Jauch, R., Cojocar, V., Schöler, H.R. & Velychko, S. (2024) Highly cooperative chimeric super-SOX induces naive pluripotency across species. *Cell Stem Cell*. 31 (1), 127-147.e9. doi:10.1016/j.stem.2023.11.010.
- Macfarlane, N.B.W., Adams, J., Bennett, E.L., Brooks, T.M., Delborne, J.A., Eggermont, H., Endy, D., Esvelt, K.M., Kolodziejczyk, B., Kuiken, T., Oliva, M.J., Peña Moreno, S., Slobodian,

- L., Smith, R.B., Thizy, D., Tompkins, D.M., Wei, W. & Redford, K.H. (2022) Direct and indirect impacts of synthetic biology on biodiversity conservation. *iScience*. 25 (11), 105423. doi:10.1016/j.isci.2022.105423.
- Makri Pistikou, A.-M., Cremers, G.A.O., Nathalia, B.L., Meuleman, T.J., Bögels, B.W.A., Eijkens, B. V., de Dreu, A., Bezembinder, M.T.H., Stassen, O.M.J.A., Bouten, C.C. V., Merks, M., Jerala, R. & de Greef, T.F.A. (2023) Engineering a scalable and orthogonal platform for synthetic communication in mammalian cells. *Nature Communications*. 14 (1), 7001. doi:10.1038/s41467-023-42810-5.
- Manhas, J., Edelstein, H.I., Leonard, J.N. & Morsut, L. (2022) The evolution of synthetic receptor systems. *Nature Chemical Biology*. 18 (3), 244–255. doi:10.1038/s41589-021-00926-z.
- Mansouri, M. & Fussenegger, M. (2022) Therapeutic cell engineering: designing programmable synthetic genetic circuits in mammalian cells. *Protein & Cell*. 13 (7), 476–489. doi:10.1007/s13238-021-00876-1.
- Marschall, A.L.J. (2021) Targeting the Inside of Cells with Biologicals: Chemicals as a Delivery Strategy. *BioDrugs*. 35 (6), 643–671. doi:10.1007/s40259-021-00500-y.
- Martinez, M. & Moon, E.K. (2019) CAR T Cells for Solid Tumors: New Strategies for Finding, Infiltrating, and Surviving in the Tumor Microenvironment. *Frontiers in Immunology*. 10. doi:10.3389/fimmu.2019.00128.
- Martinez, N.J. & Walhout, A.J.M. (2009) The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *BioEssays*. 31 (4), 435–445. doi:10.1002/bies.200800212.
- Mathur, M., Xiang, J.S. & Smolke, C.D. (2017) Mammalian synthetic biology for studying the cell. *Journal of Cell Biology*. 216 (1), 73–82. doi:10.1083/jcb.201611002.
- Matsui, M., Chu, Y., Zhang, H., Gagnon, K.T., Shaikh, S., Kuchimanchi, S., Manoharan, M., Corey, D.R. & Janowski, B.A. (2013) Promoter RNA links transcriptional regulation of inflammatory pathway genes. *Nucleic Acids Research*. 41 (22), 10086–10109. doi:10.1093/nar/gkt777.
- Matsuura, S., Ono, H., Kawasaki, S., Kuang, Y., Fujita, Y. & Saito, H. (2018) Synthetic RNA-based logic computation in mammalian cells. *Nature Communications*. 9 (1), 4847. doi:10.1038/s41467-018-07181-2.
- McFadden, J. (2023) Razor sharp: The role of Occam's razor in science. *Annals of the New York Academy of Sciences*. doi:10.1111/nyas.15086.
- Michaels, Y.S., Barnkob, M.B., Barbosa, H., Baeumler, T.A., Thompson, M.K., Andre, V., Colin-York, H., Fritzsche, M., Gileadi, U., Sheppard, H.M., Knapp, D.J.H.F., Milne, T.A., Cerundolo, V.

- & Fulga, T.A. (2019) Precise tuning of gene expression levels in mammalian cells. *Nature Communications*. 10 (1), 818. doi:10.1038/s41467-019-08777-y.
- Mohammadi, P., Beerenwinkel, N. & Benenson, Y. (2017) Automated Design of Synthetic Cell Classifier Circuits Using a Two-Step Optimization Strategy. *Cell Systems*. 4 (2), 207-218.e14. doi:10.1016/j.cels.2017.01.003.
- Morishita, R., Higaki, J., Tomita, N. & Ogihara, T. (1998) Application of Transcription Factor “Decoy” Strategy as Means of Gene Therapy and Study of Gene Expression in Cardiovascular Disease. *Circulation Research*. 82 (10), 1023–1028. doi:10.1161/01.RES.82.10.1023.
- Morrison, M.S., Podracky, C.J. & Liu, D.R. (2020) The developing toolkit of continuous directed evolution. *Nature Chemical Biology*. 16 (6), 610–619. doi:10.1038/s41589-020-0532-y.
- Morse, N.J. & Alper, H.S. (2016) Gene Expression Engineering. In: *Metabolic Engineering for Bioprocess Commercialization*. Cham, Springer International Publishing. pp. 7–30. doi:10.1007/978-3-319-41966-4_2.
- Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A. & van Oudenaarden, A. (2011) MicroRNAs can generate thresholds in target gene expression. *Nature Genetics*. 43 (9), 854–859. doi:10.1038/ng.905.
- Mustafina, K., Fukunaga, K. & Yokobayashi, Y. (2020) Design of Mammalian ON-Riboswitches Based on Tandemly Fused Aptamer and Ribozyme. *ACS Synthetic Biology*. 9 (1), 19–25. doi:10.1021/acssynbio.9b00371.
- Nagy, A. (2000) Cre recombinase: The universal reagent for genome tailoring. *genesis*. 26 (2), 99–109. doi:10.1002/(SICI)1526-968X(200002)26:2<99::AID-GENE1>3.0.CO;2-B.
- Nair, S., Ameen, M., Sundaram, L., Pampari, A., Schreiber, J., Balsubramani, A., Wang, Y.X., Burns, D., Blau, H.M., Karakikes, I., Wang, K.C. & Kundaje, A. (2023) Transcription factor stoichiometry, motif affinity and syntax regulate single-cell chromatin dynamics during fibroblast reprogramming to pluripotency. *bioRxiv*. 2023.10.04.560808. doi:10.1101/2023.10.04.560808.
- Nam, J.W., Rissland, O.S., Koppstein, D., Abreu-Goodger, C., Jan, C.H., Agarwal, V., Yildirim, M.A., Rodriguez, A. & Bartel, D.P. (2014) Global Analyses of the Effect of Different Cellular Contexts on MicroRNA Targeting. *Molecular Cell*. 53 (6), 1031–1043. doi:10.1016/J.MOLCEL.2014.02.013.
- Nambiar, T.S., Baudrier, L., Billon, P. & Ciccio, A. (2022) CRISPR-based genome editing through the lens of DNA repair. *Molecular Cell*. 82 (2), 348–388. doi:10.1016/j.molcel.2021.12.026.
- Nanopore (n.d.) *Nanopore DNA sequencing*. <https://nanoporetech.com/applications/dna-nanopore-sequencing> [Accessed: 13 January 2024].

- Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J.K. & Wysocka, J. (2023) Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nature Genetics*. 55 (5), 841–851. doi:10.1038/s41588-023-01366-2.
- Nevozhay, D., Zal, T. & Balázsi, G. (2013) Transferring a synthetic gene circuit from yeast to mammalian cells. *Nature Communications*. 4 (1), 1451. doi:10.1038/ncomms2471.
- Normanno, D., Boudarène, L., Dugast-Darzacq, C., Chen, J., Richter, C., Proux, F., Bénichou, O., Voituriez, R., Darzacq, X. & Dahan, M. (2015) Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nature Communications*. 6 (1), 7357. doi:10.1038/ncomms8357.
- Noskov, V.N., Lee, N.C.O., Larionov, V. & Kouprina, N. (2011) Rapid generation of long tandem DNA repeat arrays by homologous recombination in yeast to study their function in mammalian genomes. *Biological Procedures Online*. 13 (1). doi:10.1186/1480-9222-13-8.
- O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. (2018) Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*. 9. doi:10.3389/fendo.2018.00402.
- Oliynyk, R.T. & Church, G.M. (2022) Efficient modification and preparation of circular DNA for expression in cell culture. *Communications Biology*. 5 (1), 1393. doi:10.1038/s42003-022-04363-z.
- Orlandini von Niessen, A.G., Poleganov, M.A., Rechner, C., Plaschke, A., Kranz, L.M., Fesser, S., Diken, M., Löwer, M., Vallazza, B., Beissert, T., Bukur, V., Kuhn, A.N., Türeci, Ö. & Sahin, U. (2019) Improving mRNA-Based Therapeutic Gene Delivery by Expression-Augmenting 3' UTRs Identified by Cellular Library Screening. *Molecular Therapy*. 27 (4), 824–836. doi:10.1016/j.ymthe.2018.12.011.
- Ortega, A.D. (2023) Real-Time Assessment of Intracellular Metabolites in Single Cells through RNA-Based Sensors. *Biomolecules*. 13 (5), 765. doi:10.3390/biom13050765.
- Paulson, H. (2018) Repeat expansion diseases. *Handbook of clinical neurology*. 147, 105–123. doi:10.1016/B978-0-444-63233-3.00009-9.
- Perl, K., Ushakov, K., Pozniak, Y., Yizhar-Barnea, O., Bhonker, Y., Shivatzki, S., Geiger, T., Avraham, K.B. & Shamir, R. (2017) Reduced changes in protein compared to mRNA levels across non-proliferating tissues. *BMC Genomics*. 18 (1), 305. doi:10.1186/s12864-017-3683-9.
- Pollak, B., Cerda, A., Delmans, M., Álamos, S., Moyano, T., West, A., Gutiérrez, R.A., Patron, N.J., Federici, F. & Haseloff, J. (2019) Loop assembly: a simple and open system for recursive fabrication of <scp>DNA</scp> circuits. *New Phytologist*. 222 (1), 628–640. doi:10.1111/nph.15625.

- Pop, R.T., Pisante, A., Nagy, D., Martin, P.C.N., Mikheeva, L.A., Hayat, A., Ficz, G. & Zabet, N.R. (2023) Identification of mammalian transcription factors that bind to inaccessible chromatin. *Nucleic Acids Research*. 51 (16), 8480–8495. doi:10.1093/nar/gkad614.
- Porello, I. & Cellesi, F. (2023) Intracellular delivery of therapeutic proteins. New advancements and future directions. *Frontiers in Bioengineering and Biotechnology*. 11. doi:10.3389/fbioe.2023.1211798.
- Potvin-Trottier, L., Lord, N.D., Vinnicombe, G. & Paulsson, J. (2016) Synchronous long-term oscillations in a synthetic gene circuit. *Nature*. 538 (7626), 514–517. doi:10.1038/nature19841.
- Pratt, H.E., Andrews, G.R., Phalke, N., Huey, J.D., Purcaro, M.J., van der Velde, A., Moore, J.E. & Weng, Z. (2022) Factorbook: an updated catalog of transcription factor motifs and candidate regulatory motif sites. *Nucleic Acids Research*. 50 (D1), D141–D149. doi:10.1093/nar/gkab1039.
- Pryor, J.M., Potapov, V., Kucera, R.B., Bilotti, K., Cantor, E.J. & Lohman, G.J.S. (2020) Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLOS ONE*. 15 (9), e0238592. doi:10.1371/journal.pone.0238592.
- Puerta-Fernández, E., Romero-López, C., Barroso-delJesus, A. & Berzal-Herranz, A. (2003) Ribozymes: recent advances in the development of RNA tools. *FEMS Microbiology Reviews*. 27 (1), 75–97. doi:10.1016/S0168-6445(03)00020-2.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. & Lim, W.A. (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*. 152 (5), 1173–1183. doi:10.1016/j.cell.2013.02.022.
- Qiu, S. (2005) A computational study of off-target effects of RNA interference. *Nucleic Acids Research*. 33 (6), 1834–1847. doi:10.1093/nar/gki324.
- Qu, H., Zheng, L., Pu, J., Mei, H., Xiang, X., Zhao, X., Li, D., Li, S., Mao, L., Huang, K. & Tong, Q. (2015) miRNA-558 promotes tumorigenesis and aggressiveness of neuroblastoma cells through activating the transcription of heparanase. *Human Molecular Genetics*. 24 (9), 2539–2551. doi:10.1093/hmg/ddv018.
- Quarton, T., Ehrhardt, K., Lee, J., Kannan, S., Li, Y., Ma, L. & Bleris, L. (2018) Mapping the operational landscape of microRNAs in synthetic gene circuits. *npj Systems Biology and Applications*. 4 (1), 6. doi:10.1038/s41540-017-0043-y.
- Ramazi, S. & Zahiri, J. (2021) Post-translational modifications in proteins: resources, tools and prediction methods. *Database*. 2021. doi:10.1093/database/baab012.
- Rando, O.J. & Verstrepen, K.J. (2007) Timescales of Genetic and Epigenetic Inheritance. *Cell*. 128 (4), 655–668. doi:10.1016/j.cell.2007.01.023.

De Roeck, A., De Coster, W., Bossaerts, L., Cacace, R., De Pooter, T., Van Dongen, J., D'Hert, S., De Rijk, P., Strazisar, M., Van Broeckhoven, C. & Slegers, K. (2019) NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biology*. 20 (1), 239. doi:10.1186/s13059-019-1856-3.

Roig-Merino, A., Urban, M., Bozza, M., Peterson, J.D., Bullen, L., Büchler-Schäff, M., Stäble, S., van der Hoeven, F., Müller-Decker, K., McKay, T.R., Milsom, M.D. & Harbottle, R.P. (2022) An episomal DNA vector platform for the persistent genetic modification of pluripotent stem cells and their differentiated progeny. *Stem Cell Reports*. 17 (1), 143–158. doi:10.1016/j.stemcr.2021.11.011.

Sahin, U., Karikó, K. & Türeci, Ö. (2014) mRNA-based therapeutics — developing a new class of drugs. *Nature Reviews Drug Discovery*. 13 (10), 759–780. doi:10.1038/nrd4278.

Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I. & Overington, J.P. (2017) A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*. 16 (1), 19–34. doi:10.1038/nrd.2016.230.

Sasaki, A. & Kinjo, M. (2010) Monitoring intracellular degradation of exogenous DNA using diffusion properties. *Journal of Controlled Release*. 143 (1), 104–111. doi:10.1016/j.jconrel.2009.12.013.

Saunders, T.L. (2011) *Inducible Transgenic Mouse Models*. In: pp. 103–115. doi:10.1007/978-1-60761-974-1_7.

Saxena, P., Bojar, D. & Fussenegger, M. (2017) *Design of Synthetic Promoters for Gene Circuits in Mammalian Cells*. In: pp. 263–273. doi:10.1007/978-1-4939-7223-4_19.

Sayeg, M.K., Weinberg, B.H., Cha, S.S., Goodloe, M., Wong, W.W. & Han, X. (2015) Rationally Designed MicroRNA-Based Genetic Classifiers Target Specific Neurons in the Brain. *ACS Synthetic Biology*. 4 (7), 788–795. doi:10.1021/acssynbio.5b00040.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. & Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature*. 473 (7347), 337–342. doi:10.1038/nature10098.

Scior, A., Preissler, S., Koch, M. & Deuerling, E. (2011) Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnology*. 11 (1), 87. doi:10.1186/1472-6750-11-87.

Scully, R., Panday, A., Elango, R. & Willis, N.A. (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology*. 20 (11), 698–714. doi:10.1038/s41580-019-0152-0.

- Sengoku, T., Shiina, M., Suzuki, K., Hamada, K., Sato, K., et al. (2022) Structural basis of transcription regulation by CNC family transcription factor, Nrf2. *Nucleic Acids Research*. 50 (21), 12543–12557. doi:10.1093/nar/gkac1102.
- Shen, C.K. & Hu, W.S. (1986) DNA supercoiling of recombinant plasmids in mammalian cells. *Proceedings of the National Academy of Sciences*. 83 (6), 1641–1645. doi:10.1073/pnas.83.6.1641.
- Spagnou, S., Miller, A.D. & Keller, M. (2004) Lipidic Carriers of siRNA: Differences in the Formulation, Cellular Uptake, and Delivery with Plasmid DNA. *Biochemistry*. 43 (42), 13348–13356. doi:10.1021/bi048950a.
- Sriram, K. & Insel, P.A. (2018) G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Molecular Pharmacology*. 93 (4), 251–258. doi:10.1124/mol.117.111062.
- Stanton, B.C., Siciliano, V., Ghodasara, A., Wroblewska, L., Clancy, K., Trefzer, A.C., Chesnut, J.D., Weiss, R. & Voigt, C.A. (2014) Systematic Transfer of Prokaryotic Sensors and Circuits to Mammalian Cells. *ACS Synthetic Biology*. 3 (12), 880–891. doi:10.1021/sb5002856.
- Szenk, M., Yim, T. & Balázsi, G. (2020) Multiplexed Gene Expression Tuning with Orthogonal Synthetic Gene Circuits. *ACS Synthetic Biology*. 9 (4), 930–939. doi:10.1021/acssynbio.9b00534.
- Tanguay, R.L. & Gallie, D.R. (1996) Translational Efficiency Is Regulated by the Length of the 3' Untranslated Region. *Molecular and Cellular Biology*. 16 (1), 146–156. doi:10.1128/MCB.16.1.146.
- Tellechea-Luzardo, J., Hobbs, L., Velázquez, E., Pelechova, L., Woods, S., de Lorenzo, V. & Krasnogor, N. (2022) Versioning biological cells for trustworthy cell engineering. *Nature Communications*. 13 (1), 765. doi:10.1038/s41467-022-28350-4.
- Thakore, P.I. & Gersbach, C.A. (2016) *Design, Assembly, and Characterization of TALE-Based Transcriptional Activators and Repressors*. In: pp. 71–88. doi:10.1007/978-1-4939-2932-0_7.
- The Council for Science and Technology (2023) *Report on engineering biology: opportunities for the UK economy and national goals*. 19 May 2023. UK Gov. <https://www.gov.uk/government/publications/advice-on-engineering-biology/report-on-engineering-biology-opportunities-for-the-uk-economy-and-national-goals-html#issues-and-recommendations> [Accessed: 23 November 2023].
- Tihanyi, B. & Nyitray, L. (2020) Recent advances in CHO cell line development for recombinant protein production. *Drug Discovery Today: Technologies*. 38, 25–34. doi:10.1016/j.ddtec.2021.02.003.

- Torres, M., Ortuzar, V., Dickson, A.J. & Hussain, H. (2021) *An Omic's Data-Driven Approach Towards Engineering Mammalian Cell Factories and Bioprocesses for Biopharmaceutical Production*. In: pp. 93–128. doi:10.1007/978-3-030-79871-0_4.
- Ung, T.H., Madsen, H.J., Hellwinkel, J.E., Lencioni, A.M. & Graner, M.W. (2014) Exosome proteomics reveals transcriptional regulator proteins with potential to mediate downstream pathways. *Cancer Science*. 105 (11), 1384–1392. doi:10.1111/cas.12534.
- Di Ventura, B. & Mootz, H.D. (2019) Switchable inteins for conditional protein splicing. *Biological Chemistry*. 400 (4), 467–475. doi:10.1515/hsz-2018-0309.
- Vicente-García, C., Hernández-Camacho, J.D. & Carvajal, J.J. (2022) Regulation of myogenic gene expression. *Experimental Cell Research*. 419 (1), 113299. doi:10.1016/j.yexcr.2022.113299.
- Visschers, V.H.M. & Siegrist, M. (2018) Differences in Risk Perception Between Hazards and Between Individuals. In: *Psychological Perspectives on Risk and Risk Analysis*. Cham, Springer International Publishing. pp. 63–80. doi:10.1007/978-3-319-92478-6_3.
- Vorontsov, I.E., Eliseeva, I.A., Zinkevich, A., Nikonov, M., Abramov, S., Boytsov, A., Kamenets, V., Kasianova, A., Kolmykov, S., Yevshin, I.S., Favorov, A., Medvedeva, Y.A., Jolma, A., Kolpakov, F., Makeev, V.J. & Kulakovskiy, I.V. (2024) HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Research*. 52 (D1), D154–D163. doi:10.1093/nar/gkad1077.
- Wan, X., Pinto, F., Yu, L. & Wang, B. (2020) Synthetic protein-binding DNA sponge as a tool to tune gene expression and mitigate protein toxicity. *Nature Communications*. 11 (1). doi:10.1038/s41467-020-19552-9.
- Wang, C., Tang, K., Li, Z., Chen, Z., Xu, H. & Ye, Z. (2016) Targeted p21WAF1/CIP1 activation by miR-1236 inhibits cell proliferation and correlates with favorable survival in renal cell carcinoma. *Urologic Oncology: Seminars and Original Investigations*. 34 (2), 59.e23-59.e34. doi:10.1016/J.UROLONC.2015.08.014.
- Wang, G., Kong, Q., Wang, D. & Asmi, F. (2023) Ethical and social insights into synthetic biology: predicting research fronts in the post-COVID-19 era. *Frontiers in Bioengineering and Biotechnology*. 11. doi:10.3389/fbioe.2023.1085797.
- Wang, H., La Russa, M. & Qi, L.S. (2016) CRISPR/Cas9 in Genome Editing and Beyond. *Annual Review of Biochemistry*. 85 (1), 227–264. doi:10.1146/annurev-biochem-060815-014607.
- Wang, H., Wang, L., Zhong, B. & Dai, Z. (2022a) Protein Splicing of Inteins: A Powerful Tool in Synthetic Biology. *Frontiers in Bioengineering and Biotechnology*. 10. doi:10.3389/fbioe.2022.810180.

- Wang, J., Lawry, S.T., Cohen, A.L. & Jia, S. (2014) Chromosome boundary elements and regulation of heterochromatin spreading. *Cellular and Molecular Life Sciences*. 71 (24), 4841–4852. doi:10.1007/s00018-014-1725-x.
- Wang, K., Escobar, M., Li, J., Mahata, B., Goell, J., Shah, S., Cluck, M. & Hilton, I.B. (2022b) Systematic comparison of CRISPR-based transcriptional activators uncovers gene-regulatory features of enhancer–promoter interactions. *Nucleic Acids Research*. 50 (14), 7842–7855. doi:10.1093/nar/gkac582.
- Wang, N.B., Beitz, A.M. & Galloway, K. (2020) Engineering cell fate: Applying synthetic biology to cellular reprogramming. *Current Opinion in Systems Biology*. 24, 18–31. doi:10.1016/j.coisb.2020.09.002.
- Wang, P., Zhou, Y. & Richards, A.M. (2021) Effective tools for RNA-derived therapeutics: siRNA interference or miRNA mimicry. *Theranostics*. 11 (18), 8771–8796. doi:10.7150/thno.62642.
- Wang, X.-W., Hu, L.-F., Hao, J., Liao, L.-Q., Chiu, Y.-T., Shi, M. & Wang, Y. (2019) A microRNA-inducible CRISPR–Cas9 platform serves as a microRNA sensor and cell-type-specific genome regulation tool. *Nature Cell Biology*. 21 (4), 522–530. doi:10.1038/s41556-019-0292-7.
- Waryah, C.B., Moses, C., Arooj, M. & Blancafort, P. (2018) *Zinc Fingers, TALEs, and CRISPR Systems: A Comparison of Tools for Epigenome Editing*. In: pp. 19–63. doi:10.1007/978-1-4939-7774-1_2.
- Watve, M.M., Dahanukar, N. & Watve, M.G. (2010) Sociobiological Control of Plasmid Copy Number in Bacteria. *PLoS ONE*. 5 (2), e9328. doi:10.1371/journal.pone.0009328.
- Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE*. 6 (2), e16765. doi:10.1371/journal.pone.0016765.
- Whitmarsh, A.J. & Davis*, R.J. (2000) Regulation of transcription factor function by phosphorylation. *Cellular and Molecular Life Sciences*. 57 (8), 1172–1183. doi:10.1007/PL00000757.
- Wong, C. (2023) UK first to approve CRISPR treatment for diseases: what you need to know. *Nature*. doi:10.1038/d41586-023-03590-6.
- World Health Organization (WHO) (n.d.) *Messenger RNA vaccines*. WHO. <https://www.who.int/teams/health-product-policy-and-standards/standards-and-specifications/vaccine-standardization/mrna-vaccines> [Accessed: 18 December 2023].
- Xiang, J.S., Kaplan, M., Dykstra, P., Hinks, M., McKeague, M. & Smolke, C.D. (2019) Massively parallel RNA device engineering in mammalian cells with RNA-Seq. *Nature Communications*. 10 (1), 4327. doi:10.1038/s41467-019-12334-y.

- Xu, C., Hu, S. & Chen, X. (2016) Artificial cells: from basic science to applications. *Materials Today*. 19 (9), 516–532. doi:10.1016/j.mattod.2016.02.020.
- Yao, F., Svensjö, T., Winkler, T., Lu, M., Eriksson, C. & Eriksson, E. (1998) Tetracycline Repressor, tetR, rather than the tetR–Mammalian Cell Transcription Factor Fusion Derivatives, Regulates Inducible Gene Expression in Mammalian Cells. *Human Gene Therapy*. 9 (13), 1939–1950. doi:10.1089/hum.1998.9.13-1939.
- Yarnall, M.T.N., Ioannidi, E.I., Schmitt-Ulms, C., Krajewski, R.N., Lim, J., et al. (2023) Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nature Biotechnology*. 41 (4), 500–512. doi:10.1038/s41587-022-01527-4.
- Yesbolatova, A., Saito, Y., Kitamoto, N., Makino-Itou, H., Ajima, R., Nakano, R., Nakaoka, H., Fukui, K., Gamo, K., Tominari, Y., Takeuchi, H., Saga, Y., Hayashi, K. & Kanemaki, M.T. (2020) The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nature Communications*. 11 (1), 5701. doi:10.1038/s41467-020-19532-z.
- Young, R., Haines, M., Storch, M. & Freemont, P.S. (2021) Combinatorial metabolic pathway assembly approaches and toolkits for modular assembly. *Metabolic Engineering*. 63, 81–101. doi:10.1016/j.ymben.2020.12.001.
- Yu, J., Zhang, L., Hwang, P.M., Rago, C., Kinzler, K.W. & Vogelstein, B. (1999) Identification and classification of p53-regulated genes. *Proceedings of the National Academy of Sciences*. 96 (25), 14517–14522. doi:10.1073/pnas.96.25.14517.
- Zang, J., Zhu, Y., Zhou, Y., Gu, C., Yi, Y., et al. (2021) Yeast-produced RBD-based recombinant protein vaccines elicit broadly neutralizing antibodies and durable protective immunity against SARS-CoV-2 infection. *Cell Discovery*. 7 (1), 71. doi:10.1038/s41421-021-00315-9.
- Zhang, R., Goetz, H., Melendez-Alvarez, J., Li, J., Ding, T., Wang, X. & Tian, X.-J. (2021) Winner-takes-all resource competition redirects cascading cell fate transitions. *Nature Communications*. 12 (1), 853. doi:10.1038/s41467-021-21125-3.
- Zhang, Y., Fan, M., Zhang, X., Huang, F., Wu, K., Zhang, J., Liu, J., Huang, Z., Luo, H., Tao, L. & Zhang, H. (2014) Cellular microRNAs up-regulate transcription via interaction with promoter TATA-box motifs. *RNA*. 20 (12), 1878–1889. doi:10.1261/rna.045633.114.
- Zhang, Z., Dey, S. & Singh, A. (2022) The impact of decoys on a genetic oscillator based on coupled positive-negative feedbacks. *IFAC-PapersOnLine*. 55 (40), 127–132. doi:10.1016/j.ifacol.2023.01.060.

Zhao, Y. (2023) TFSyntax: a database of transcription factors binding syntax in mammalian genomes. *Nucleic Acids Research*. 51 (D1), D306–D314. doi:10.1093/nar/gkac849.

Zhou, X., Vink, M., Klaver, B., Berkhout, B. & Das, A.T. (2006) Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Therapy*. 13 (19), 1382–1390. doi:10.1038/sj.gt.3302780.

Zhu, Y., Zhu, L., Wang, X. & Jin, H. (2022) RNA-based therapeutics: an overview and prospectus. *Cell Death & Disease*. 13 (7), 644. doi:10.1038/s41419-022-05075-2.

Zimak, J., Wagoner, Z.W., Nelson, N., Waechtler, B., Schlosser, H., Kopecky, M., Wu, J. & Zhao, W. (2021) Epigenetic silencing directs expression heterogeneity of stably integrated multi-transcript unit genetic circuits. *Scientific Reports*. 11 (1), 2424. doi:10.1038/s41598-021-81975-1.

Zoueiri, A., Di Blasi, R., Ellis, T. & Ceroni, F. (2021) Genetic Toolkits to Design and Build Mammalian Synthetic Systems. *Trends in Biotechnology*. 39 (10), 1004–1018. doi:10.1016/j.tibtech.2020.12.007.