Building to understand miRNA circuits

Thesis by
Michael Flynn

In Partial Fulfillment of the Requirements for the
Degree of
Ph.D. in Applied Physics

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended September 10, 2025

Michael Flynn
ORCID: 0009-0003-1186-957X

ACKNOWLEDGEMENTS

This has been a long 8-year journey, longer than I had anticipated. Without help from others I wouldn't have made it. First of all, thank you to my parents, James and Michelle, who followed my progress always with support and encouragement, always calling when I did not call home on the weekends. Thank you to my siblings, Matthew, Emma, Daniel, Genevieve, Elizabeth, Marie, Elyse, and Lily for always believing in me. Thank you to the friends I met along the way, providing welcome relief: Lauren Breeyear, Dylan Bannon, Eduardo Beltrame, James Ousey, Zhewei Chen, Tom Naragon, Aditi Narayanan, Ben Hoscheit, Aiden Aceves, Tatyana Dobreva, Ali Vogelaar, Sumit Goel, Fabian Wiesemuller, Megan Schill, and Ian McArthur. Thank you to the BMB Bombers softball team for being a welcome breath of fresh air during summers— many good memories. Thank you to the hard-working members of the graduate student and postdoc union including Nadia Suryawinata, Rohan Shenoy, Jasmine Emtage, Marina Lecouche, Rahma Elsiesy, and Noor Naji for standing up for the rights of other student workers, an important mission. Thank you to Rongrong Du and Kevin Luo for the fun late night conversations and the occasional hotpot.

Thank you to Xiaojing Gao, for training me in experimental methods in molecular biology, which I had never done before joining the Elowitz lab, and for including me in a rotation project. Thanks again to Rongrong Du for joining my miRNA projects and doing an incredible amount of work. Thank you to Yodai Takei, Martin Tran, Duncan Chadley, and James Linton for technical help in single molecule protocols, segmentation, and imaging. Thank you to Shang Wang, Yitong Ma, and Bo Gu for stimulating conversation. Thank you to all the other members of the Elowitz lab for providing a pleasant environment in which to work, and holding things together in a complex working environment. Thank you to Acacia Mayfield and Karan Mahe for conducting crucial mouse experiments during this work.

Thank you to Paul Bellan and Yi Zhou for a pleasant rotation in a plasma physics lab.

Thank you to Professor Daniel Aalberts for getting me started in biophysics. Thank you to Michael Elowitz for providing a supportive environment in which to complete these projects, and for believing in me, that I would achieve the ambitious goals that I had set out to accomplish.

ABSTRACT

MicroRNA (miRNA) regulation is ubiquitous in human biology, with miRNAs playing a role in every developmental process. Despite the fact that deletion of miRNA genes typically derepress their targets by only 20%-50%, such deletions are often lethal. However, this raises the question of how such modest derepression can lead to severe phenotypic consequences. To better understand miRNA regulation, I took a "build to understand" approach: by building synthetic biological circuits using miRNA in two engineering projects, I put models of miRNA regulation to the test and created biological devices with practical applications. First, I developed miRNA incoherent feedforward loop (IFFL) circuits that enable precise control of therapeutic transgene expression to augment Rett syndrome gene therapy. Second, my colleagues and I systematically varied miRNA target complementarity and cooperativity to generate a toolkit of modular IFFL circuits, termed DIMMERs, that enabled precise, tunable control of transgene expression across diverse cell types to facilitate imaging, editing, and gene therapy. Together, these projects provided evidence that canonically-sized miRNAs can repress gene expression by more than 10-fold in the presence of three or more co-repressing miRNAs, but achieve little repression individually. This challenges previous models of miRNAs as subtle fine-tuners of gene expression, which may have underestimated miRNA potency by focusing on individual targets rather than those of cooperative groups.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] M. J. Flynn et al. "Synthetic dosage-compensating miRNA circuits for quantitative gene therapy." *bioRxiv* (Mar. 2024). DOI: 10.1101/2024.03.13.584179. URL: http://dx.doi.org/10.1101/2024.03.13.584179.

Contributions: M.J.F. and A.M.H. contributed equally to this study. M.J.F. and M.B.E. conceived the project. M.J.F. identified the incoherent-feedforward-loop motif, performed mathematical modeling, and implemented the circuit using miRNA inside an AAV gene expression cassette. M.J.F. validated the circuit in cell culture using flow cytometry, sm-FISH, and HCR, wrote the image analysis pipeline, and compared endogenous and ectopic expression of MeCP2 in the mouse brain. M.B.E. provided guidance in circuit design and data analysis. A.M.M. and V.G. selected viral capsids and designed the mouse behavioral study. A.M.M. prepped and injected virus, and performed mouse tissue preparation. A.M.M established Rett model mouse colonies, and performed mouse behavioral studies, including time-course of behavioral phenotyping. R.D. performed mRNA sequencing in response to miRNA expression. M.J.F. and M.B.E. wrote the manuscript with input from all authors. M.B.E. supervised the in vitro work and V.G. supervised the in vivo work. M.B.E. and V.G. funded the project.

[2] R. Du et al. "miRNA circuit modules for precise, tunable control of gene expression." *bioRxiv* (2024). DOI: 10.1101/2024.03.12.583048. URL: https://doi.org/10.1101/2024.03.12.583048.

Contributions: R.D. and M.F. contributed equally to this study. R.D., M.F., and M.B.E. conceived and designed the study. M.B.E. directed and supervised the study. R.D., M.F., K.M., M.H., B.G., D.L., and S.E.M. performed or assisted with experiments and data analysis. R.D. and M.B.E. wrote the manuscript with input from all authors.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Overview

In his autobiography "Disturbing the Universe" Freeman Dyson said, "A good scientist is a person with original ideas. A good engineer is a person who makes a design that works with as few original ideas as possible." [1] Although Dyson framed this as a compliment to engineers, it implied that engineering involves less originality. I argue instead that while this distinction between scientists and engineers may hold in well-established engineering domains, it breaks down in fields where the underlying principles are still being discovered. In this context, engineering problems can serve as laboratories where a scientific theory is subjected to a predictive, empirical test. If the predictions fail, this forces the scientific theory to be revised, which is the goal of the scientific process.

This is the essential philosophy of the field of synthetic biology, which seeks to "build to understand" by reconstructing biological circuits in living cells using orthogonal components [2, 3]. In doing so, we often discover gaps in our understanding while simultaneously developing tools applicable to problems such as gene and cell therapy.

One area of biology where fundamental principles are still being uncovered is microRNA (miRNA) regulation. MicroRNAs are small RNAs produced from RNA hairpin loops that form inside introns of nascent transcripts [4]. These hairpins are cleaved out of the transcript, transported to the cytoplasm, and loaded into an Argonaute protein to guide the RNA-induced silencing complex (RISC) to mRNA targets [5, 6, 7, 8].

MicroRNA regulation via the RISC is mediated by two separate mechanisms. The first mechanism, called "slicing," is the direct cleavage of the mRNA by Argonaute, which happens only for targets that are fully complementary to the 21-nucleotide miRNA guide [9]. The second mechanism, the canonical regulation, acts through 6-8 nucleotide "seed" matches complementary to the 5' end of the miRNA [10, 11, 12]. Repression is mediated by the scaffolding protein TNRC6 linking Argonaute to the CCR4-NOT complex, which recruits translation inhibition factors [13] and deadenylates the mRNA, destabilizing it [14]. TNRC6 can bind to three Argonaute proteins simultaneously, providing a structural basis for cooperative miRNA regulation, where multiple bound Argonautes reinforce one another's binding to the mRNA target [15]. The great majority of miRNA target sites are regulated by the canonical pathway in humans, where only 1% of conserved miRNAs make use of non-canonical pathways [16].

MicroRNA regulation is ubiquitous and plays a role in virtually every developmental process [16]. Each of the 90 conserved miRNA families has greater than 300 7-8 nucleotide target sites under selective pressure to maintain pairing to the miRNA [17]. Approximately 58% of human mRNAs have conserved miRNA targets in their 3' untranslated region (UTR), and within this group, the average mRNA has four to five miRNA binding sites [17].

The effect of a miRNA on its target is thought to be subtle, with mRNA and protein expression only changing by 20%-50% upon adding or deleting miRNA genes [18, 19]. Despite this, the phenotypic consequences of miRNA deletion can be quite severe, with lethality often observed in mice [16]. This raises a central question: how do such subtle effects cause such severe phenotypic consequences?

One answer is that biological systems are highly sensitive to small but systematic changes in gene expression, despite being generally robust to perturbations from stochastic gene expression, heat shock, and other insults. In this thesis, I argue for another explanation: that the paradigm of miRNAs as subtle regulators overlooks the critical role of cooperativity in miRNA regulation, with strong repression emerging only for mRNA targeted by at least three miRNAs.

To provide evidence towards this thesis, I took a "build to understand" approach. In two engineering projects I constructed and optimized synthetic miRNA circuits and evaluated their behavior, while simultaneously creating devices with practical applications.

In the first project, I designed and tested synthetic miRNA incoherent feedforward loop (IFFL) circuits that can be used in Rett syndrome gene therapy. Rett syndrome is a severe neurodevelopmental disorder caused by mutations in MeCP2, a methyl-CpG binding protein essential for brain development [20, 21]. Rett syndrome gene therapy has a "Goldilocks problem": duplications of the MECP2 locus cause MECP2 Duplication syndrome, a different genetic disease [22], and previous attempts at gene therapy have all confirmed severe toxicity if MeCP2 is overexpressed [23, 24, 25, 26, 27]. To solve this problem I incorporated a synthetic miRNA-based IFFL circuit in an adeno-associated virus (AAV)-delivered Rett syndrome gene therapy. In this gene therapy, the viral vector expressed both a therapeutic copy of Mecp2 as well as a synthetic miRNA that down-regulated the expressed copy of MeCP2 via a target in the Mecp2 3' UTR. This circuit regulated ectopic Mecp2 expression to below the endogenous Mecp2 level even at a maximal viral dose in a mouse model of Rett syndrome, and led to an improvement in the behavioral Rett phenotype. This project demonstrated that synthetic miRNA IFFLs, previously validated only in vitro [28], could function in a living animal model to improve gene therapy, and that they function optimally when multimerized targets enable cooperative regulation.

In the second project, together with my colleague Rongrong Du, I varied the sequence, complementarity, and cooperativity of miRNAs and their targets in order to create a toolkit of tunable, orthogonal miRNA circuit modules. Wide-ranging applications such as intracellular imaging [29] and gene editing [30] benefit from miRNA IFFL modules that reduce background signal or activity. This project also demonstrated that miRNAs can mediate strong repression (>10-fold) on 8 nucleotide targets when acting cooperatively, but only weak repression when acting individually.

In a final chapter, I compare different models of miRNA regulation. The data from these projects is most consistent with a model of regulation in which there is little to no repression from a single miRNA acting alone on canonical seed targets, but strong repression emerges when at least three miRNAs cooperatively target the same mRNA. This provides evidence that the paradigm of the miRNA as a subtle regulator should shift to a model in which miRNAs act strongly through cooperative groups of at least three.

Before describing these projects and conclusion, below I give a concise primer on biological circuit models and derive a thermodynamic model for miRNA regulation.

1.2 Primer on biological circuits

Most physicists are familiar with electronic circuits, as they are a core part of standard undergraduate curricula and underpin virtually all scientific devices used today. Fewer are familiar with biological circuits, despite the fascinating potential of this field of study. Biological cells are capable of extraordinary feats of distributed computation, including the development of the body plan, the wiring of the brain, the micro-evolutionary selection of antibody genes, the immune response, and wound healing. All these feats are mediated by complex networks of biological circuits.

Just as electronic circuits are systems of interacting electrical components connected by metal conductors, biological circuits are systems of interacting biological molecules: DNA and RNA elements, transcription factors, RNA-binding proteins, proteases, kinases, and others interacting at several different levels [31]. These molecules are encoded by DNA, which is bound to by transcription factor proteins in promoter regions, which attracts RNA polymerase to bind and transcribe genes into mRNA, which is then translated by ribosomes into proteins. Proteins and mRNA are inherently unstable, degraded actively and stochastically by the proteasome and nucleases. All circuit behavior builds on top of these basic processes.

Modeling production and degradation

A simple model of biological transcription is therefore a gene transcribed at rate α mRNA molecules per minute, with each mRNA molecule degraded at a rate γ per minute. This produces an ordinary

differential equation for the number of mRNA molecules over time $m(t)$:

$$\frac{dm}{dt} = \alpha - \gamma m. \tag{1.1}$$

This differential equation has the solution

$$m(t) = \frac{\alpha}{\gamma}\left(1 - e^{-\gamma t}\right) \tag{1.2}$$

which converges exponentially to the steady-state concentration $m = \alpha/\gamma$ with time constant $1/\gamma$. This establishes the fundamental framing of all biological circuits, which exist in the space between production and degradation.

Modeling transcription activation and repression using kinetics

Transcription rates can be modulated by activator and repressor proteins. For example, a gene may only transcribe at the rate α when an activating transcription factor protein A is bound. In this model, the gene can be in one of two states: active (G_{on}) or inactive (G_{off}). Binding and unbinding of a transcription factor at concentration A causes G_{off} to switch on at rate $k_{on}A$ and G_{on} to switch off at rate k_{off}. This can be modeled by the following system of ODEs:

$$\frac{dG_{off}}{dt} = -k_{on}AG_{off} + k_{off}G_{on}, \tag{1.3}$$

$$\frac{dG_{on}}{dt} = k_{on}AG_{off} - k_{off}G_{on}, \tag{1.4}$$

$$\frac{dm}{dt} = \alpha G_{on} - \gamma m. \tag{1.5}$$

Because binding and unbinding are typically much faster than transcription or degradation, we assume quasi-equilibrium, $dG_{off}/dt = dG_{on}/dt = 0$. This yields

$$\frac{G_{on}}{G_{off}} = \frac{Ak_{on}}{k_{off}} = \frac{A}{K_D} \tag{1.6}$$

where $K_D = k_{off}/k_{on}$ is the dissociation constant. Using $G_{tot} = G_{on} + G_{off}$, we solve for G_{on} in terms of G_{tot}:

$$G_{on} = G_{tot}\frac{A/K_D}{1 + A/K_D}. \tag{1.7}$$

Substituting into the mRNA equation gives:

$$\frac{dm}{dt} = \alpha G_{tot}\frac{A/K_D}{1 + A/K_D} - \gamma m. \tag{1.8}$$

Thus the production rate is now a fraction of the original "constitutive-on" circuit, with the fraction controlled by the concentration of the activator A, in a saturating-linear relationship.

A similar derivation can be made for a transcriptional repressor, where the gene is active only when the repressor R is unbound, so that

$$G_{on} = G_{tot} \frac{1}{1 + R/K_D} \tag{1.9}$$

and the differential equation for $m(t)$ becomes:

$$\frac{dm}{dt} = \alpha G_{tot} \frac{1}{1 + R/K_D} - \gamma m. \tag{1.10}$$

Assuming the binding of R and A are independent, the two situations can be combined (using K_A for the K_D of A and K_R for the K_D of R):

$$\frac{dm}{dt} = \alpha G_{tot} \frac{A/K_A}{1 + A/K_A} \frac{1}{1 + R/K_R} - \gamma m. \tag{1.11}$$

This equation describes a model in which transcription occurs at rate α only when the activator is bound and the repressor is unbound.

Modeling transcription activation and repression using thermodynamics

An alternative means of deriving the transcription rate is using statistical mechanics. The quasi-equilibrium assumption is equivalent to assuming that the system can be described as being in thermal equilibrium. At thermal equilibrium, the probability p_i that a system is in state i is given by the Boltzmann distribution, namely that $p_i = \frac{1}{Z} W_i e^{-E_i/kT}$ where E_i is the energy of that state, W_i is that state's multiplicity, and Z is the partition function, $Z = \sum_i W_i e^{-E_i/kT}$. The mean transcription rate is then the expected value of α over the equilibrium distribution.

$$E[\alpha] = \frac{\sum_i \alpha_i W_i e^{-E_i/kT}}{\sum_i W_i e^{-E_i/kT}}. \tag{1.12}$$

To evaluate the expectation, it is convenient to construct a "states and weights" table that enumerates the possible gene states and defines their associated transcription rates, multiplicities, and Boltzmann factors. For the current system:

- there are four states of the gene: unbound, activator bound, repressor bound, and both bound;
- transcription only occurs at rate α when only the activator is bound, and zero otherwise;
- for the multiplicity of the states, each unbound activator or repressor molecule has $\rho_w V$ positions which it can occupy, where ρ_w is the concentration of water ($\sim$ 55 M), which corresponds to the fundamental grid size of this model, and V is the volume of the cell. Thus the multiplicity of the unbound state is $(\rho_w V)^{AV+RV}$ where A and R are the concentrations

of the activator and repressor molecules, respectively. For the "activator-bound" state, there are AV molecules to choose from, and only $AV - 1$ molecules left in the unbound state, and so its multiplicity is $AV(\rho_w V)^{AV+RV-1}$. Likewise the repressor-bound multiplicity is $RV(\rho_w V)^{AV+RV-1}$, and the dual bound multiplicity is $(AV)(RV)(\rho_w V)^{AV+RV-2}$. The multiplicities simplify, after normalization to the unbound state, to 1, A/ρ_w, R/ρ_w, and AR/ρ_w^2.

- For the energy, setting the energy of the unbound state as $E = 0$ makes the Boltzmann factor for the unbound state equal to 1, and $e^{-\Delta E_A/kT}$ and $e^{-\Delta E_R/kT}$ are the activator and repressor Boltzmann factors, respectively, where ΔE_A is the binding energy of the activator and ΔE_R the binding energy of the repressor.

The resulting states, rates, multiplicities, and Boltzmann factors are summarized in the following table:

State	Rate	Multiplicity	Boltzmann factor
Unbound	0	1	1
Activator Bound	α	A/ρ_w	$e^{-\Delta E_A/kT}$
Repressor Bound	0	R/ρ_w	$e^{-\Delta E_R/kT}$
Both Bound	0	AR/ρ_w^2	$e^{-(\Delta E_A+\Delta E_R)/kT}$

Evaluating the expected transcription rate over the four states:

$$\begin{aligned} E[\alpha] &= \frac{\alpha \frac{A}{\rho_w} e^{-\Delta E_A/kT}}{1 + \frac{A}{\rho_w} e^{-\Delta E_A/kT} + \frac{R}{\rho_w} e^{-\Delta E_R/kT} + \frac{AR}{\rho_w^2} e^{-(\Delta E_A+\Delta E_R)/kT}} \\ &= \alpha \frac{\frac{A}{\rho_w} e^{-\Delta E_A/kT}}{1 + \frac{A}{\rho_w} e^{-\Delta E_A/kT}} \frac{1}{1 + \frac{R}{\rho_w} e^{-\Delta E_R/kT}}. \end{aligned} \tag{1.13}$$

This expression matches the form derived from the ODE approach (with $G_{tot} = 1$), namely:

$$\alpha \frac{A/K_A}{1 + A/K_A} \frac{1}{1 + R/K_R} \tag{1.14}$$

with

$$K_A = \rho_w e^{\Delta E_A/kT}, \tag{1.15}$$

$$K_R = \rho_w e^{\Delta E_R/kT}. \tag{1.16}$$

Thus both kinetic and thermodynamic derivations yield the same functional form for the production rate. In both these cases, activators and repressors control the transcription rate in a fractional, saturating-linear function.

Oscillators

With transcriptional regulation established, we can now consider circuits of multiple genes that regulate each other to generate diverse dynamic behaviors. One of the first synthetic circuits to appear was the repressilator [32], an oscillator composed of 3 genes in a ring, where each gene represses the next. Sustained oscillations require an effective Hill coefficient greater than one, which can be achieved by using homodimeric repressors that bind proportionally to $(R/K_R)^2$ instead of just (R/K_R).

A simple model of this three-gene ring can be expressed by the system of equations below. For simplicity, all dissociation constants and degradation rates are assumed to be equal ($K_A = K_B = K_C = K$ and $\gamma_A = \gamma_B = \gamma_C = \gamma$).

$$\frac{dA}{dt} = \frac{\alpha}{1 + \left(\frac{C}{K}\right)^2} - \gamma A, \tag{1.17}$$

$$\frac{dB}{dt} = \frac{\alpha}{1 + \left(\frac{A}{K}\right)^2} - \gamma B, \tag{1.18}$$

$$\frac{dC}{dt} = \frac{\alpha}{1 + \left(\frac{B}{K}\right)^2} - \gamma C. \tag{1.19}$$

To find oscillations, we can examine how the system responds to small perturbations from a steady state. Let $\mathbf{x} = (A, B, C)$ denote the vector of gene expression levels, $\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$ describe the system of equations, and $\mathbf{x_0} = (A_0, B_0, C_0)$ denote the steady state solution satisfying $\frac{d\mathbf{x}_0}{dt} = 0$. By symmetry it must be that $A_0 = B_0 = C_0 = \kappa$ where κ satisfies $\gamma\kappa = \alpha/(1 + (\kappa/K)^2)$. The response to infinitesimal deviation from the steady state $\mathbf{x} = \mathbf{x}_0 + \delta\mathbf{x}$ evolves according to the linearized system

$$\frac{d}{dt}(\delta\mathbf{x}) = \frac{df}{d\mathbf{x}}\Big|_{\mathbf{x}_0} \cdot \delta\mathbf{x}. \tag{1.20}$$

Oscillatory behavior arises when this linear equation has complex eigenvalues with positive real part.

To find the eigenvalues, first the matrix can be computed from the original equations:

$$\frac{df}{d\mathbf{x}}\Big|_{\mathbf{x}_0} = \begin{pmatrix} -\gamma & 0 & -\beta \\ -\beta & -\gamma & 0 \\ 0 & -\beta & -\gamma \end{pmatrix}, \quad \beta = \frac{2\alpha\frac{\kappa}{K^2}}{\left(1 + \left(\frac{\kappa}{K}\right)^2\right)^2}. \tag{1.21}$$

This matrix has eigenvalues equal to $\lambda = -\gamma + \beta\omega$ where $\omega \in \{-1, e^{i\pi/3}, e^{-i\pi/3}\}$. The roots with the largest real part correspond to the $e^{\pm i\pi/3}$ solutions. Since those eigenvalues have imaginary

parts, the solution is oscillatory. Oscillations grow when the real part is positive: $-\gamma + \frac{\beta}{2}$. This criterion is equivalent to $\frac{\beta}{2} > \gamma$, which can be satisfied provided α is large enough.

Incoherent feedforward loops

One method of dissecting the complex natural circuitry of living cells is by breaking large networks into smaller, 3-4 node networks interpreted as elementary "circuit motifs" whose essential topologies are replicated across several systems as modules [33, 34]. One of the most important of the 3-node networks is the incoherent feedforward loop (IFFL), which consists of an input node driving both an output node and a negative regulator of that output node [35].

Let I denote the input, R the repressor, and O the output. All IFFLs are described by an ODE of the form

$$\frac{dR}{dt} = \alpha I - \gamma_R R, \tag{1.22}$$

$$\frac{dO}{dt} = \beta I - \gamma_O O - kOR, \tag{1.23}$$

where k is a second-order rate constant capturing the repression strength of R on O (such as k_{cat}/K_M for non-saturated Michaelis Menten kinetics). The steady state solution has the relatively simple form:

$$R_0 = \frac{\alpha I}{\gamma_R}, \tag{1.24}$$

$$O_0 = \beta \frac{\gamma_R}{k\alpha} \frac{(I/\gamma_O)}{\frac{\gamma_R}{k\alpha} + (I/\gamma_O)}. \tag{1.25}$$

The essential property of the IFFL at steady state is the fact that the expression for O_0 is an asymptotic function in I which approaches a maximum of $\beta\gamma_R/(k\alpha)$. This means that there is a large "flat region," a range of inputs for which the output changes very little. In this thesis, I make use of this property in critical circuits for controlling gene expression in the context of noisy delivery mechanisms such as viral delivery and transient transfection.

Dynamically, the IFFL can act as a pulse or difference detector [36]. Because of the delay on the repression arm of the network, fast pulses or changes are transmitted through the IFFL before the circuit returns to a steady state. These properties of the IFFL are important in living systems for tasks such as chemotaxis (directed movement up/down chemical gradients), where a biological organism seeks to react to increases or decreases in chemoattractant concentration in order to move towards a higher concentration source.

Other circuits

While not inside the scope of this primer, there are a plethora of additional synthetic circuit types that have been demonstrated: bistable switches [37], multi-stable state systems [38], many-to-many

computation networks [39], spatial patterning [40], lineage tracers [41] and others.

1.3 Models of microRNA regulation

Now that basic methods for modeling biological circuits have been established, we can use these tools to address ambiguities in miRNA regulation. As discussed previously, the standard picture of miRNA regulation consists of individual miRNAs acting on individual targets and modestly repressing their expression by 20% to 50%. However, it is also acknowledged that the TNRC6 scaffolding domain allows for cooperativity in regulation, but the consequences remain poorly understood.

The behavior of the miRNA circuits implemented in the projects presented here will differ qualitatively depending on which model reflects the true mechanism. In this section, I present different models of miRNA regulation to be compared against the empirical data from these projects. In the final chapter, I evaluate which model best fits the data.

To highlight cooperativity, all the models I consider here focus on the behavior of miRNA regulation in the presence of three miRNAs (R_1, R_2, R_3) acting on independent sites of the same target. As previously described, miRNA regulation is mediated by two separate pathways: (1) slicing the mRNA, and (2) TNRC6-mediated recruitment of deadenylation and translational repression factors. I begin with the simpler slicing pathway before considering the more complex TNRC6-mediated regulation.

Slicing models are additive

In the slicing mode of regulation, miRNA-loaded Argonaute can be modeled as an enzyme that catalytically degrades mRNA according to Michaelis-Menten kinetics (with a k_{cat} and a K_M). Since binding and cleavage occur independently for each of the miRNAs, the degradation rate is the sum of their individual contributions:

$$\text{rate} = \frac{k_{cat,1} R_1}{K_{M,1} + R_1} + \frac{k_{cat,2} R_2}{K_{M,2} + R_2} + \frac{k_{cat,3} R_3}{K_{M,3} + R_3}. \tag{1.26}$$

If each of the R_i has the same k_{cat} and K_M, and is expressed far below the K_M of the reaction, then the concentrations directly add:

$$\text{rate} \approx \frac{k_{cat}}{K_M}(R_1 + R_2 + R_3). \tag{1.27}$$

Thus, for the slicing mode of regulation, the effects of multiple miRNAs combine additively.

TNRC6-based models display different forms of cooperativity

For TNRC6-mediated repression, TNRC6 binds indirectly to mRNAs via binding to up to three Argonaute (Ago) proteins that are bound to the mRNA. Degradation occurs through the recruitment

of the CCR4-NOT deadenylase by TNRC6. Unlike slicing, TNRC6-mediated repression can be expected to have inherent nonlinearities because TNRC6 binding on one mRNA-bound Argonaute can be greatly stabilized by the binding of additional Argonautes.

This complicated system is best modeled using a thermodynamic approach to analyze the assembly of up to 3 Argonaute proteins and 1 TNRC6 on a single mRNA. As before, the overall degradation rate is the Boltzmann-weighted average over all possible microstates:

$$\text{Degradation rate} = \frac{\sum_i \text{rate}_i W_i e^{-\Delta E_i/kT}}{\sum_i W_i e^{-\Delta E_i/kT}}. \tag{1.28}$$

In this model, I make several assumptions:

- possible states include: (i) the fully unbound state, (ii) 1-3 Argonaute proteins bound without TNRC6, (iii) 1-3 Argonautes bound with TNRC6;
- TNRC6 has total concentration T, the Argonautes are at the same concentration $R = R_1 = R_2 = R_3$;
- each Argonaute binds its target mRNA site with the binding energy ΔE_R;
- each Argonaute binds to TNRC6 with binding energy ΔE_T;
- the miRNA-driven degradation rate of mRNA is k_{deg} when TNRC6 is bound, 0 when unbound;
- as before, the binding of a free factor appends a term equal to concentration of that factor divided by ρ_w to the (normalized) multiplicity.

With these assumptions, the different states, their degradation rate, their multiplicity, and Boltzmann weight (normalized to the unbound state) are enumerated in the table below:

State	Rate	Multiplicity	Boltzmann Weight
Unbound	0	1	1
1 Ago Bound, TNRC6 Unbound	0	$3R/\rho_w$	$e^{-\Delta E_R/kT}$
2 Ago Bound, TNRC6 Unbound	0	$3R^2/\rho_w^2$	$e^{-2\Delta E_R/kT}$
3 Ago Bound, TNRC6 Unbound	0	R^3/ρ_w^3	$e^{-3\Delta E_R/kT}$
1 Ago Bound, TNRC6 Bound	k_{deg}	$3(T/\rho_w)R/\rho_w$	$e^{-(\Delta E_R+\Delta E_T)/kT}$
2 Ago Bound, TNRC6 Bound	k_{deg}	$3(T/\rho_w)R^2/\rho_w^2$	$e^{-2(\Delta E_R+\Delta E_T)/kT}$
3 Ago Bound, TNRC6 Bound	k_{deg}	$(T/\rho_w)R^3/\rho_w^3$	$e^{-3(\Delta E_R+\Delta E_T)/kT}$

Let $r = \frac{R}{\rho_w} e^{-\Delta E_R/kT}$, $\tau = \frac{T}{\rho_w} e^{-\Delta E_T/kT}$. The expected value can now be computed by summing the terms:

$$E[\text{rate}] = k_{deg}\tau \frac{3r + 3r^2 e^{-\Delta E_T/kT} + r^3 e^{-2\Delta E_T/kT}}{1 + 3r + 3r^2 + r^3 + \tau\left[3r + 3r^2 e^{-\Delta E_T/kT} + r^3 e^{-2\Delta E_T/kT}\right]}. \tag{1.29}$$

The factor of $e^{-2\Delta E_T/kT}$ greatly augments the cubic term in the numerator, even if the binding of TNRC6 is relatively weak. The default functional behavior in this case (relatively weak TNRC6 binding) is therefore

$$E[\text{rate}] \approx k_{deg} \frac{\tau r^3 e^{-2\Delta E_T/kT}}{1 + \tau r^3 e^{-2\Delta E_T/kT}}. \tag{1.30}$$

Another model can be created assuming strong TNRC6 binding and that neighboring Argonauts weakly bind to each other with energy J. The expected value of the Argonaut occupancy on TNRC6 is:

$$E[\text{Ago bound to TNRC6}] = \frac{3\left(\frac{R}{\rho_w}\right) e^{-\Delta E_T/kT} + 6\left(\frac{R^2}{\rho_w^2}\right) e^{-(2\Delta E_T + J)/kT} + 3\left(\frac{R^3}{\rho_w^3}\right) e^{-(3\Delta E_T + 2J)/kT}}{1 + 3\left(\frac{R}{\rho_w}\right) e^{-\Delta E_T/kT} + 3\left(\frac{R^2}{\rho_w^2}\right) e^{-(2\Delta E_T + J)/kT} + \left(\frac{R^3}{\rho_w^3}\right) e^{-(3\Delta E_T + 2J)/kT}} \tag{1.31}$$

$$\approx 3\left(\frac{R^3}{\rho_w^3}\right) e^{-(3\Delta E_T + 2J)/kT}, \tag{1.32}$$

due to the energy term J promoting the cubic term to the leading order.

In this model, we account for the finite supply of R using the following conservation law:

$$R_{tot} = R + TE[\text{Ago bound to TNRC6}] \tag{1.33}$$

$$\approx 3T\left(\frac{R^3}{\rho_w^3}\right) e^{-(3\Delta E_T + 2J)/kT} \tag{1.34}$$

in the limit of strong TNRC6 binding.

We can reconstruct the previous partition function with the new energy terms:

$$E[\text{rate}] = k_{deg}\tau \frac{3r + 3r^2 e^{-(\Delta E_T + J)/kT} + r^3 e^{-(2\Delta E_T + 2J)/kT}}{1 + 3r + 3r^2 e^{-J/kT} + r^3 e^{-2J/kT} + \tau\left[3r + 3r^2 e^{-(\Delta E_T + J)/kT} + r^3 e^{-(2\Delta E_T + 2J)/kT}\right]}. \tag{1.35}$$

In the strong binding limit the cubic terms dominate once again:

$$E[\text{rate}] = k_{deg}\tau \frac{r^3 e^{-(2\Delta E_T + 2J)/kT}}{1 + \tau r^3 e^{-(2\Delta E_T + 2J)/kT}} \tag{1.36}$$

but now we can make the substitution $3T\left(\frac{R^3}{\rho_w^3}\right)e^{-(3\Delta E_T+2J)/kT} = R_{tot}$:

$$E[\text{rate}] = k_{deg} \frac{\frac{1}{3}\frac{R_{tot}}{\rho_w} e^{-3\Delta E_R/kT}}{1 + \frac{1}{3}\frac{R_{tot}}{\rho_w} e^{-3\Delta E_R/kT}}. \tag{1.37}$$

In this model, the cooperative binding of Argonaut to TRNC6 saturates and they act as one single complex, binding to mRNA with binding energy $-3\Delta E_R$. Notice also that similarly tight binding with only two miRNAs is weaker by a factor $e^{-\Delta E_R/kT}$ which can be quite large, which would mean that there is essentially only regulation when all 3 miRNAs are present.

Model comparison

We have therefore derived three different modes of regulation.

Mechanism	Approximate rate	behavior
Slicing	$f(R_1, R_2, R_3) = \frac{k_{cat}}{K_M}(R_1 + R_2 + R_3)$	**additive**
TNRC6 weak	$f(R_1, R_2, R_3) = k_{deg}\frac{\tau r^3 e^{-2\Delta E_T/kT}}{1+\tau r^3 e^{-2\Delta E_T/kT}}$	**cooperative-cubic**
TNRC6 strong	$f(R_1, R_2, R_3) = k_{deg}\frac{\frac{1}{3}\frac{R_{tot}}{\rho_w} e^{-3\Delta E_R/kT}}{1+\frac{1}{3}\frac{R_{tot}}{\rho_w} e^{-3\Delta E_R/kT}}.$	**cooperative-linear**

These modes of regulation will lead to qualitatively different behavior of a miRNA IFFL. If the repression strength of a repressor is $f(R)$, then the IFFL equation is:

$$\frac{dR}{dt} = \alpha I - \gamma_R R, \tag{1.38}$$

$$\frac{dO}{dt} = \beta I - \gamma_O O - f(R)O. \tag{1.39}$$

At steady state:

$$R = \frac{\alpha I}{\gamma_R}, \tag{1.40}$$

$$O = \frac{\beta I}{\gamma_O + f(\frac{\alpha I}{\gamma_R})}. \tag{1.41}$$

The different functional forms are shown in **Figure 1.1**. If $f(R)$ is not linear in R, then the output will not be asymptotically flat. For instance, the cooperative regulation scales with R^3, so the IFFL asymptotically scales like I^{-2} as it goes to infinity (until the circuit binding saturates).

Importantly, in the cooperative mode, there is very little regulation when there is only one miRNA present, but significant regulation when multiple are present. This behavior, if confirmed empirically, would affirm the thesis that miRNAs are not subtle regulators, but act strongly in groups of three.

In the next chapters, I will present the synthetic miRNA-IFFL gene therapy and a toolkit of miRNA IFFLs. Using data from these projects, I will then select which of these models best fit the data.

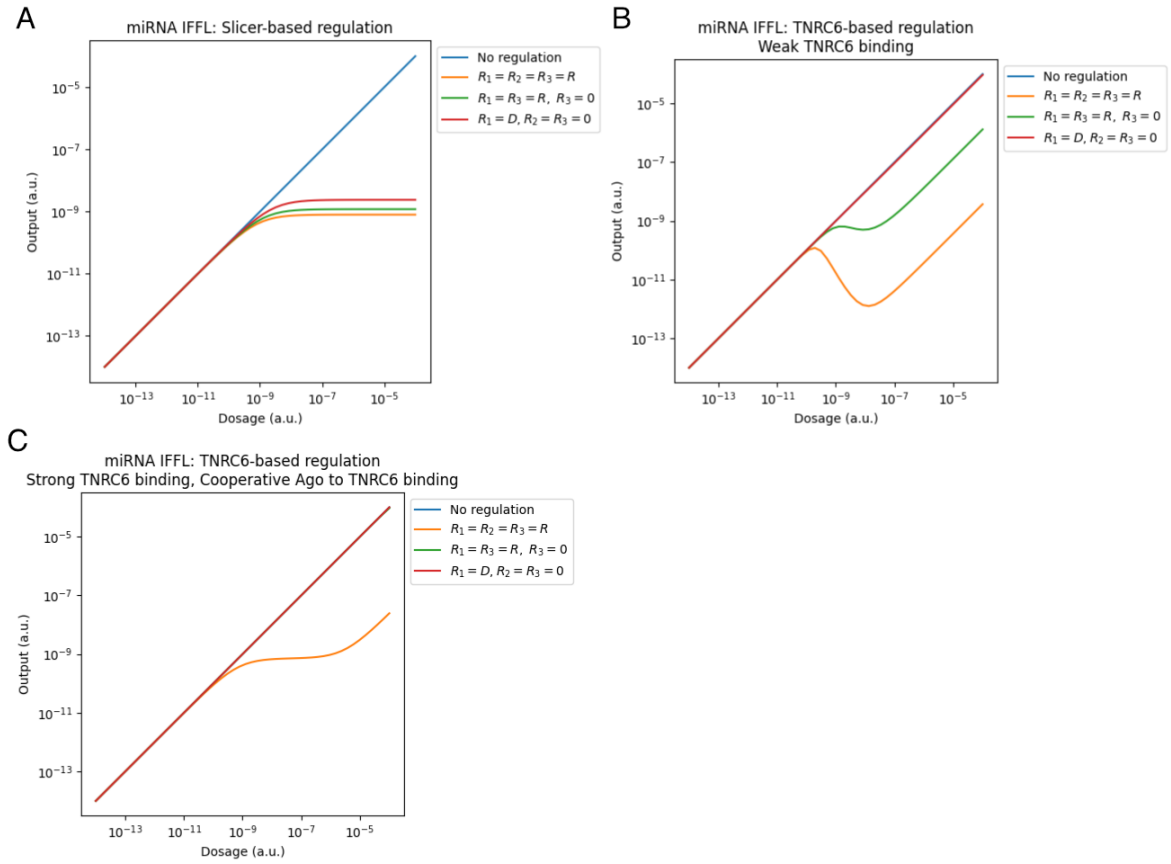

Figure 1.1: **miRNA regulation model comparison using IFFLs.** (A) Model of a miRNA IFFL using the slicer mode of regulation. Importantly, deletion of a miRNA has very little effect **(B)** miRNA IFFL using the TNRC6 pathway with weak binding. Notably, the regulation is eliminated when only one miRNA present. **(C)** miRNA IFFL using the TNRC6 pathway with strong TNRC6 binding and cooperative binding of Ago to TNRC6.

References

[1] F. Dyson and H. Bethe. *Disturbing the universe*. 1979.

[2] M. Elowitz and W. A. Lim. "Build life to understand it." *Nature* 468.7326 (2010), pp. 889–890.

[3] C. J. Bashor and J. J. Collins. "Understanding biological regulation through synthetic biology." *Annual Review of Biophysics* 47.1 (2018), pp. 399–423.

[4] X. Cai, C. H. Hagedorn, and B. R. Cullen. "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." *RNA* 10.12 (2004), pp. 1957–1966.

[5] T. A. Nguyen et al. "Functional anatomy of the human microprocessor." *Cell* 161.6 (2015), pp. 1374–1387.

[6] R. Yi et al. "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." *Genes & Development* 17.24 (2003), pp. 3011–3016.

[7] G. Hutvagner et al. "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." *Science* 293.5531 (2001), pp. 834–838.

[8] S. Iwasaki et al. "Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes." *Molecular Cell* 39.2 (2010), pp. 292–299.

[9] Y. Wang et al. "Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes." *Nature* 461.7265 (2009), pp. 754–761.

[10] B. P. Lewis, C. B. Burge, and D. P. Bartel. "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." *Cell* 120.1 (2005), pp. 15–20.

[11] D. P. Bartel. "MicroRNAs: target recognition and regulatory functions." *Cell* 136.2 (2009), pp. 215–233.

[12] V. Agarwal et al. "Predicting effective microRNA target sites in mammalian mRNAs." *eLife* 4 (2015), e05005.

[13] A. Kamenska et al. "Human 4E-T represses translation of bound mRNAs and enhances microRNA-mediated silencing." *Nucleic Acids Research* 42.5 (2014), pp. 3298–3313.

[14] S. Jonas and E. Izaurralde. "Towards a molecular understanding of microRNA-mediated gene silencing." *Nature Reviews Genetics* 16.7 (2015), pp. 421–433.

[15] E. Elkayam et al. "Multivalent recruitment of human argonaute by GW182." *Molecular Cell* 67.4 (2017), pp. 646–658.

[16] D. P. Bartel. "Metazoan microRNAs." *Cell* 173.1 (Mar. 2018), pp. 20–51. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.03.006. URL: http://dx.doi.org/10.1016/j.cell.2018.03.006.

[17] R. C. Friedman et al. "Most mammalian mRNAs are conserved targets of microRNAs." *Genome Research* 19.1 (2009), pp. 92–105.

[18] D. Baek et al. “The impact of microRNAs on protein output.” *Nature* 455.7209 (2008), pp. 64–71.

[19] M. Selbach et al. “Widespread changes in protein synthesis induced by microRNAs.” *Nature* 455.7209 (2008), pp. 58–63.

[20] R. E. Amir et al. “Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.” *Nature Genetics* 23.2 (1999), pp. 185–188.

[21] R. Tillotson and A. Bird. “The molecular basis of MeCP2 function in the brain.” *Journal of Molecular Biology* 432.6 (2020), pp. 1602–1623.

[22] M. B. Ramocki et al. “Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome.” *Annals of Neurology* 66.6 (2009), pp. 771–782.

[23] K. K. Gadalla et al. “Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice.” *Molecular Therapy* 21.1 (2013), pp. 18–30.

[24] K. K. Gadalla et al. “Development of a novel AAV gene therapy cassette with improved safety features and efficacy in a mouse model of Rett syndrome.” *Molecular Therapy Methods & Clinical Development* 5 (2017), pp. 180–190.

[25] S. E. Sinnett et al. “Improved MECP2 gene therapy extends the survival of MeCP2-null mice without apparent toxicity after intracisternal delivery.” *Molecular Therapy Methods & Clinical Development* 5 (2017), pp. 106–115.

[26] V. Matagne et al. “A codon-optimized Mecp2 transgene corrects breathing deficits and improves survival in a mouse model of Rett syndrome.” *Neurobiology of Disease* 99 (2017), pp. 1–11.

[27] V. Matagne et al. “Severe offtarget effects following intravenous delivery of AAV9-MECP2 in a female mouse model of Rett syndrome.” *Neurobiology of Disease* 149 (2021), p. 105235.

[28] L. Bleris et al. “Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template.” *Molecular Systems Biology* 7.1 (2011), p. 519.

[29] B. Chen et al. “Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.” *Cell* 155.7 (Dec. 2013), pp. 1479–1491. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.12.001. URL: http://dx.doi.org/10.1016/j.cell.2013.12.001.

[30] G. Liu et al. “The CRISPR-Cas toolbox and gene editing technologies.” *Molecular Cell* 82.2 (2022), pp. 333–347.

[31] A. R. Thurm, G. L. Janer Carattini, and L. Bintu. “Human synthetic biology and programmable gene regulation control.” *Annual Review of Genomics and Human Genetics* 26 (2025).

[32] M. B. Elowitz and S. Leibler. “A synthetic oscillatory network of transcriptional regulators.” *Nature* 403.6767 (2000), pp. 335–338.

[33] R. Milo et al. “Network motifs: simple building blocks of complex networks.” *Science* 298.5594 (2002), pp. 824–827.

[34] U. Alon. “Network motifs: Theory and experimental approaches.” *Nature Reviews Genetics* 8 (2007), pp. 450–461. DOI: 10.1038/nrg2102.

[35] S. Mangan and U. Alon. “Structure and function of the feed-forward loop network motif.” *Proceedings of the National Academy of Sciences* 100.21 (2003), pp. 11980–11985.

[36] L. Goentoro et al. “The incoherent feedforward loop can provide fold-change detection in gene regulation.” *Molecular Cell* 36.5 (2009), pp. 894–899.

[37] T. S. Gardner, C. R. Cantor, and J. J. Collins. “Construction of a genetic toggle switch in Escherichia coli.” *Nature* 403.6767 (2000), pp. 339–342.

[38] R. Zhu et al. “Synthetic multistability in mammalian cells.” *Science* 375.6578 (2022), eabg9765.

[39] H. E. Klumpe et al. “The computational capabilities of many-to-many protein interaction networks.” *Cell Systems* 14.6 (2023), pp. 430–446.

[40] P. Li et al. “Morphogen gradient reconstitution reveals Hedgehog pathway design principles.” *Science* 360.6388 (2018), pp. 543–548.

[41] K.-H. K. Chow et al. “Imaging cell lineage with a synthetic digital recording system.” *Science* 372.6538 (2021), eabb3099.

Chapter 2

SYNTHETIC DOSAGE-COMPENSATING MIRNA CIRCUITS ALLOW PRECISION GENE THERAPY FOR RETT SYNDROME

The following chapter is reproduced from:

[1] M. J. Flynn et al. "Synthetic dosage-compensating miRNA circuits for quantitative gene therapy." *bioRxiv* (Mar. 2024). DOI: 10.1101/2024.03.13.584179. URL: http://dx.doi.org/10.1101/2024.03.13.584179.

2.1 Summary

A longstanding challenge in gene replacement therapy is expressing a dosage-sensitive therapeutic gene within a tight therapeutic window. For example, loss of *MECP2* function causes Rett syndrome, while its duplication causes *MECP2* duplication syndrome. Viral gene delivery methods generate variable numbers of gene copies in individual cells, creating a need for gene dosage-invariant expression systems. Here, we introduce a compact, miRNA-based, incoherent feed-forward loop circuit that reduces *Mecp2* expression in cells and mouse brains. Single molecule analysis of endogenous and ectopic *Mecp2* mRNA revealed sustained expression not exceeding endogenous *Mecp2* in 95% of cells at a high viral dose. Delivered systemically in a brain-targeting AAV capsid, the circuit improved Rett behavioral symptoms for over 24 weeks, outperforming an unregulated gene therapy. These results demonstrate that synthetic miRNA-based regulatory circuits can augment a gene therapy by enabling precise in vivo expression.

2.2 Introduction

Gene replacement therapy promises to enable lasting cures for genetic diseases by delivering corrected copies of genes mutated within a patient. Recently, the field has achieved clinical successes addressing spinal muscular atrophy (SMA) [1], hemophilia [2, 3, 4], and inherited retinal dystrophy [5]. New technologies, such as engineered adeno-associated viral vectors (AAVs) that noninvasively transduce the CNS at high efficiencies [6] are enabling development for CNS disorders including STXBP1 Developmental Epileptic Encephalopathy and Parkinson's Disease [7, 8, 9, 10].

However, many gene therapies face a 'Goldilocks' problem: too little expression of the gene leads to the disease phenotype, but too much expression can induce other disease phenotypes. For example, loss of function of the *SYNGAP1* gene results in non-syndromic intellectual disability and

epilepsy [11] while overexpression of *SYNGAP1* leads to a pronounced depression of excitatory signaling in vitro [12]. Similarly, *UBE3A* deficiency causes Angelman's syndrome, but duplication and triplications of the locus are associated with autism spectrum disorder [13]. Overexpression toxicity is also a concern in clinical gene replacement. For example, loss of function mutations in *SMN1* cause spinal muscular atrophy, but overexpression of *SMN1* through gene therapy led to clinically silent but concerning dorsal root ganglion pathology in mouse and non-human primate studies of SMA gene therapy [14, 15, 16].

Rett syndrome presents the prototypical Goldilocks problem. It is a severe neurodevelopmental disease caused by loss-of-function mutations in the gene encoding MeCP2, a methyl-CpG/A binding protein [17]. MeCP2 binds to methylated regions of the genome and serves as a binding hub and bridge to the NCoR/SMRT co-repressor complex to repress methylated genes, an essential function for brain maturation [18]. Since *MECP2* is on the X chromosome, mutations in the gene lead to different symptoms between males and females. Human males cannot survive with a single non-functional copy of *MECP2* [19]. In human females heterozygous for a mutation in *MECP2*, Rett syndrome is characterized by a severe developmental regression at 7-18 months of age, progressive loss of speech and hand use, ataxia, and acquired microcephaly, among other symptoms [20]. MeCP2-deficient male mice show reductions in lifespan, brain size, neuron soma size, synapse counts, dendritic spine density, and electrophysiological activity [21]. However, mild overexpression of MeCP2 also leads to a disease phenotype [22] and duplication of the *MECP2* gene causes another disorder, *MECP2* duplication syndrome [23]. In engineered mice, Cre-based reactivation of *Mecp2* expression from its endogenous genomic context alleviates disease phenotypes, suggesting that the condition is reversible [24]. However, gene therapies based on AAV-mediated delivery of *Mecp2* to Rett model mice have induced toxicity [25, 26, 27, 28, 29, 30], specifically from overexpression of the MeCP2 protein [26]. Thus, a critical requirement for Rett syndrome gene replacement therapy is to express MeCP2 within a narrow therapeutic window.

Synthetic biological circuits, specifically incoherent feed-forward loops (IFFLs) in which the therapeutic gene is co-expressed with its own negative regulator, could help to address this problem. Sinnett et al. exploited endogenous miRNAs that are upregulated in response to *Mecp2* expression to reduce dosage dependence [31]. However, these natural miRNAs exhibited a limited, 2-fold dynamic range. We reasoned that a synthetic miRNA that is co-transcribed with the gene of interest would have a range of expression that matched the gene of interest, and therefore better compensate for gene dosage changes. In this work, we therefore introduced a synthetic compact miRNA-based IFFL circuit that limits *Mecp2* expression and reduces its sensitivity to gene dosage in cell culture. The circuit restricted ectopic *Mecp2* mRNA to levels less than those of endogenous *Mecp2* in 95% of cells in the mouse brain. Finally, a gene therapy vector containing the circuit outperformed an

unregulated gene therapy, improving behavioral symptoms over a timescale of 24 weeks. These results suggest that an integrated miRNA-based gene circuit could improve gene therapies for Rett syndrome and likely for other genetic diseases as well.

2.3 Results

Modeling predicts incoherent feedforward regulation can limit protein expression to a therapeutic window

Ectopic gene expression can deviate substantially from the endogenous expression distribution for several reasons. First, the number of gene copies delivered to an individual cell can vary by orders of magnitude due to varying uptake efficiency of different organs and cell types (**Figure 2.1A**, upper right) [32, 33], spatial gradients around a direct injection site (**Figure 2.1A**, bottom left) and Poisson-distributed stochastic vector uptake by individual cells (**Figure 2.1A**, bottom right) [34]. Second, engineered promoters typically used in gene therapy vectors often induce stronger expression than endogenous promoters, potentially leading to toxicity from even a single copy (**Figure 2.1B**, left panel) [35]. Finally, for an X-linked disease such as Rett syndrome, even an ectopic gene expressed at the physiological levels could lead to overexpression in cells that also express the endogenous copy (**Figure 2.1B**, right panel). Current approaches to limiting ectopic expression include optimizing the promoter [36]. However, while these approaches generally reduce mean expression relative to unregulated constructs, they cannot actively adapt to variations in gene dosage.

The incoherent feedforward loop (IFFL) is an adaptive biological circuit motif that can potentially address these challenges [37, 38, 39, 40]. Previous work has shown that synthetic IFFL circuits can successfully buffer gene expression against variations in gene dosage [41, 42, 43, 44], noise from upstream regulators [45], competition for cellular resources [44, 46], or general perturbations [47]. Here, we consider IFFLs in which a target gene and its negative regulator are co-transcribed, so that higher gene dosage leads to greater transcription rates of both components (**Figure 2.1C**) [37, 38, 48]. A simple mathematical model of such an IFFL (**Methods**) shows that, given unregulated single-copy production level P, gene dosage D, miRNA degradation rate γ_R, miRNA catalytic second order rate constant k, and miRNA:mRNA production ratio α, the steady state expression level of the output gene, G, approaches a limiting value as gene dosage increases:

$$G = \frac{\gamma_R}{k\alpha}\frac{DP}{\frac{\gamma_R}{k\alpha} + DP} \rightarrow \frac{\gamma_R}{k\alpha}. \tag{2.1}$$

This equation shows that beyond the unregulated expression level of $\gamma_R/k\alpha$, increasing repression from the negative regulator cancels out increasing production of the gene, maintaining an expression ceiling across a wide range of gene dosages (**Figure 2.1D**).

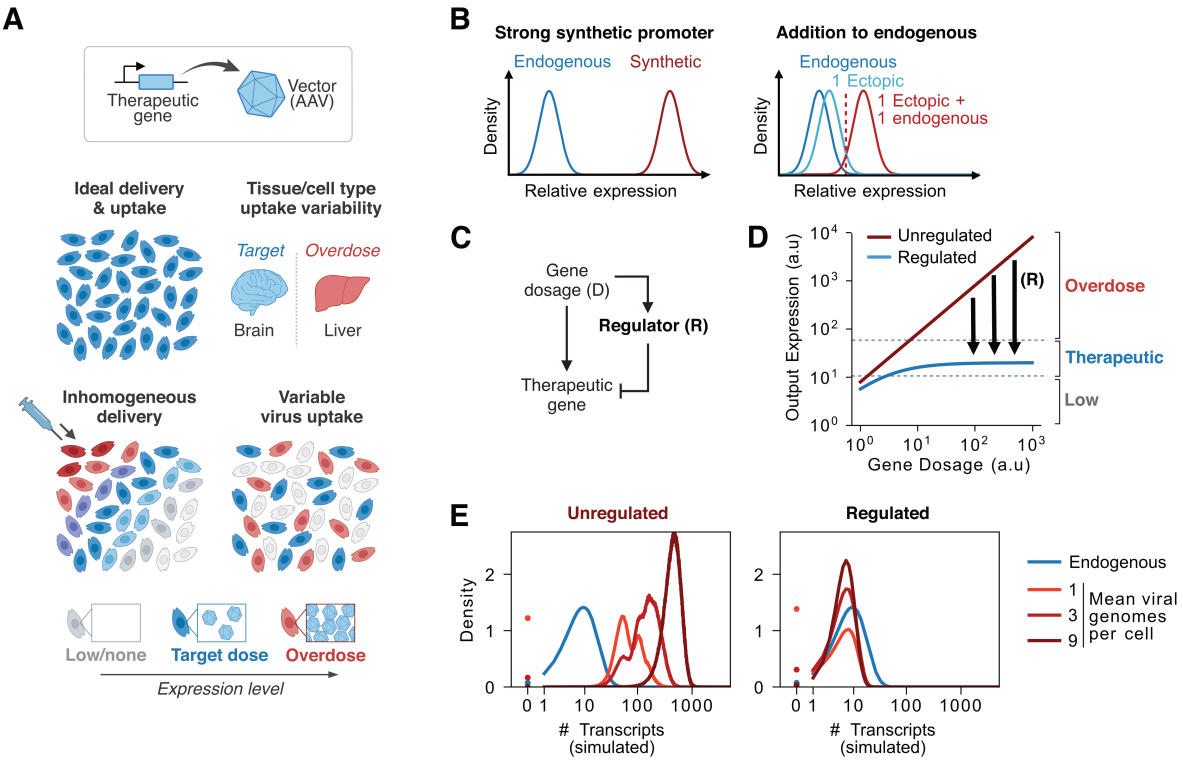

Figure 2.1: (Caption on next page)

We further extended this model to incorporate discrete gene copy numbers, stochastic vector delivery, and bursty gene expression kinetics. We also assumed that a single copy of the ectopic promoter was strong enough to exceed native expression levels (**Figure 2.1E**, **Methods**). Simulated IFFLs operating in these regimes successfully regulated the distribution of expression to be similar to that of a target endogenous distribution, across different mean copy numbers of the gene (**Figure 2.1E**, **Methods**). These results suggested that a suitably engineered IFFL could generate a more robust and precise therapeutic gene expression system.

A synthetic miRNA-based IFFL module enables dosage compensated *Mecp2* expression

In some natural genes, intronic miRNAs downregulate expression of their host gene, forming an IFFL within a single transcript [49, 50, 51]. This circuit architecture has also been demonstrated synthetically [41] and could have desirable properties for gene therapy, since intronic miRNA expression cassettes are non-immunogenic and genetically compact. However, it has not been established if synthetic miRNA-based IFFLs can match the expression level of an endogenous mRNA, or whether they can function in an AAV-based gene therapy.

To address these questions, we designed a set of miRNA-based dosage compensating IFFL con-

Figure 2.1: **Mathematical modeling shows how incoherent feedforward loop circuits can maintain gene expression within a therapeutic window. (A)** Gene therapy contends with multiple sources of variability in expression. Ideally, all cells would receive the same number of viral genome copies, and homogeneously express the correct amount of the therapeutic gene (upper left). However, viral uptake rates can vary greatly by organ and cell type (upper right), such that a dose that is therapeutic in one organ (e.g., brain, blue) may be toxic in another that takes up viral vectors at a higher rate (e.g., liver, red). With direct injection, cells close to the injection site receive more copies than cells farther away (lower left). Finally, even with correct mean delivery, viral uptake remains subject to stochastic variation (lower right). **(B)** The high level of expression induced by synthetic promoters commonly used in gene therapy may cause toxic overexpression from even a single transgene copy (left). Additionally, for X-linked genes like *MECP2*, approximately half of cells in affected females express a functional endogenous copy. The therapeutic gene must not overexpress MeCP2 when its expression is added to the wildtype allele (right). **(C)** Schematic of an incoherent feedforward loop (IFFL) motif in which a therapeutic gene is co-expressed with its own negative regulator. **(D)** Therapeutic gene expression as a function of gene dosage, as modeled for an idealized IFFL (**Methods**, 'Modeling an idealized dosage-compensating IFFL'). The increasingly negative action of the negative regulator (R, black) compensates for increases in gene dosage, leading to a plateau regime where large changes in gene dosage yield nearly the same output expression of the circuit (blue), preventing overexpression. **(E)** Simulated distributions of therapeutic gene expression at different viral MOI, either unregulated (left) or regulated by the IFFL (right), compared to a target endogenous expression distribution (blue). Simulations incorporate stochastic viral uptake, bursty transcription, and stochastic enzyme kinetics. The mean expression level of a single copy of the ectopic gene is assumed to be higher than that of the endogenous gene. The IFFL circuit compensates for these sources of variation.

structs (**Figure 2.2A**). We engineered a divergent promoter made up of the CMV enhancer flanked by the MeP229 promoter [25] in the forward direction and an intron-free Ef1α promoter in the reverse direction ('ECM promoter'). The forward promoter drives expression of a previously characterized (full-length) MeCP2_e2-EGFP protein fusion to facilitate analysis of protein expression [29]. The reverse promoter drives expression of unregulated mRuby3 as an indicator of gene dosage.

To implement IFFL regulation, we incorporated a miRNA expression cassette in a synthetic intron [52] within the 3'UTR of *Mecp2-EGFP*. This miRNA cassette is based on the strong and well-characterized miR-E backbone [53], which generates a miRNA complementary to a 21-bp sequence derived from *Renilla luciferase*, which is orthogonal to the human genome. To compare two different strengths of regulation, we inserted either 1 or 4 copies of the target sequence into the 3'UTR, upstream of the miRNA-containing intron, to create "1x" and "4x" circuits (**Figure 2.2A**). We also constructed an "unregulated" control construct lacking both the miRNA and its target sites. All 3 constructs were less than 4300 base pairs in length, and thus small enough to be efficiently packaged inside an AAV.

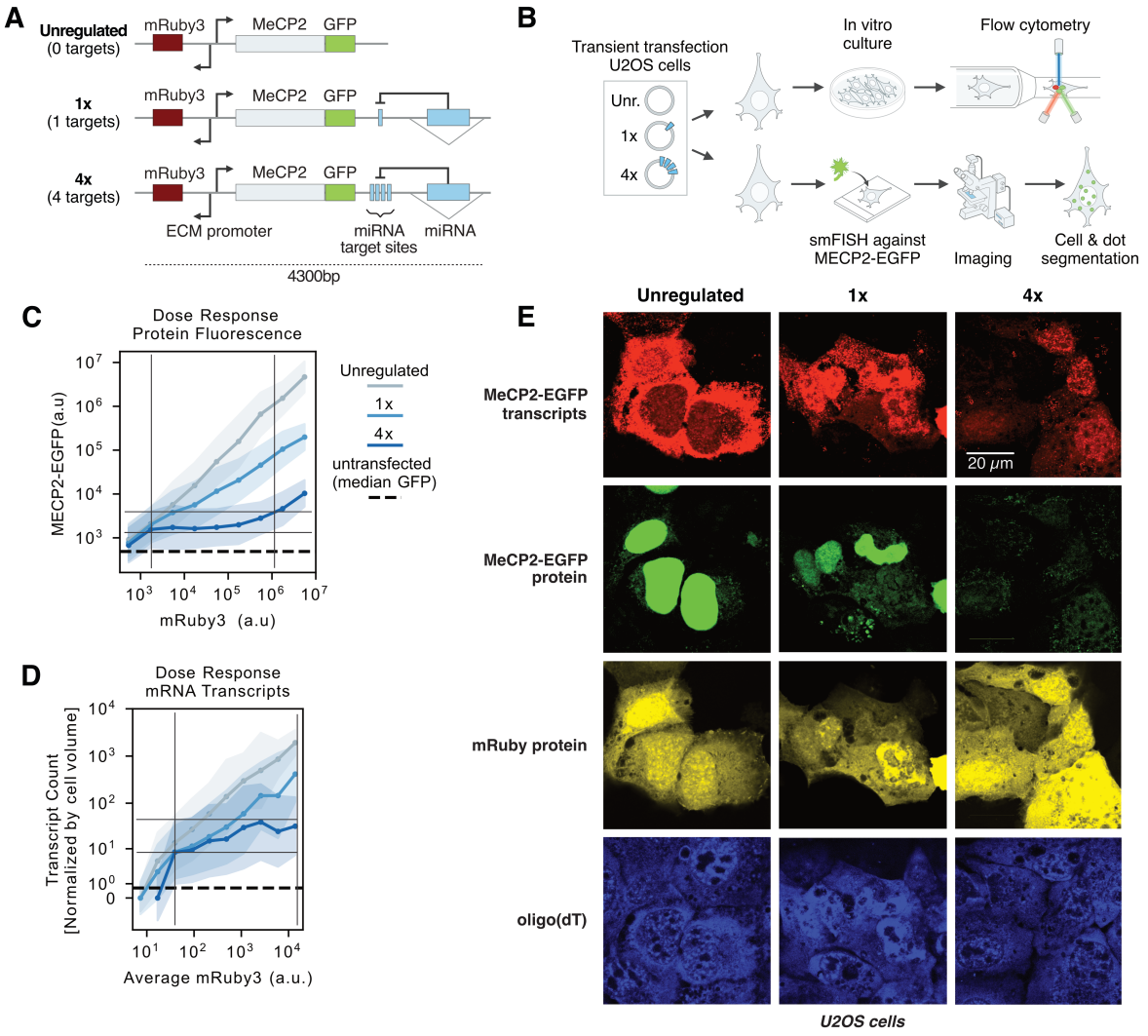

Figure 2.2: (Caption on next page)

To test the ability of these circuits to compensate for variation in gene dosage, we quantified their expression as a function of dosage at the protein and mRNA levels. For protein-level quantification, we transiently transfected U2OS cells and analyzed MeCP2 expression using flow cytometry (**Figure 2.2B**, upper path, **Methods**). Unregulated MeCP2 was expressed at a level proportional to gene dosage, as expected (**Figure 2.2C**). The 1x construct reduced MeCP2-EGFP expression and its dependence on dosage (slope of EGFP versus mRuby3) (**Figure 2.2C**). While it did not achieve ideal dosage compensation, it provided a useful intermediate-regulation condition for subsequent studies. The 4x circuit generated behavior closer to that expected from the simplified IFFL model (**Figure 2.1D**), with relatively constant (<3-fold variation) expression across a broad range (>300-fold) of gene dosage (**Figure 2.2C**). Beyond $\sim 10^6$ a.u., the curve became linear again, as expected in a model accounting for more realistic two-step degradation (**Methods**).

Figure 2.2: **Synthetic miRNA IFFLs can adapt to variations in gene dosage in cell culture.** **(A)** Designs of 3 constructs based on a divergent promoter producing MeCP2-EGFP in the forward direction and the mRuby3 dosage indicator in the reverse. The "unregulated" circuit has no miRNA target sites and no miRNA cassette and serves as a no-circuit control. The 1x and 4x circuits contain a miRNA cassette located within a synthetic intron in the 3' UTR of *Mecp2-EGFP*, which respectively targets 1 or 4 fully complementary miRNA target sites upstream of the intron. **(B)** Workflow to characterize circuit performance at the mRNA and protein levels. For protein, U2OS cells were transiently transfected and cultured for 48 hours. We then measured protein expression by flow cytometry (upper path). For mRNA, U2OS cells were transiently transfected, incubated for 48 hours, fixed, and then analyzed by smFISH and confocal microscopy (lower path). **(C)** MeCP2-EGFP protein fluorescence as a function of mRuby3 dosage indicator for the 3 constructs, as measured by flow cytometry. MeCP2-EGFP fluorescence was proportional to dosage for the unregulated construct (gray), as expected. For the 1x construct (medium blue), the slope (on a log-log plot) was reduced, indicating decreased dosage sensitivity. For the 4x construct (dark blue), MeCP2-EGFP expression was nearly independent of dosage over 2.5 orders of magnitude variation in gene dosage (framed region). This stable expression level was approximately 3-fold above the fluorescence of untransfected cells (dashed black line). Here, and in D, shaded regions represent ±1 standard deviation of the logarithmic expression values. **(D)** *Mecp2-EGFP* transcript count as a function of average mRuby3 dosage-indicator fluorescence, as measured by smFISH and confocal microscopy. smFISH dot counts were normalized by cell volume (**Methods**). The 4x construct produced an expression level that varied less than 4-fold over a greater than 300-fold range of dosage (framed region). **(E)** smFISH imaging of ectopic *Mecp2-EGFP* transcripts (first row) and protein (second row), as well as mRuby dosage indicator (third row) and oligo(dT) mRNA marker (fourth row). In the first row, transcript count is proportional to red signal intensity. Cells displayed comparable levels of mRuby protein (third row) and mRNA quality and expression (fourth row) in all conditions, while *Mecp2-EGFP* expression decreases with stronger IFFL regulation at both transcript (first row) and protein (second row) levels.

To quantify differences in expression at the mRNA level, we chose smFISH over other methods such as qPCR because it allows direct comparison of gene dosage and ectopic MeCP2 expression in single cells. We transiently transfected U2OS cells with each construct, performed smFISH [54] against the *Mecp2-EGFP* transcript, and imaged both protein fluorescence and transcripts using confocal microscopy (**Figure 2.2B**, lower path, **Methods**). Consistent with the flow cytometry results, among cells expressing similar levels of the mRuby dosage reporter (**Figure 2.2E**, third row) and an oligo(dT) mRNA marker (**Figure 2.2E**, fourth row), MeCP2-EGFP mRNA and protein expression levels decreased with increasing number of miRNA target sites (**Figure 2.2E**, first and second row).

To quantitatively measure the relationship between gene dosage and target mRNA levels, we computationally segmented cells in the smFISH images and counted individual transcripts (**Methods**). The unregulated and 1x constructs showed linear and sublinear, but still increasing, dependence on the mRuby3 dosage indicator, respectively (**Figure 2.2D**). The 4x construct exhibited the lowest level of expression, which also varied less than 4-fold over a greater than 300-fold range of dosage (**Figure 2.2D**).

To test for potential off-target effects of the miRNA regulation, we performed bulk-RNAseq and compared transcriptome expression of a BFP-miRNA (miR-E) cassette to a negative control transfection of a BFP-only expression vector (**Methods**). Few genes were up- or down-regulated in the miRNA condition relative to the negative control and none contained partial sequence matches to the miRNA (**Figure 2.5A**). These results suggest any sequence-specific perturbations from the synthetic miRNA itself were minimal.

Taken together, these results indicate that the 4x IFFL circuit can establish dosage-insensitive expression and reduce the magnitude of cell-cell variation in gene expression in cell culture, without significantly perturbing endogenous gene expression.

AAV-delivered IFFL circuits reduce *Mecp2* mRNA to endogenous levels in mouse brains

Maximizing therapeutic efficacy and safety in Rett syndrome gene therapy requires expressing AAV-delivered *Mecp2* at appropriate levels in the brain. While the precise therapeutic window of *Mecp2* expression is unknown, it presumably spans the endogenous expression range. We therefore sought to quantitatively compare expression of AAV-delivered *Mecp2* with endogenous *Mecp2* in mouse brains, focusing on the RNA level to allow single-molecule, single-cell quantitation. To distinguish ectopic and endogenous transcripts, we designed two sets of orthogonal hybridization chain reaction (HCR) probes: one set targeted the EGFP sequence exclusive to the ectopic transcript, while the other set targeted sequences in the 3' UTR of the major isoforms of *Mecp2* that are exclusive to the endogenous gene (**Figure 2.3A**).

To achieve brain-specific delivery, we took advantage of Cap.B22, a recently developed AAV capsid variant that efficiently targets the brain when systemically delivered, while de-targeting the liver and other organs [55]. We packaged each of the constructs in Cap.B22. To compare circuit-regulated expression to that produced by a typical promoter commonly used in gene therapy [35, 56, 57], we also packaged a constitutive CAG-GFP construct based on the synthetic CAG promoter. We then systemically injected each variant into WT mice at a relatively high dose of 5×10^{12} viral genomes (vg) per mouse. Using a high AAV dose provides a stringent test of dosage compensation by the circuits. 3 weeks post-injection, we collected brain sections and analyzed mRNA levels using HCR and confocal microscopy (**Figure 2.3A**, **Methods**). As expected, in the resulting images, the

highest expression of *Mecp2* mRNA and protein was seen in brains that received the CAG-GFP construct, followed by the unregulated, 1x, and 4x constructs (**Figure 2.3D**, third and fifth row).

To quantify the effect of the circuit in single cells, we segmented cells based on oligo(dT) fluorescence (**Figure 2.3D**, second row) and counted individual transcripts within individual cells (**Methods**). In a negative control condition, where primary HCR probes were not added, only zero or one transcripts were detected in 48% of cells, indicating low background signal (**Figure 2.5B**). Based on this, we set 5 MeCP2-EGFP mRNA transcripts per cell as a threshold for cell transduction. The proportion of cells in each condition that met this threshold was 100% for the CAG-GFP condition, 37% for the unregulated condition, 19% for the 1x condition, and 22% for the 4x condition. The proportion of cells that passed the threshold was influenced by the expression level, which was reduced by regulation. The CAG-GFP construct overexpressed the ectopic transcript by a median factor of 20-fold relative to endogenous *Mecp2* (**Figure 2.3B**, first panel), showing how standard synthetic promoters can greatly overexpress *Mecp2*.

The unregulated and regulated constructs exhibited distinct behaviors. All three constructs underexpressed ectopic MeCP2 relative to endogenous MeCP2 at the median (**Figure 2.3C**). However, the top decile of brains that received the unregulated construct overexpressed *Mecp2* by at least 50% (**Figure 2.3B**, second panel, **Figure 2.3C**). Since even mild overexpression of MeCP2 was previously found to be harmful [22], this could potentially lead to toxicity. Both regulated constructs underexpressed ectopic MeCP2 relative to endogenous, with a median ectopic expression level 10-fold less than the endogenous MeCP2 expression level for the 4x circuit, and the majority of cells exhibiting no ectopic transcripts for the 1x construct (**Figure 2.3C**). Neither circuit had a tail of MeCP2-overexpressing cells, and only 5% of cells exhibited more ectopic than endogenous MeCP2 (**Figure 2.3C**). The 4x circuit thus ensured that nearly the entire ectopic distribution remained below the endogenous MeCP2 expression level. A similar distribution of endogenous transcripts was measured in each condition (**Figure 2.6A**, **2B**). While these results were for cells in image fields taken randomly over the brain, similar results held when brain regions were analyzed individually (**Figures 2.6C-H**).

Together, these results indicate that the IFFL circuits can regulate *Mecp2* expression within the mouse brain at mRNA expression levels below those of endogenous *Mecp2*. They also reveal that different construct designs can generate distinct distributions of relative expression levels. Finally, they provoke the critical question of how these different distributions may ultimately impact disease progression in a model organism.

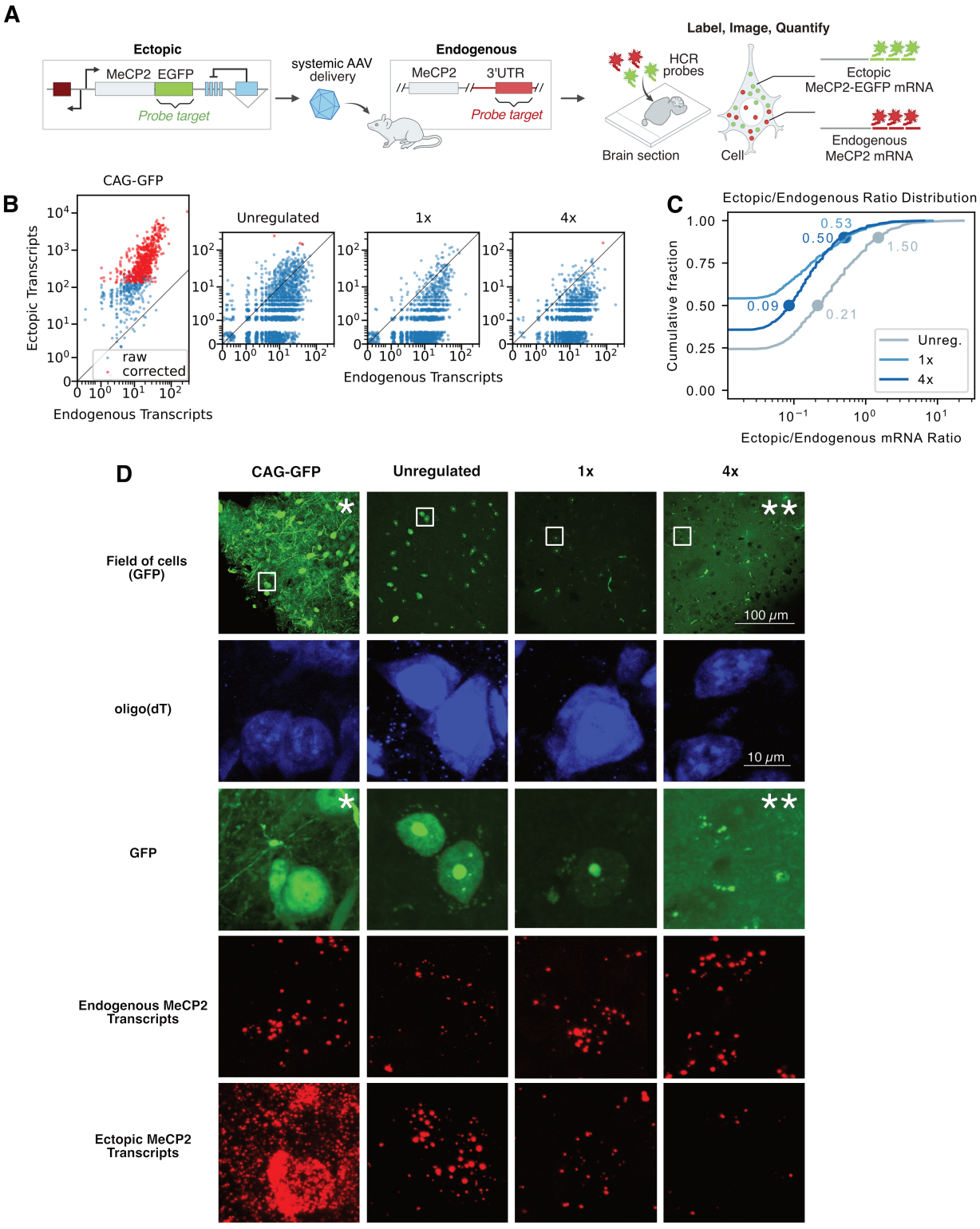

Figure 2.3: (Caption on next page)

Figure 2.3: **Synthetic miRNA IFFLs regulate mRNA expression to levels comparable to or less than endogenous *Mecp2* levels in mouse brains.** **(A)** Orthogonal HCR probes specifically target endogenous or ectopic *Mecp2*. Ectopic *Mecp2-EGFP* was targeted with HCR probes against the EGFP coding sequence. Endogenous *Mecp2* was targeted with HCR probes against sequences in the endogenous 3' UTR which do not appear in the ectopic construct. Mice were injected with viral constructs and, after 3 weeks of expression, brain slices were analyzed by HCR and confocal microscopy. **(B)** Ectopic (y-axis) vs endogenous (x-axis) transcripts measured in single cells (**Methods**). The diagonal black line denotes equal expression. Red dots denote cells whose counts have been corrected to account for dense dot spacing (**Methods**, **Figure 2.5C**). Each condition has approximately 1500 cells. CAG-EGFP expressed ectopic transcript at levels an order of magnitude greater than endogenous *Mecp2* transcripts. The unregulated, 1x, and 4x constructs showed progressively reduced ectopic expression. **(C)** Circuits reduce ectopic to endogenous mRNA ratio. Empirical cumulative distribution functions (ECDFs) of the ratio of ectopic to endogenous *Mecp2* transcripts in single cells. Annotated values denote the 50th and 90th percentiles. **(D)** Sample HCR images, focusing on individual cells in a field of cortical neurons (first row, white boxes denote enlarged areas below). All cells exhibited similar endogenous *Mecp2* expression (fourth row), but decreasing ectopic MeCP2-EGFP protein (third row) and ectopic *Mecp2* transcripts (fifth row) from CAG-GFP to unregulated to 1x to 4x constructs.
(*) Brightness of CAG-GFP image has been reduced to better distinguish cells.
(**) Brightness of the 4x-GFP image has been increased to make the dimmer fluorescence of MeCP2 nuclear puncta more visible.

Circuit-regulated AAV-MeCP2 outperforms unregulated AAV-MeCP2 in a mouse model of Rett syndrome

To assess whether regulating the expression of AAV-delivered *Mecp2* improves behavioral outcomes in Rett model mice compared to unregulated gene therapy, we evaluated the impact of the three constructs in a mouse line carrying a *Mecp2*-null allele [58]. To quantify outcomes, we used the standardized Rett score, a validated composite phenotypic performance scale including hindlimb clasping, gait analysis, mouse weight and other domains, which has been used to describe female $Mecp2^{-/X}$ mice by other groups (**Methods**) [24]. Female $Mecp2^{-/X}$ mice were treated with the unregulated, 1x, or 4x constructs or left uninjected, creating 4 groups of mice, with five or more mice per group (**Methods**). At 4 weeks of age, mice were randomly assigned to sex/genotype matched cohorts for systemic injection with AAV-CAP.B22-packaged constructs at a dose of 10^{14} vg/kg (**Methods**). Blinding was performed by a third-party who did not participate in phenotypic analysis. The baseline Rett score was measured at 4 weeks of age, with subsequent measurements at a biweekly interval. All phenotype assessments were performed blinded to genotype and treatment condition. These measurements were continued until the mice reached 28 weeks of age (**Figure 2.4A**).

The Rett phenotype score increased systematically over time, driven primarily by increases in hindlimb clasping and abnormal gait subscores (**Figure 2.7E**). These increases were not observed in wildtype littermates, as expected (**Methods**, **Figure 2.4B,C**). By contrast, the mice that received unregulated *Mecp2* exhibited a similar motor phenotype to the untreated *Mecp2*$^{-/X}$ animals, potentially due to *Mecp2* overexpression-induced toxicity (**Figure 2.4B,C**) [22]. The 1x construct produced a mild improvement in symptoms, suggesting that limiting *Mecp2* expression can reduce toxicity (**Figure 2.4B,C**). Strikingly, the mice that received the tightly regulated (4x) *Mecp2* performed more comparably to WT controls than the other constructs (**Figure 2.4B,C**), with time-averaged behavior scores ('Mean behavior score') significantly lower than those of both the uninjected mice (bootstrap p = 0.003) and the mice that received the unregulated construct (bootstrap p=0.01) (**Figure 2.4D**). Similar significant differences were found at individual timepoints, including week 22 (**Figure 2.7A**) and week 28 (**Figure 2.7B**). Notably, the standard deviation of the time-averaged scores for the 4x construct, 0.30, was also significantly lower than that of the unregulated construct, 1.04 (bootstrap, p=0.02), suggesting that the regulated gene therapy causes a less variable phenotype score. Together, these results demonstrate that tighter regulation improves the ability of AAV-delivered *Mecp2* in mitigating progression of the Rett phenotype in female *Mecp2*$^{-/X}$ mice.

The gene therapy did not improve all Rett phenotypes. Notably, female Rett model mice have an obesity phenotype that was not significantly affected by any of the injected constructs (**Figure 2.7D**). Additionally, we assessed whether regulation of *Mecp2* expression could confer any benefit to *Mecp2*-null male mice, which express no MeCP2 in the brain and exhibit a more severe phenotype, including a 12-week lifespan. However, we found no extension of the lifespan of male *Mecp2*-null male mice in any condition (**Figure 2.7C**). This contrasts with modest lifespan extensions reported with other Rett syndrome gene therapies [26], and with the strong behavioral results in the female mice in this work. Thus, while the regulated *Mecp2* construct achieved strong phenotypic improvements in females without evident toxicity, it did not fully address all Rett phenotypes.

Ectopic constructs are potentially subject to unintended regulation or epigenetic silencing that could impact their operation over time. To check for such effects, we repeated the HCR analysis on the brains of mice that had completed the behavioral testing, after 24 or more weeks of observation. All three constructs exhibited similar expression profiles as they did in the initial analysis (**Figure 2.3**). Specifically, the unregulated construct produced an expression distribution whose median cell underexpressed ectopic *Mecp2* by a factor of about 3, while 10% of cells overexpressed ectopic *Mecp2* by 4-fold or more (**Figure 2.4E,F**). The 1x and 4x constructs both expressed ectopic *Mecp2* at levels significantly lower than endogenous *Mecp2*, with the median cell expressing 0 transcripts and only 4% of cells expressing ectopic *Mecp2* at the endogenous level or higher (**Figure 2.4E,F**).

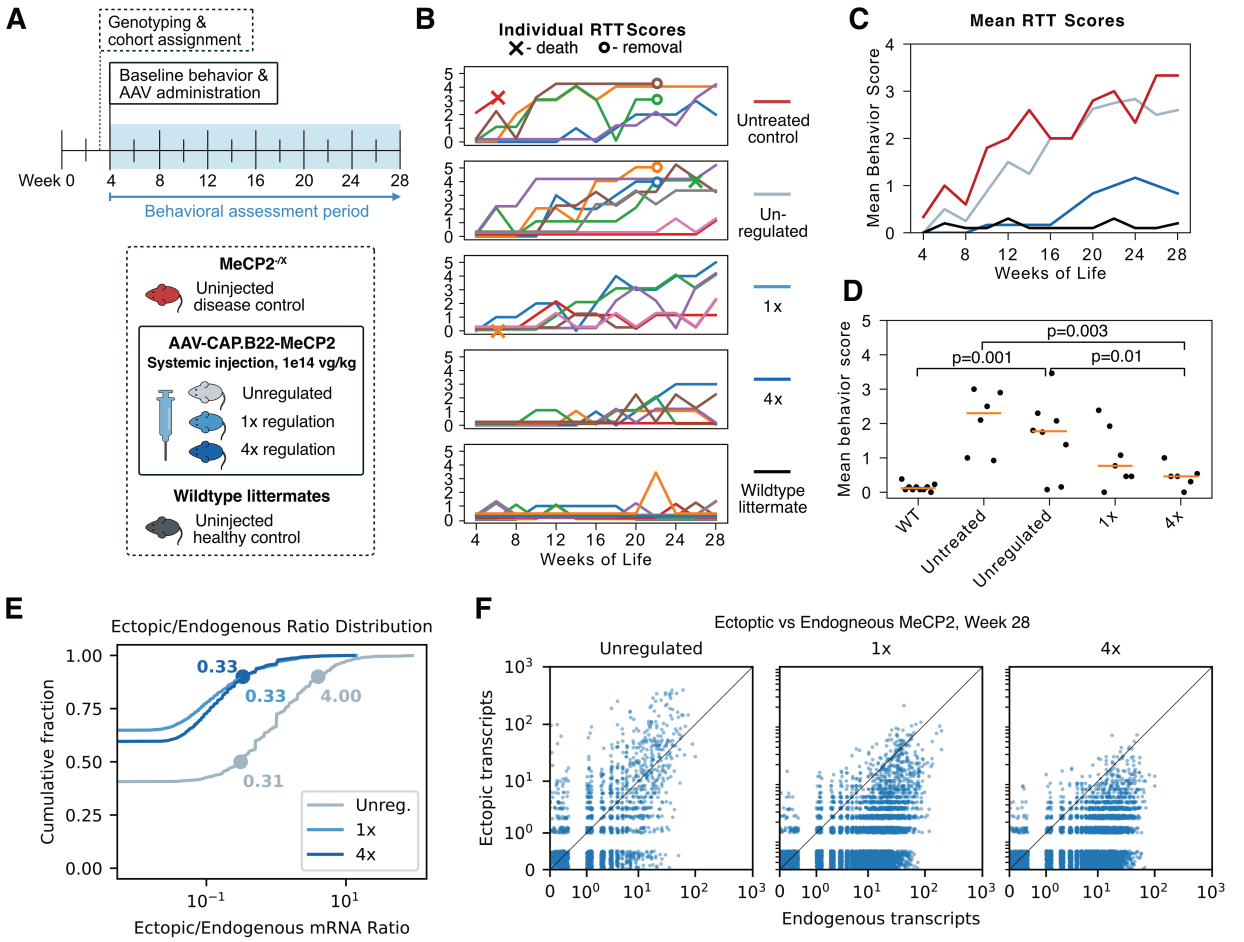

Figure 2.4: (Caption on next page)

It should be noted that, while the median cell may express 0 transcripts at a given point in time, it is expected that cells stochastically sample different points of the expression distribution, likely evening out protein-level expression over time. Overall, these data suggest that the circuits robustly maintained reduced expression over the 24 weeks of the study.

Taken together, these results demonstrate that a synthetic miRNA-based IFFL circuit can significantly and durably augment a model of Rett syndrome gene therapy in a Rett mouse model, over periods of at least 24 weeks.

Discussion

Without regulation, therapeutic genes are expressed at varying levels across cells. These levels could range from insufficient to toxic overexpression (**Figure 2.1**). Here, we designed and optimized a compact synthetic miRNA-based incoherent feedforward loop circuit that can compensate for variations in gene dosage and other sources of expression variation to ensure that total mRNA and protein expression remains within a therapeutic window (**Figure 2.2**). When incorporated

Figure 2.4: **IFFL-regulated gene therapy outperforms unregulated gene therapy in a mouse model of Rett syndrome.** **(A)** Experimental design: female *Mecp2*$^{-/X}$ mice were divided into 4 treatment groups (dashed box): uninjected or injected with one of the 3 constructs (**Figure 2.2A**) packaged in AAV-CAP.B22. Female wildtype littermates were included as healthy controls. Timeline: at 4 weeks of age, baseline Rett behavior scores were recorded and the mice were injected with 1×10^{14} vg/kg. Rett phenotype scores [24] were then measured biweekly for 24 weeks (tick marks on timeline). **(B)** Rett behavior (RTT) scores over time for individual mice (colored tracks) in each group. In each of the uninjected control, unregulated, and 1x groups, 1 mouse died during the study ("X" marker). In the 4x and wildtype groups, no mice died. After 22 weeks, 2 mice were removed from both the uninjected control group and the unregulated group for another experiment ("O" marker). **(C)** Mean score trajectories for each group in (B). Colors indicate treatment groups, as shown in the legend to the right of (B). **(D)** Rett behavior scores averaged across all timepoints for each mouse ('Mean behavior score') by condition. Orange bars denote the median of each group. Mice that received the 4x construct had mean scores significantly lower than both the uninjected mice (bootstrap p=0.003) and mice that received the unregulated construct (bootstrap p=0.01). The standard deviation of the 4x group, 0.30, was also significantly lower than that of the unregulated group, 1.04 (bootstrap p=0.02). The mean scores of the wildtype controls were significantly lower than the uninjected controls (bootstrap p=0.0003) and mice receiving the unregulated (bootstrap p=0.001) and the 4x (bootstrap p=0.01) constructs. **(E)** Empirical cumulative distribution functions (ECDFs) of the ratios of ectopic to endogenous transcripts in single cells after 24 weeks of expression (cf. **Figure 2.3C**). **(F)** Ectopic vs endogenous *Mecp2* transcripts quantified in single cells of mouse brains after behavior experiments were completed (week 28). Brain slices were analyzed by HCR, as in **Figure 2.3**. Despite continuous expression for 24 weeks, ectopic and endogenous levels remained similar for each construct.

---

in an AAV vector, the circuit maintained expression in mouse brains at or below the level of endogenous *Mecp2* (**Figure 2.3**). In the context of MeCP2 gene therapy, the tightly regulated ("4x") circuit variant improved behavioral phenotype score in a Rett female mouse model, outperforming unregulated and more weakly regulated constructs (**Figure 2.4**). This work benefited from a brain-targeting systemic AAV capsid, Cap.B22 [55], which allows higher viral titers because it avoids delivery to the liver, where toxicity has been observed in previous studies [25]. However, comparison between unregulated and regulated circuit variants shows that the circuit provides additional benefit beyond that achieved by capsid targeting alone. Synthetic miRNA IFFL circuits could potentially improve gene therapy for Rett syndrome and other genetic disorders.

The single-molecule analysis employed here enabled us to quantitatively compare ectopic and endogenous expression levels in the same cell. This showed that the unregulated construct generated a meaningful subpopulation of cells that overexpressed MeCP2, while the regulated constructs did not yield such a population (**Figure 2.3C**). The superior performance of the tightly regulated circuit in rescuing the Rett phenotype (**Figure 2.4B**) may be due to this lack of higher expression.

Additionally, we note that both circuits reduced expression in the brain to levels far lower than that of a CAG-GFP construct, which used a more typical synthetic promoter (**Figure 2.3B**).

In contrast to previous studies, the regulated *Mecp2* gene therapy described here did not improve the objective criteria of survival of male *Mecp2*-null mice or the obesity phenotype of heterozygous females. Only a subjective metric, the Rett phenotype score, was improved. This difference may reflect our use of the e2 MeCP2 isoform rather than the more active e1 isoform used in other studies [25, 26, 27, 29, 59]. While some studies have observed phenotype relief with the e2 isoform [59], later studies found a deficiency in phenotype relief using only the e2 isoform [60, 61]. Another potential explanation may be our choice of a central nervous system (CNS)-targeting capsid, whereas previous studies used AAV9 vectors, which deliver genes more broadly. Several Rett phenotypes in humans, including breathing irregularities, cardiovascular dysfunction, and decreased pain sensitivity [62], have been attributed to functions in the peripheral nervous system (PNS). Applying the gene therapy to the PNS and other tissues with targeted AAVs [63] could help to improve efficacy and clarify the role of MeCP2 in non-CNS cells. In future studies, it will also be important to compare to previously characterized benchmark vectors which have successfully shown improvements to these objective criteria and also an opportunity to show improvements on liver toxicity.

During revisions of this manuscript, Ross et al. published a study on an AAV9 Rett syndrome gene therapy using a miRNA-based IFFL to regulate *Mecp2* expression [64]. Their approach extended lifespan of male $Mecp2^{-/y}$ mice to a median of 37 weeks, eliminated previously observed toxicity, and quantitatively demonstrated MeCP2 protein regulation, but did not assess long-term effects on Rett syndrome phenotypes in females. It will be interesting to find out whether the difference in male lifespan extension between their study and the present work can be explained by different isoform usage (e1 and e2, respectively) or other factors [59, 60, 61]. Despite differences in isoform choice, both studies demonstrate benefits from IFFL circuits, underscoring their therapeutic potential.

A potential concern with miRNA IFFL circuits is the possibility of off-target effects of the miRNA. However, observations here and in previous work suggest such effects can be minimal. First, in previous studies using shRNAs, toxicity was observed using a Pol III based promoter, but not for a Pol II based promoter like the one used in this study, suggesting that off-target effects require higher expression levels [65]. The promoter we use is 2 orders of magnitude lower in expression level than the popular CAG promoter (**Figure 2.3B**), which is similar to the CMV promoter used in the Ehlert et al. study that yielded safe Pol II based synthetic miRNAs [65]. Also, Ross et al. showed reduction in toxicity scores and significant phenotypic benefit using a promoter to express synthetic miRNA at higher levels than those in this paper [64]. Second, mice that received high doses of the regulated, miRNA-producing construct did not exhibit apparent lifespan deficits or obvious signs of

severe toxicity. Future designs could further optimize miRNA sequences to match ectopic MeCP2 protein levels to endogenous levels, and minimize in vivo off-target effects on gene expression. In the future, it will be important to develop genome-wide assessments of on- and off-target circuit effects over timescales comparable to animal lifespan.

Finally, we note that the compactness of IFFL dosage compensating circuit makes it broadly compatible with AAV-based gene therapies for other genetic diseases, such as SMA, Angelman syndrome, Fragile X syndrome, and monogenic autism spectrum disorders [65, 66]. It could also extend the durability of gene therapies, by permitting higher initial vector copy numbers, and therefore a longer duration of therapeutic protein expression, while avoiding overexpression, in dividing, diluting cells. Similar circuits could also improve safety and efficacy of other therapies based on ectopic expression, such as gene editing [67]. We therefore anticipate increasing use of synthetic regulatory circuits to enable more accurate, precise, and robust control in gene therapies and other emerging therapeutic approaches.

2.4 Methods

Modeling an idealized dosage-compensating IFFL

To understand the functional behavior of the IFFL circuit, we considered a simple model of IFFL-regulated expression, in which a gene produces both a mRNA gene product, G, and its negative regulator (miRNA), R, which acts catalytically to degrade the target mRNA. We assume both species are removed through a combination of degradation and dilution with rate constants of γ_G and γ_R, respectively. We denote the expression level of G, in units of molecules per cell, produced by a single unregulated gene copy as P. This means that at steady state, P molecules are produced per cell during the mean life-time, $\tau = \gamma_G^{-1}$, of the gene product. In other words, the production rate for a single copy of the gene is $P/\tau = P\gamma_G$ molecules per cell per unit time. When there are multiple copies of the gene, we introduce D to denote the gene dosage, i.e., the number of gene copies per cell. With D gene copies, the production rate is $DP\gamma_G$.We also assume that R is produced proportionally to G, with α being the stoichiometric ratio of R to G production rates. Finally, we assume that the miRNA catalytically degrades G at a rate kGR. This assumes that the gene product is far from saturation, i.e., $G \ll K_M$ where K_M denotes the Michaelis constant for miRNA-mediated mRNA degradation. Together, these reactions can be described by a pair of ordinary differential equations:

$$\frac{dG}{dt} = DP\gamma_G - \gamma_G G - kGR, \tag{2.2}$$

$$\frac{dR}{dt} = \alpha DP\gamma_G - \gamma_R R. \tag{2.3}$$

Solving for the steady state of this system, $\frac{dG}{dt} = \frac{dR}{dt} = 0$, gives

$$G = \frac{\gamma_R}{k\alpha} \frac{DP}{\frac{\gamma_R}{k\alpha} + DP}. \tag{2.4}$$

When DP is much less than $\gamma_R/k\alpha$, the circuit has a negligible effect, and the steady-state expression level, G, is approximately equal to DP, the unregulated expression level. When DP approaches or exceeds $\gamma_R/k\alpha$, expression approaches a constant limiting value $G_* = \gamma_R/k\alpha$, insensitive to both D and P. The ideal IFFL thus caps expression at the value $\gamma_R/k\alpha$. This cap can be decreased by strengthening the catalytic rate constant of the catalytic repression (increasing k), by decreasing the repressor degradation rate γ_R, or by increasing the stoichiometric ratio α.

Modeling a dosage-compensating miRNA IFFL with more realistic two-step catalysis

A more realistic model of a miRNA IFFL can be obtained by including a two-step model of degradation. In this model the mRNA G and RISC complex R reversibly form a complex C with an on-rate k_{on} and off-rate k_{off}, and then the complex cleaves the mRNA at a rate k_{cat}, leaving just R. This set of reactions can be mathematically described by the following system of ordinary differential equation:

$$\frac{dG}{dt} = DP\gamma_R - \gamma_G G - k_{on} GR + k_{off} C, \tag{2.5}$$

$$\frac{dR}{dt} = \alpha DP\gamma_G - \gamma_R R - k_{on} GR + k_{off} C + k_{cat} C, \tag{2.6}$$

$$\frac{dC}{dt} = k_{on} GR - k_{off} C - k_{cat} C. \tag{2.7}$$

At steady state, the unbound RISC complex R_{ss} has concentration

$$R_{ss} = \frac{\alpha DP\gamma_G}{\gamma_R}, \tag{2.8}$$

the unbound mRNA G_{ss} has concentration

$$G_{ss} = \frac{K_M \gamma_R}{k_{cat}\alpha} \frac{DP}{\frac{K_M\gamma_R}{k_{cat}\alpha} + DP} \tag{2.9}$$

where $K_M = (k_{cat} + k_{off})/k_{on}$.

Then the total mRNA concentration is the sum of the unbound mRNA and the mRNA bound to the RISC:

$$G_{total} = G_{ss} + C_{ss} = G_{ss}\left(1 + \frac{\alpha\gamma_G}{K_M\gamma_R} DP\right) = \frac{K_M\gamma_R}{k_{cat}\alpha} \frac{DP}{\frac{K_M\gamma_R}{k_{cat}\alpha} + DP}\left(1 + \frac{\alpha\gamma_G}{K_M\gamma_R} DP\right). \tag{2.10}$$

As a function of DP, the term outside the parentheses approaches an asymptote of $K_M\gamma_R/k_{cat}\alpha$ when $DP > K_M\gamma_R/k_{cat}\alpha$. Inside the parentheses, the term is roughly equal to unity until $DP >$

$K_M\gamma_R/\alpha\gamma_G$, when it becomes linear. Thus while $K_M\gamma_R/k_{cat}\alpha < DP < K_M\gamma_R/\alpha\gamma_G$ the curve should be roughly flat, before returning to linearity, with a k_{cat}/γ_G-fold range of dosage invariant expression. Since the degradation rate of mRNA γ_G is orders of magnitude slower than the k_{cat} of the RISC complex, there should be dosage invariance over multiple orders of magnitude in DP, as we observe in experiments (**Figure 2.2C**).

Stochastic simulations

To simulate the miRNA IFFL circuit in the presence of Poissonian viral uptake and high ectopic promoter strength, we assumed the same reactions as the ODE model above. In place of D, we assume a given cell has N viral genomes, where N is drawn from a Poisson distribution. To accurately simulate bursty transcription, of these N genomes, we assume $n \leq N$ copies are in a transcriptionally active state at any given time, while the remaining $N - n$ copies are off (silent). We assume individual viral copies switch from the off to the on state at a rate of l and from on to off at a rate of μ. The total set of reactions are displayed in Table 2.1.

Table 2.1: Gillespie Simulation reactions

Reaction	**Rate**
$(G, R) \rightarrow (G + 1, R + 1)$	$nD\gamma_G\mu/l$
$G \rightarrow G + 1$	γ_G
$R \rightarrow R + 1$	γ_R
$G \rightarrow G - 1$	kRG
$n \rightarrow n + 1$	$(N - n)l$
$n \rightarrow n - 1$	$n\mu$

Table 2.2: Gillespie Simulation: Conditions and reaction rates

Condition	D(#)	γ_G (hour^{-1})	γ_R (hour^{-1})	l (hour^{-1})	μ (hour^{-1})	k (#/hour^{-1})	Mean N(#)
Endogenous	10	1/8	1/24	2.5/8	20	0	1 (fixed)
Low transduction (unregulated)	100	1/8	1/24	1	1	0	0.69
Medium transduction (unregulated)	100	1/8	1/24	1	1	0	3
High transduction (unregulated)	100	1/8	1/24	1	1	0	9
Low transduction (regulated)	100	1/8	1/24	1	1	1/160	0.69
Medium transduction (regulated)	100	1/8	1/24	1	1	1/160	3
High transduction (regulated)	100	1/8	1/24	1	1	1/160	9

The code (available online[1]) conducts a stochastic (Gillespie) simulation of this system [68]. The code was run for 2,000,000 samples for each of the distributions in Table 2.2. All rates are assumed to be per hour.

Vector Constructs and Cloning

Vector constructs were made from several components by PCR cloning, de novo synthesis, and InFusion Snap assembly (Takara #638948). The vector backbone containing antibiotic resistance and AAV ITRs came from AAV:ITR-U6-sgRNA(backbone)-hSyn-Cre-2A-EGFP-KASH-WPRE-shortPA-ITR, a gift from Feng Zhang (Addgene plasmid #60231; http://n2t.net/addgene:60231; RRID:Addgene_60231) [69]. The ECM promoter, mRuby3, synthetic polyadenylation sequences, and miRNA cassette were cloned from various in house plasmids or de novo synthesized via IDT gBlocks and inserted with InFusion Snap Assembly over several intermediate steps.

Transient transfection

For imaging experiments, glass bottom plates were coated overnight by diluting human laminin (ThermoFisher #A29249) to a surface density of 0.5 μ g/cm^2. The next day, U2OS cells were seeded at a density of 50,000 cells in each well of a 24 well plate, either Falcon Polystyrene Microplates (Fisher Scientific #08-772-1) for flow cytometry or MatTek No 1.5 uncoated glass bottom plates (Fisher Scientific #NC9988706) for smFISH imaging experiments, and cultured under standard conditions overnight. The following day, the cells were transiently transfected using Fugene HD (Promega #E2311), according to the manufacturer's protocol, and incubated for 48 hours before analysis.

Flow cytometry

Cells in a 24-well plate were trypsinized with 50 μL of 0.25% Trypsin (ThermoFisher #25200056) for 5 minutes at 37 ℃. Next, 150 μ L of HBSS with 2.5 mg/mL BSA and 1 mM EDTA was used to wash down the cells. Cells were then filtered through a 40 μm cell strainer (Fisher Scientific #08-771-1) into a 96-well round bottom plate (Fisher Scientific #08-772-17). This plate was then analyzed with a CytoFLEX flow cytometer.

In vitro smFISH

To analyze *Mecp2* expression in brain slices, we used a previously published smFISH protocol [54].

Probe Design: 27 primary probes were constructed by tiling the EGFP segment of the *MeCP2-EGFP* RNA, extending the length of each probe until the melting temperature reached 42 Celsius in the presence of 50% formaldehyde according to the MeltingTemp utility in the BioPython

[1] https://github.com/labowitz/DosageCompensatedGeneTherapy/blob/main/Fig1E%20Gillespie%20simulation.ipynb

package. Each probe was then flanked with two copies (one each side) of a readout sequence 'GCGCAATAAACCCTA' separated by a 'TT' element. Readout probes complementary to the readout sequence were ordered separately, conjugated to Alexa594. For visualizing the cytoplasm, oligo(dT) (30 'T' nucleotides) was conjugated to Alexa Fluor 405.

Protocol: Protocol: Cells were washed twice with PBS before fixation with 4% formaldehyde for 10 minutes at room temperature. Cells were then washed twice with PBS and then permeabilized with cold 70% ethanol and incubated at 4 ℃ for 1 hour. The ethanol was then aspirated and the cells were allowed to air dry. The cells were then washed 3 times with 2xSSC and incubated overnight at 37 ℃ with 50% hybridization buffer (5 mL formamide, 4 mL Ultrapure RNase/DNase free water, 1 mL 20x SSC, and 1 g of high MW dextran sulfate Sigma-Aldrich #D8906) with 10 nM of smFISH probes. The next day, the cells were washed 4 times with 30% wash buffer (3 mL formamide, 5.9 mL Ultrapure RNase/DNase free water, 1 mL 20xSSC, and 0.1 mL Triton-X) at 37 ℃ and then once with 4x SSC at room temperature.

To add the readout probes, 50 nM readout probes in 10% EC Buffer (1 mL of ethylene carbonate, 7 mL Ultrapure RNase/DNase free water, 2 mL 20x SSC, and 1 g Low MW dextran sulfate Sigma #D4911) was added to the cells and incubated for 20 minutes at room temperature. After, the cells were washed 3 times with 10% wash buffer (1 mL formamide, 7.9 mL Ultrapure RNase/DNase free water, 1 mL 20xSSC, and 0.1 mL Triton-X) for 5 minutes at room temperature. Then, the cells were washed briefly with 4xSSC, then 2xSSC, and then the cells were incubated with 500 nM oligo(dT)-Alexa405 in 2xSSC for 1 hour at room temperature. Next, the cells were washed with 2xSSC twice, and then twice with antibody staining buffer (50 mL PBS, 0.5 g Ultrapure BSA, 1.13 g glycine, 50 μL Tween 20). Then, the cells were incubated with antibody staining buffer for 30 minutes for blocking. Then, the cells were incubated with 0.5 μg/mL Alexa-Fluor-647-conjugated Na/K ATPase antibody (Abcam #ab196695) in antibody staining buffer for 2 hours at room temperature. Finally, the cells were washed twice with antibody staining buffer, and twice with 2xSSC, and mounted in 50 μL of Prolong Gold (ThermoFisher #P10144) before imaging.

Bulk RNA sequencing sample preparation

To assess off-target effect of the miRNA cassette, U2OS cells were plated on 6-well plates with 300,000 cells per well. Cells were transfected the following day with 1,000 ng of either the control plasmid or the BFP-miR-L plasmid using Fugene HD (Promega #E2311) according to the manufacturer's instructions. Media was replaced with 2 mL of fresh media 24 hours post-transfection. Cells were harvested 48 hours post-transfection by digestion with 0.25% Trypsin-EDTA, centrifugation at 300 g for 5 minutes, and removal of the supernatant by aspiration. The cell pellet was stored in -80 ℃ prior to the purification. RNA was extracted using the RNeasy kit

(Qiagen #74106) according to the manufacturer's instructions. RNA was treated with Turbo DNase (ThermoFisher #AM2238) and purified using the RNeasy kit RNA cleanup protocol. mRNA sequencing libraries were prepared by Novogene.

Differential gene expression analysis

To characterize the perturbations that synthetic miRNA brought to the endogenous transcriptome, differential expression analysis was performed using DESeq2 (1.40.1) [70] in R (4.3.1), comparing transcript counts in miRNA-transfected cells and BFP-only cells. Preprocessing and further analysis of RNA expression was performed as in ref. [71].

AAV production and purification for in vivo assessment of constructs

Plasmids used for AAV preparation include the single stranded (ss) rAAV genomes containing the cassettes described in the main text, pHelper (Addgene) and plasmids encoding AAV-CAP.B22 [55, 72]. Viruses were prepared through triple-transient transfection in adherent HEK293 cell culture (ATCC) and purified by ultracentrifugation as previously described [73]. Viruses were concentrated in sterile saline for injection into rodents, and viral titers were measured by qPCR.

Animals

All experiments involving animals were approved by the Institutional Animal Care and Use Committee (IACUC) and Office of Laboratory Animal Research (OLAR) at the California Institute of Technology. The animals used in this study are the Rett syndrome model commonly known as *Mecp2*-null (Jackson Laboratories 003890). The colony was maintained on a C57Bl/6J background for compatibility with the engineered AAV capsids used in this study. Mice in behavioral cohorts were group housed when possible at 71-75 °F under a reverse light cycle (12h on, 12h off) with enrichment materials such as blocks and shepherd shacks.

Intravenous AAV administration and behavioral assessment in *Mecp2*-null mice

To assess the performance of the AAV-MeCP2 candidate constructs against the disease phenotype of the *Mecp2*-null rodent model of Rett Syndrome, animals were genotyped at week 3 of life and randomized into groups controlled for sex, age and breeding pair origin. Mice in treatment groups were injected intravenously through the retro-orbital vein as previously described [73]. Heterozygous females and hemizygous males were divided into balanced cohorts to receive no injection (n=12; 6 male and 6 female), or were dosed by weight to receive 1e14 viral genomes/kilogram of the AAV-CAP.B22-MeCP2 unregulated (n=13; 5 male and 8 female), 1x (1 target site; n=13; 6 male and 7 female) or 4x (4 target sites; n=11; 5 male and 6 female) *Mecp2* therapeutic constructs. Experimental blinding was then conducted by a third party. These mice, along with their wildtype littermates, were assessed at baseline for motor and neurological performance using the RTT Phe-

notype Scoring Scale as previously described [24], Testing continued until the study end at 18 or 24 weeks post injection, or until humane endpoint criteria were reached, resulting in withdrawal of an animal from the study. All behavioral scoring was done individually in a large plastic arena outside of the home cage, and was performed during the dark phase under red light. We observed a phenotype increase earlier than some studies [24, 58], but consistent with timing observed in others [74].

Mecp2 RNA histological analysis

Following 3 weeks of expression, injected rodents were deeply sedated via intraperitoneal injection of Euthasol (pentobarbital sodium and phenytoin sodium solution, Virbac AC) prior to cardiac perfusion using RNAse-free, heparinized saline and 4% paraformaldehyde (PFA) in 0.1 M phosphate buffered saline (PBS). Tissues were post-fixed in 4% PFA for 24-48 hours and either sliced on a vibratome at 50 μM for immediate analysis or cryoprotected with RNAse-free 10% and 30% sucrose solution, frozen in OCT and stored at -80 ℃. Prior to analysis, tissues were sliced to a thickness of 50 μM on a Leica cryostat. Tissue slices collected for FISH analysis were incubated in ice cold RNAse-free 70% ethanol prior to probing.

Hybridization chain reaction

Probes and Buffers: Probe sets were ordered from Molecular Technologies[2] against the coding sequence of EGFP and endogenous mus Musculus Mecp2 isoform 1 (NM_001081979), with specific instructions to only include the 3' UTR for the latter. Molecular Technologies hybridization buffer, wash buffer, and amplification buffer were included as part of the order.

Protocol: Paraformaldehyde-fixed fresh or frozen brain tissue was sliced to a thickness of 50 μm, mounted onto glass coverslips and dried in a fume hood. Slices obtained from frozen brain tissue were rinsed with RNAse-free PBS to remove OCT compound prior to drying. Ice cold RNAse-free 70% ethanol was applied to the dried tissue slices. After incubating at 4 ℃ for 1 hour, the ethanol was removed and samples were dried in a fume hood. Once dry, the coverslips were washed with PBS before incubating in 8% SDS in PBS for 20 minutes at room temperature. The SDS solution was then poured off and the coverslips were rinsed 3 times with PBS for 2 minutes. The area of the coverslips around the tissue slices was then dabbed dry with a Kimwipe before mounting a SecureSeal hybridization chamber (Grade Biolabs #621502) on top of the tissue slices. Using the hybridization chambers, the samples were washed with 5xSSCT (5xSSC with 0.1% Tween 20) and then the samples were pre-incubated with 30% HCR probe hybridization buffer for 30 minutes at 37 ℃. After this incubation, the tissue slices were incubated with 0.2 μL of 2 μM stock of each of the odd and even HCR probe mixtures for both ectopic and endogenous target in 100 μL 30%

[2]https://www.moleculartechnologies.org/

HCR probe hybridization buffer for 3 days at 37 ℃. After, the tissue slices were washed 4 times with 30% HCR probe wash buffer at 37 ℃ for 15 minutes. Then the tissue slices were washed twice with 5xSSCT for 5 minutes at room temperature. Next, the tissue slices were pre-incubated with HCR amplification buffer for 30 minutes at room temperature. In parallel, 2 μL of each of the 3 μM stocks of hairpins H1 and H2 for both endogenous and ectopic targets (4 total) was added to separate PCR tubes (4 total) and heated to 95 ℃ for 90 seconds before being allowed to cool to room temperature in a dark drawer for 30 minutes. After this was completed, the 2 μL from each of the PCR tubes was added to 100 μL of amplification buffer and the tissue slices were incubated in this solution for 90 minutes at room temperature in the dark. Once complete, the samples were washed 4 times for 15 minutes with 5xSSCT. The samples were then washed once with 2xSSC, and then the slices were incubated with 500nM oligo(dT)-Alexa405 in 2xSSC for 1 hour at room temperature. The samples were then washed 4 times with 2xSSC, and finally mounted in Prolong Gold (ThermoFisher #P10144).

Confocal microscopy

FISH and HCR samples were analyzed on a Nikon Eclipse Ti inverted microscope with a Yokogawa CSU-W1 spinning disc unit (Andor) and a Zyla 4.2 sCMOS camera (Andor), using a 60x oil immersion objective (1.4 NA). Z-stacks were taken from the bottom of the cell slice to 30 microns into the tissue, with a step size of 0.3 microns. Exposures were acquired using laser illumination with the following wavelengths and exposure times: 405 nm (500 ms), 488 nm (300 ms), 561 nm (500 ms), 594 nm (500 ms), and 640 nm (500 ms). Imaging data sets consisted of variable depth stacks of images composed of 2048x2048 0.1 μm x 0.1 μm x 0.3 μm voxels digitized to 16-bit fluorescence intensities. Raw data available at doi:10.22002/hncr2-9pt15.

Cell segmentation

Cell segmentation was performed using the Cellpose program (https://www.cellpose.org/) [75, 76] applied to the oligo(dT) signal (**Figure 2.3**) or a Na/K ATPase membrane marker (**Figure 2.3**). The basic 'cyto' model proved sufficient to accurately segment cell bodies in the vast majority of cases. The masks generated by Cellpose were then passed to the next parts of the program.

Dot counting algorithm

In **Figures 2.2D, 2.3B, 2.3C, 2.4E, and 2.4F**, dots representing single transcripts were detected using the following algorithm. First, each fluorescence image was thresholded using a minimum intensity parameter (see below). Next, the local maxima of the images were identified. Maxima that overlapped with a mask created by Cellpose were labeled as dots inside that cell. In the case of smFISH dots, cell counts were normalized to cell volume by multiplying the raw counts by the

mean cell volume divided by the individual cell's volume, where cell volume is defined by the number of voxels contained within the CellPose mask defining the cell.

In brain samples, autofluorescence also exhibited dots. However, the broader spectrum of autofluorescence caused these to be detected in multiple fluorescence channels, unlike dots associated with specific transcripts, which predominantly appeared only in a single fluorescence channel. To avoid miscategorizing these as transcripts, any dots that occurred at the same point in the 594 and 640 channels were labeled as autofluorescent background and discarded.

Images of these cells are included as Image Analysis Supplements (**Supplementary Material**), each row of which depicts an individual cell. The oligo(dT) images depict a maxproject of the oligo(dT) signal within the z-range of the cell boundaries, with a white outline depicting the segmentation boundary at the z-level of the centroid of the segmented cell. For HCR, images depicting endogenous and ectopic transcripts are next, followed by an overlay of both. In these images, masks are taken at each z-level of the segmented cell, and then a max-project of the masked z-slices is taken. Dots in each channel are denoted by white circles, while background dots are denoted by cyan circles. Autofluorescent puncta are not true dots, and show up in the overlay. Some dots are missed either due to being too dim, or due to having their local maxima outside the segmentation boundary (**Figure 2.8E**). The receiver-operating-characteristic (ROC) analysis (below) showed that there will be a small false negative rate.

Optimizing dot counting threshold

The accuracy of the dot counting method was sensitive to the fluorescence threshold. We therefore sought to optimize the threshold value by comparison to a ground truth. To establish a ground truth data set, we manually analyzed 4 cells expressing the unregulated construct and 10 negative control cells for which no fluorescent secondary probes were hybridized. We then ran the image analysis pipeline on these samples, scanning the threshold value from a pixel intensity of 100 to 1100. The number of dots identified at each threshold value was compared to the ground truth values. After scanning the threshold, we applied a receiver-operating-characteristic (ROC) analysis to identify an optimal threshold value (**Figure 2.8A, 2.8B**). More specifically, we compared the number of true positives and false positives for each threshold. Taking P to be the ground truth number of positive dot calls, TP as the number of true positive dots identified by the algorithm, and FP as the number of false positive dots, we searched for the threshold value that minimized $(P - (TP + FP))^2$, i.e., the squared error in the total dots called (**Figure 2.8C, 2.8D**).

Correcting dense dot counts

In some cells, high densities of HCR dots made it impossible to distinguish individual dots. Because of this, the dot counts saturated, so that even if the total dot fluorescence in the cell increased, there

was no increase in dot count. To correct for this regime, we corrected dot counts if they were past the threshold of fluorescence beyond which there was little change in dot count, and assigned them a count that was dependent on a linear function of the total dot fluorescence (**Figure 2.5C**). To get this linear function, we find the median dot count n_0 of cells with fluorescence close to the threshold f_0, and beyond that threshold assign cells an adjusted count n as a linear function of their observed fluorescence, f :

$$n = f n_0 / f_0. \tag{2.11}$$

Bootstrap significance testing

To test the statistical significance of the difference between the means of two sets of samples, we followed a standard bootstrap hypothesis testing procedure. Briefly, samples from both groups were pooled into a single set. Two new sets with the same size as the original sets were then constructed by sampling with replacement from the pooled set and the means of the resampled sets were compared. This was repeated 105 times, and the proportion of resampled set pairs with means that differed by more than the observed difference of the original pair was reported as the p-value.

2.5 Data availability

Raw data is available at doi:10.22002/hncr2-9pt15. Source code for the dot counting pipeline is available at: GitHub - labowitz/DosageCompensatedGeneTherapy.

2.6 Supplementary figures

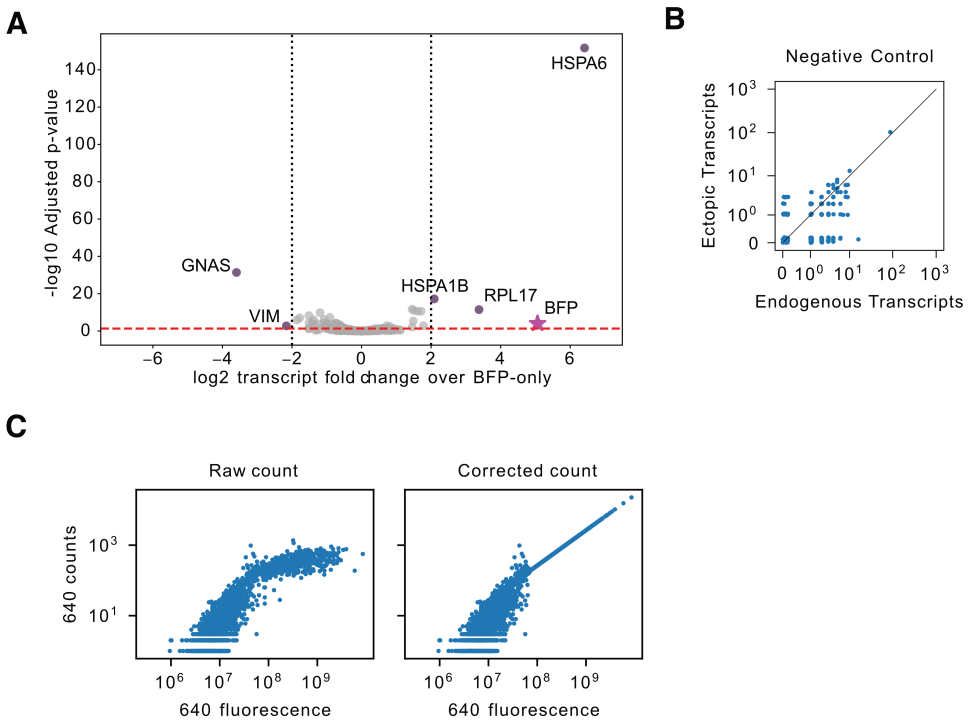

Figure 2.5: **miRNA off-target analysis, HCR negative controls, and HCR saturation correction.** **(A)** Volcano plot of transcriptomic changes in U2OS cells transfected with mTagBFP2–miRNA compared to a control plasmid lacking the miRNA (bulk RNA-seq). The x-axis shows $\log_2$ fold change in expression, the y-axis shows $-\log_{10}$ adjusted p-value. The dashed black lines indicate where the absolute value of the $\log_2$ fold change exceeded a value of 2. The red dashed line indicates an adjusted p-value of 0.05. Six genes (purple dots) showed significantly differential expression relative to the construct without miRNA, possibly due to stress from overexpression of transfected BFP-miRNA (star). No upregulated or downregulated transcript contained target sites for the miRNA. **(B)** HCR analysis of ectopic Mecp2 (y-axis) vs endogenous Mecp2 (x-axis) in negative control neurons corresponding to **Figures 2.3B and 2.3C**. Each dot represents one individual cell. Plotted values represent number of dots identified in each channel. The no-primary negative control had no probes against Mecp2 added and thus no true HCR dots. 48% of cells have either 1 or no counts in both the ectopic and endogenous channels. **(C)** Dot counts in individual cells as a function of total dot fluorescence both before ('Raw count') and after ('Corrected count') correcting for highly dense cell dots (**Methods**).

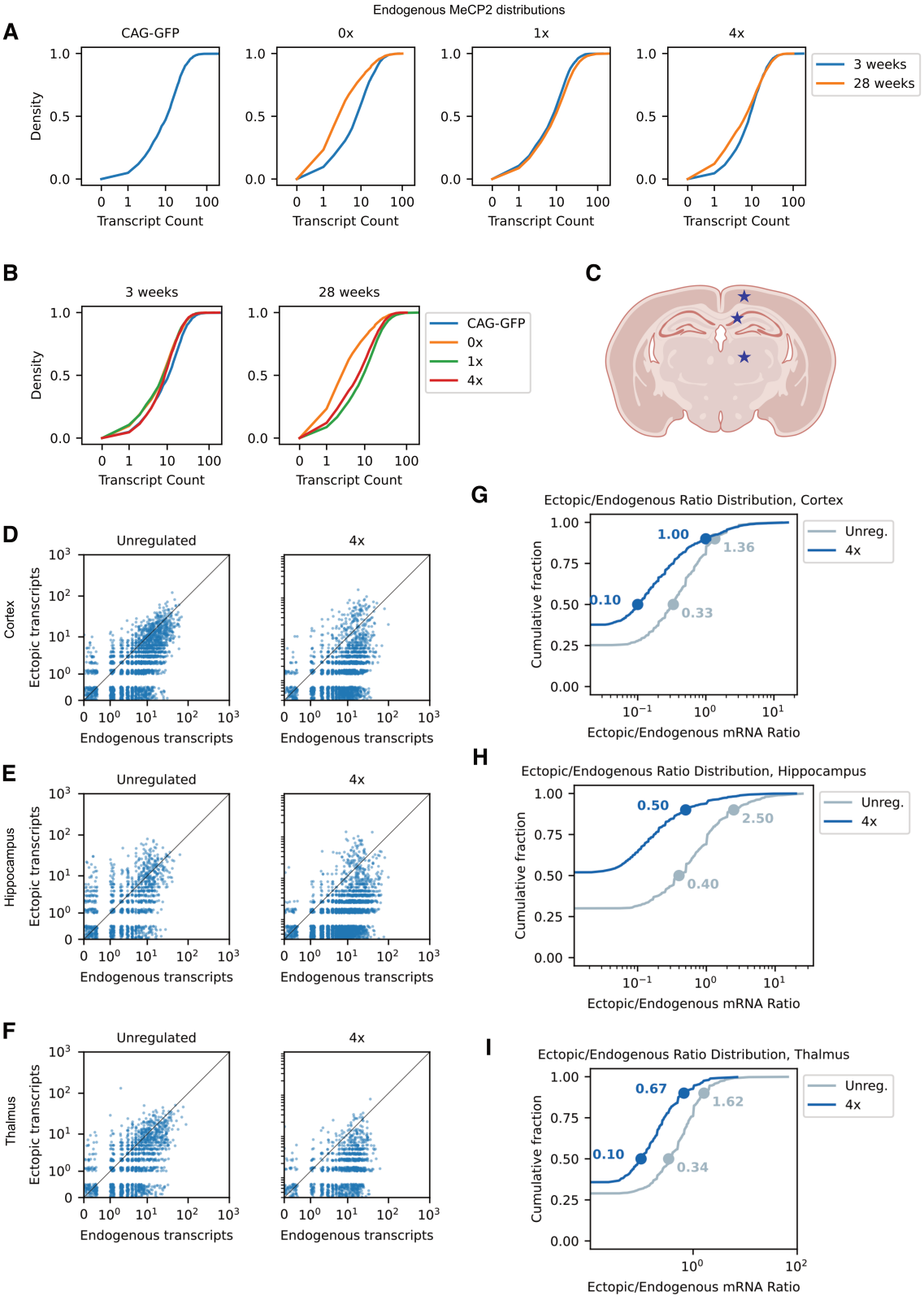

Figure 2.6: (Caption on next page)

Figure 2.6: **Mouse Brain HCR Supplement.** **(A)** ECDFs of endogenous Mecp2 in mice at 3 and 28 weeks after AAV delivery for the datasets in **Figures 2.3B, 2.3C, 2.4E, 2.4F**. Each plot compares different time points for the indicated construct. **(B)** ECDFs of endogenous Mecp2 in mice at 3 and 28 weeks after AAV delivery for the datasets in **Figures 2.3B, 2.3C, 2.4E, 2.4F**. Each plot compares the different constructs at the indicated time point. **(C)** Stars denote the locations in the brain sampled in panels D-I: cortex, hippocampus, and thalamus. **(D-F)** Quantified HCR dots for ectopic Mecp2 (y-axis) vs endogenous Mecp2 (x-axis) for individual neurons in the cortex (D), hippocampus (E), and thalamus (F). The left plot shows the scatter for mouse brains that had received the unregulated construct. The right plot shows those that received the 4x construct. **(G-I)** ECDFs of the ectopic-to-endogenous Mecp2 HCR dot ratio in individual neurons in the cortex (G), hippocampus (H), and thalamus (I). The 4x construct ECDFs were shifted to lower ectopic-to-endogenous ratios relative to the unregulated constructs.

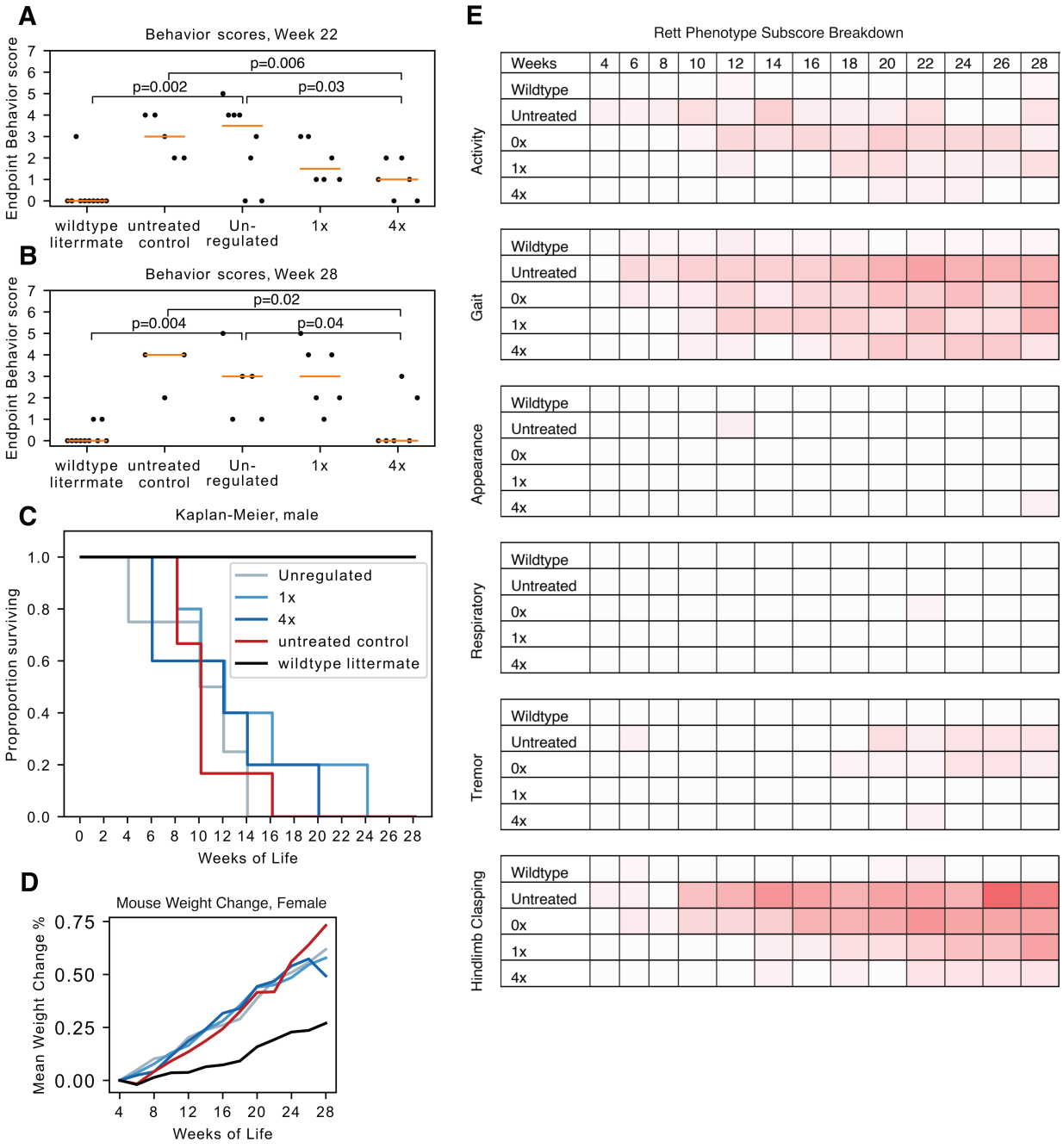

Figure 2.7: (Caption on next page)

Figure 2.7: **Mouse Behavior Supplement.** **(A)** Endpoint Rett behavior scores at week 22. Dots indicate behavior scores for individual mice, orange lines indicate median behavior score by group. Mice that received the 4x construct scored significantly lower than uninjected controls (bootstrap p=0.006) and the mice that received the unregulated construct (bootstrap p=0.03). The standard deviation of the 4x group's scores, 0.82, was significantly lower than the standard deviation of the unregulated group's scores, 1.79 (bootstrap p=0.03). Wildtype littermates scored significantly lower than the uninjected controls (bootstrap p=0.001) and mice that received the unregulated construct (bootstrap p=0.002). No significant difference was observed between wildtype and 4x mice (bootstrap p=0.08). **(B)** Endpoint Rett behavior scores at week 28, as in A. Mice that received the 4x construct scored significantly lower than uninjected controls (bootstrap p=0.02) and the mice that received the unregulated construct (bootstrap p=0.04). The standard deviation of the 4x scores, 1.21, was not significantly lower than the standard deviation of the unregulated scores, 1.49 (bootstrap p=0.29). Wildtype littermates scored significantly lower than the uninjected control (bootstrap p=0.002) and mice receiving the unregulated construct (bootstrap p=0.004). The scores of wildtype and 4x mice were not significantly different (bootstrap p=0.08). **(C)** Kaplan-Meier survival curves for male Rett mice. Mean lifespans were not significantly different between treatment groups. **(D)** Mean percentage weight change over time in female mice. All mutant cohorts gained weight similarly, indicating the obesity phenotype was unaffected by the gene therapy. **(E)** Subscore breakdown of Rett Phenotype score over time. Red shading intensity is proportional to severity for each category.

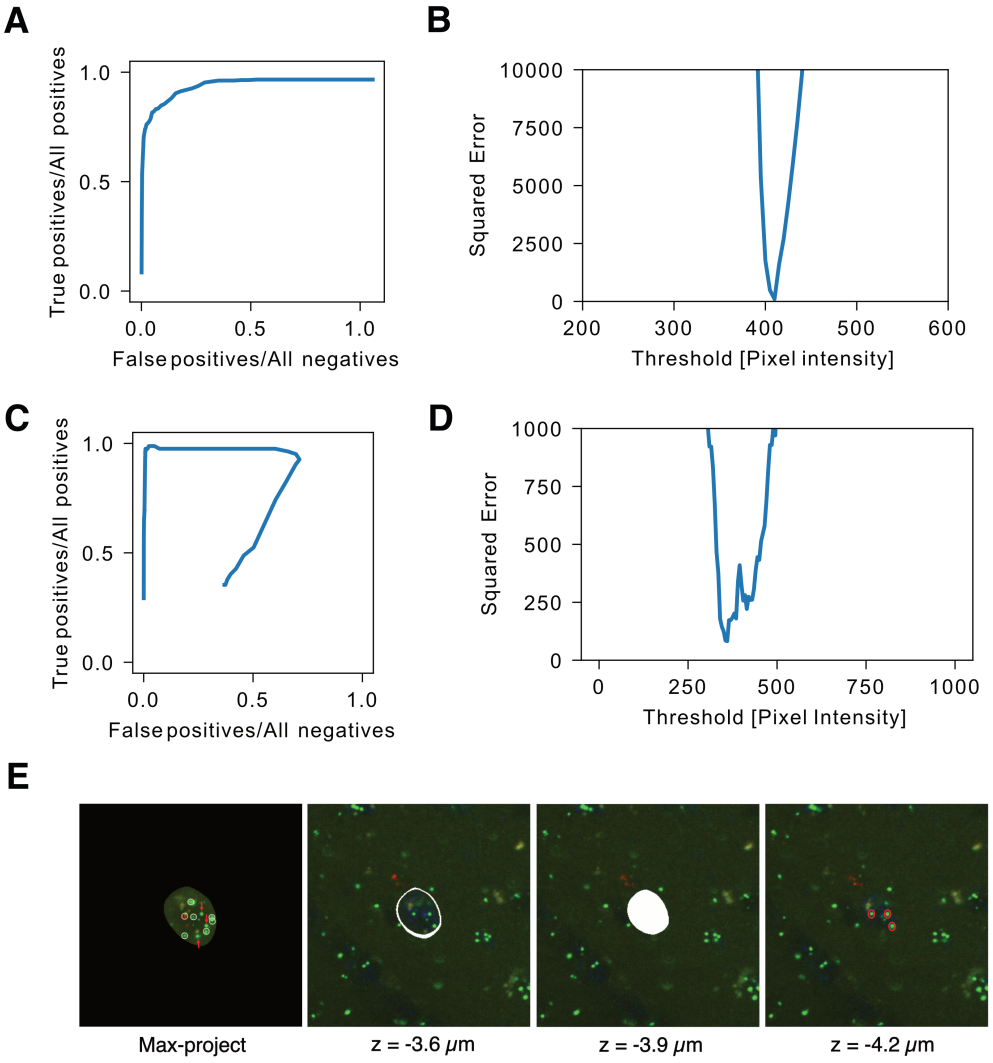

Figure 2.8: (Caption on next page)

Figure 2.8: **Optimization of Dot Detection.** **(A)** True positive rate vs false positive rate as a function of the dot counting threshold, which increases from pixel intensity 100 (top right) to 1100 (bottom left), in U2OS cells analyzed with smFISH. True positives and false positives were defined relative to manually annotated ground truth images. An ideal detection algorithm would count all the true positives and no false positives, i.e., would lie in the upper left corner. Some thresholds give results close to this ideal corner, and define the optimal values of the threshold, but perfect classification is not possible. **(B)** The squared difference between the total detected counts and the total true counts (defined by manual annotation) as a function of threshold, in data from (A), showed a narrow window of acceptable thresholds that give a low error around a threshold pixel intensity of 410. **(C)** True positive rate vs false positive rate as a function of the dot counting threshold, which increases from 100 (blue curve, top right) to 1100 (blue curve, bottom left), in mouse brain slices analyzed by HCR. The true positive and false positive rate were defined relative to manually annotated ground truth images. In contrast to the smFISH data, a significant component of the background consisted of autofluorescent puncta prevalent in the brain. To correct for this, false positive dots were eliminated by discarding dots that appeared in both channels. This led to a decrease in true positive rate and false positive rate at low thresholds, leading to the curve away from the top right. Some threshold values closely approached a true positive rate of 1 and false positive rate of 0 (upper left corner). **(D)** The squared difference between the detected counts and the total true counts as a function of threshold in data from (C), showed a window of threshold that had minimal error around a pixel intensity of 360. **(E)** In the Image Analysis Supplements, some dots appear to not be counted (max-project, red arrows point to missed dots). In some of these cases, this is due to the dot intensity maximum lying outside of the range of z-slices encompassed by the cell. Z-levels of -3.6 μm, -3.9 μm, and -4.2 μm relative to the cell centroid are displayed here. The solid white mass indicates the "bottom" of the detected cell at slice z = -3.9 μm. However the maxima of the missed dots are at z = -4.2 μm or below.

References

[1] J. R. Mendell et al. "Single-dose gene-replacement therapy for spinal muscular atrophy." *New England Journal of Medicine* 377.18 (2017), pp. 1713–1722.

[2] A. C. Nathwani et al. "Long-term safety and efficacy of factor IX gene therapy in hemophilia B." *New England Journal of Medicine* 371.21 (2014), pp. 1994–2004.

[3] L. A. George et al. "Hemophilia B gene therapy with a high-specific-activity factor IX variant." *New England Journal of Medicine* 377.23 (2017), pp. 2215–2227.

[4] M. C. Ozelo et al. "Valoctocogene roxaparvovec gene therapy for hemophilia A." *New England Journal of Medicine* 386.11 (2022), pp. 1013–1025.

[5] S. Russell et al. "Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: A randomised, controlled, open-label, phase 3 trial." *The Lancet* 390.10097 (2017), pp. 849–860.

[6] R. C. Challis et al. "Adeno-associated virus toolkit to target diverse brain cells." *Annual Review of Neuroscience* 45.1 (2022), pp. 447–469.

[7] J. R. Mendell et al. "Current clinical applications of in vivo gene therapy with AAVs." *Molecular Therapy* 29.2 (2021), pp. 464–488.

[8] D. Samanta et al. "Current and emerging precision therapies for developmental and epileptic encephalopathies." *Pediatric Neurology* (2025).

[9] L. Levinson. "FDA clears CAP-003 Parkinson's gene therapy to be tested in trials." *Parkinson's News Today* (2025). URL: https://parkinsonsnewstoday.com/news/fda-clears-cap-003-parkinsons-gene-therapy-tested-trials/.

[10] R. Novy. "STXBP1 foundation president and executive eirector comment on Capsida's FDA approval to start gene therapy clinical trial." *STXBP1 Foundation* (2025). URL: https://www.stxbp1disorders.org/blog/capsida-progress-updates-on-cap-002-stxbp1-dee.

[11] M. H. Berryer et al. "Mutations in SYNGAP1 cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency." *Human Mutation* 34.2 (2013), pp. 385–394.

[12] G. Rumbaugh et al. "SynGAP regulates synaptic strength and mitogen-activated protein kinases in cultured neurons." *Proceedings of the National Academy of Sciences* 103.12 (2006), pp. 4344–4351.

[13] S. E. Smith et al. "Increased gene dosage of Ube3a results in autism traits and decreased glutamate synaptic transmission in mice." *Science Translational Medicine* 3.103 (2011), 103ra97–103ra97.

[14] M. Van Alstyne et al. "Gain of toxic function by long-term AAV9-mediated SMN overexpression in the sensorimotor circuit." *Nature Neuroscience* 24.7 (2021), pp. 930–940.

[15] C. Hinderer et al. “Severe toxicity in nonhuman primates and piglets following high-dose intravenous administration of an adeno-associated virus vector expressing human SMN.” *Human Gene Therapy* 29.3 (2018), pp. 285–298.

[16] J. Hordeaux et al. “Adeno-associated virus-induced dorsal root ganglion pathology.” *Human Gene Therapy* 31.15-16 (2020), pp. 808–818.

[17] R. E. Amir et al. “Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.” *Nature Genetics* 23.2 (1999), pp. 185–188.

[18] R. Tillotson and A. Bird. “The molecular basis of MeCP2 function in the brain.” *Journal of Molecular Biology* 432.6 (2020), pp. 1602–1623.

[19] L. Villard et al. “Two affected boys in a Rett syndrome family: clinical and molecular findings.” *Neurology* 55.8 (2000), pp. 1188–1193.

[20] B. Hagberg et al. “A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett’s syndrome: Report of 35 cases.” *Annals of Neurology* 14.4 (1983), pp. 471–479.

[21] D. Tropea et al. “Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice.” *Proceedings of the National Academy of Sciences* 106.6 (2009), pp. 2029–2034.

[22] A. L. Collins et al. “Mild overexpression of MeCP2 causes a progressive neurological disorder in mice.” *Human Molecular Genetics* 13.21 (2004), pp. 2679–2689.

[23] M. B. Ramocki et al. “Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome.” *Annals of Neurology* 66.6 (2009), pp. 771–782.

[24] J. Guy et al. “Reversal of neurological defects in a mouse model of Rett syndrome.” *Science* 315.5815 (2007), pp. 1143–1147.

[25] K. K. Gadalla et al. “Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice.” *Molecular Therapy* 21.1 (2013), pp. 18–30.

[26] K. K. Gadalla et al. “Development of a novel AAV gene therapy cassette with improved safety features and efficacy in a mouse model of Rett syndrome.” *Molecular Therapy Methods & Clinical Development* 5 (2017), pp. 180–190.

[27] S. E. Sinnett et al. “Improved MECP2 gene therapy extends the survival of MeCP2-null mice without apparent toxicity after intracisternal delivery.” *Molecular Therapy Methods & Clinical Development* 5 (2017), pp. 106–115.

[28] V. Matagne et al. “A codon-optimized Mecp2 transgene corrects breathing deficits and improves survival in a mouse model of Rett syndrome.” *Neurobiology of Disease* 99 (2017), pp. 1–11.

[29] R. Tillotson et al. “Radically truncated MeCP2 rescues Rett syndrome-like neurological defects.” *Nature* 550.7676 (2017), pp. 398–401.

[30] V. Matagne et al. “Severe offtarget effects following intravenous delivery of AAV9-MECP2 in a female mouse model of Rett syndrome.” *Neurobiology of Disease* 149 (2021), p. 105235.

[31] S. E. Sinnett et al. “Engineered microRNA-based regulatory element permits safe high-dose mini MECP2 gene therapy in Rett mice.” *Brain* 144.10 (2021), pp. 3005–3019.

[32] A. C. Nathwani et al. “Long-term safety and efficacy following systemic administration of a self-complementary AAV vector encoding human FIX pseudotyped with serotype 5 and 8 capsid proteins.” *Molecular Therapy* 19.5 (2011), pp. 876–885.

[33] S. K. Wang et al. “In situ detection of adeno-associated viral vector genomes with SABER-FISH.” *Molecular Therapy Methods & Clinical Development* 19 (2020), pp. 376–386.

[34] K.-M. Prasad et al. “Robust cardiomyocyte-specific gene expression following systemic injection of AAV: In vivo gene delivery follows a Poisson distribution.” *Gene Therapy* 18.1 (2011), pp. 43–52.

[35] C. Huichalaf et al. “In vivo overexpression of frataxin causes toxicity mediated by iron-sulfur cluster deficiency.” *Molecular Therapy Methods & Clinical Development* 24 (2022), pp. 367–378.

[36] S. Powers et al. “Novel MECP2 gene therapy is effective in a multicenter study using two mouse models of Rett syndrome and is safe in non-human primates.” *Molecular Therapy* 31.9 (2023), pp. 2767–2782.

[37] T. Frei and M. Khammash. “Adaptive circuits in synthetic biology.” *Current Opinion in Systems Biology* 28 (2021), p. 100399.

[38] S. Mangan and U. Alon. “Structure and function of the feed-forward loop network motif.” *Proceedings of the National Academy of Sciences* 100.21 (2003), pp. 11980–11985.

[39] W. Ma et al. “Defining network topologies that can achieve biochemical adaptation.” *Cell* 138.4 (2009), pp. 760–773.

[40] K. S. Love et al. “Model-guided design of microRNA-based gene circuits supports precise dosage of transgenic cargoes into diverse primary cells.” *Cell Systems* 16.6 (2025).

[41] L. Bleris et al. “Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template.” *Molecular Systems Biology* 7.1 (2011), p. 519.

[42] T. H. Segall-Shapiro, E. D. Sontag, and C. A. Voigt. “Engineered promoters enable constant gene expression at any copy number in bacteria.” *Nature Biotechnology* 36.4 (2018), pp. 352–358.

[43] G. Lillacci, Y. Benenson, and M. Khammash. “Synthetic control systems for high performance gene expression in mammalian cells.” *Nucleic Acids Research* 46.18 (2018), pp. 9855–9863.

[44] R. D. Jones et al. “An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells.” *Nature Communications* 11.1 (2020), p. 5690.

[45] M. Osella et al. “The role of incoherent microRNA-mediated feedforward loops in noise buffering.” *PLoS Computational Biology* 7.3 (2011), e1001101.

[46] T. Frei et al. “Characterization and mitigation of gene expression burden in mammalian cells.” *Nature Communications* 11.1 (2020), p. 4641.

[47] T. J. Strovas et al. “MicroRNA-based single-gene circuits buffer protein synthesis rates against perturbations.” *ACS Synthetic Biology* 3.5 (2014), pp. 324–331.

[48] M. H. Khammash. “Perfect adaptation in biology.” *Cell Systems* 12.6 (2021), pp. 509–521.

[49] A. Chamorro-Jorganes et al. “Autoregulation of glypican-1 by intronic microRNA-149 fine tunes the angiogenic response to FGF2 in human endothelial cells.” *Journal of Cell Science* 127.6 (2014), pp. 1169–1178.

[50] J. Tsang, J. Zhu, and A. van Oudenaarden. “MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals.” *Molecular Cell* 26.5 (2007), pp. 753–767.

[51] M. Megraw et al. “Isoform specific gene auto-regulation via miRNAs: A case study on miR-128b and ARPP-21.” *Theoretical Chemistry Accounts* 125.3 (2010), pp. 593–598.

[52] L. Qiu et al. “A construct with fluorescent indicators for conditional expression of miRNA.” *BMC Biotechnology* 8.1 (2008), p. 77.

[53] C. Fellmann et al. “An optimized microRNA backbone for effective single-copy RNAi.” *Cell Reports* 5.6 (2013), pp. 1704–1713.

[54] Y. Takei et al. “Single-cell nuclear architecture across cell types in the mouse brain.” *Science* 374.6567 (2021), pp. 586–594.

[55] D. Goertsen et al. “AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset.” *Nature Neuroscience* 25.1 (2022), pp. 106–115.

[56] M. Perdomini et al. “Prevention and reversal of severe mitochondrial cardiomyopathy by gene therapy in a mouse model of Friedreich’s ataxia.” *Nature Medicine* 20.5 (2014), pp. 542–547.

[57] M. Hocquemiller et al. “Adeno-associated virus-based gene therapy for CNS diseases.” *Human Gene Therapy* 27.7 (2016), pp. 478–496.

[58] J. Guy et al. “A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome.” *Nature Genetics* 27.3 (2001), pp. 322–326.

[59] S. Luikenhuis et al. “Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice.” *Proceedings of the National Academy of Sciences* 101.16 (2004), pp. 6033–6038.

[60] M. Alvarez-Saavedra et al. “Cell-specific expression of wild-type MeCP2 in mouse models of Rett syndrome yields insight about pathogenesis.” *Human Molecular Genetics* 16.19 (2007), pp. 2315–2325.

[61] B. Kerr et al. “Transgenic complementation of MeCP2 deficiency: phenotypic rescue of Mecp2-null mice by isoform-specific transgenes.” *European Journal of Human Genetics* 20.1 (2012), pp. 69–76.

[62] L. Weaving et al. “Rett syndrome: clinical review and genetic update.” *Journal of Medical Genetics* 42.1 (2005), pp. 1–7.

[63] X. Chen et al. “Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems.” *Neuron* 110.14 (2022), pp. 2242–2257.

[64] P. D. Ross et al. “Self-regulating gene therapy ameliorates phenotypes and overcomes gene dosage sensitivity in a mouse model of Rett syndrome.” *Science Translational Medicine* 17.792 (2025), eadq3614.

[65] E. M. Ehlert et al. “Cellular toxicity following application of adeno-associated viral vector-mediated RNA interference in the nervous system.” *BMC Neuroscience* 11.1 (2010), p. 20.

[66] M. Benger, M. Kinali, and N. D. Mazarakis. “Autism spectrum disorder: prospects for treatment using gene therapy.” *Molecular Autism* 9.1 (2018), p. 39.

[67] G. Liu et al. “The CRISPR-Cas toolbox and gene editing technologies.” *Molecular Cell* 82.2 (2022), pp. 333–347.

[68] D. T. Gillespie. “Exact stochastic simulation of coupled chemical reactions.” *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361.

[69] R. J. Platt et al. “CRISPR-Cas9 knockin mice for genome editing and cancer modeling.” *Cell* 159.2 (2014), pp. 440–455.

[70] M. I. Love, W. Huber, and S. Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology* 15.12 (2014), p. 550.

[71] R. Du et al. “miRNA circuit modules for precise, tunable control of gene expression.” *bioRxiv* (2024).

[72] M. R. Chuapoco et al. “Adeno-associated viral vectors for functional intravenous gene transfer throughout the non-human primate brain.” *Nature Nanotechnology* 18.10 (2023), pp. 1241–1251.

[73] R. C. Challis et al. “Systemic AAV vectors for widespread and targeted gene delivery in rodents.” *Nature Protocols* 14.2 (2019), pp. 379–414.

[74] B. E. Collins et al. “Safety and efficacy of genetic MECP2 supplementation in the R294X mouse model of Rett syndrome.” *Genes, Brain and Behavior* 21.1 (2022), e12739.

[75] C. Stringer et al. “Cellpose: A generalist algorithm for cellular segmentation.” *Nature Methods* 18.1 (2021), pp. 100–106.

[76] M. Pachitariu and C. Stringer. “Cellpose 2.0: how to train your own model.” *Nature Methods* 19.12 (2022), pp. 1634–1641.

Chapter 3

MIRNA MODULES FOR PRECISE, TUNABLE CONTROL OF GENE EXPRESSION

The following chapter is reproduced from:

[1] R. Du et al. "miRNA circuit modules for precise, tunable control of gene expression." *bioRxiv* (2024).

3.1 Summary

Accurate control of transgene expression is important for research and therapy but challenging to achieve in most settings. miRNA-based regulatory circuits can be incorporated within transgenes for improved control. However, the design principles, performance limits, and applications of these circuits in research and biotechnology have not been systematically determined. Here, combining modeling and experiments, we introduce miRNA-based circuit modules, termed DIMMERs, that establish precise, tunable control of transgene expression across diverse cell types to facilitate imaging, editing, and gene therapy. The circuits use multivalent miRNA regulatory interactions to achieve nearly uniform, tunable, protein expression over two orders of magnitude variation in gene dosage. They function across diverse cell types, and can be multiplexed for independent regulation of multiple genes. DIMMERs reduce off-target CRISPR base editing, improve single-molecule imaging, and allow live tracking of AAV-delivered transgene expression in mouse cortical neurons. DIMMERs thus enable accurate regulation for research and biotechnology applications.

3.2 Introduction

Biomedical research and biotechnology heavily rely on expression of transgenes in living cells. The ability to accurately establish transgene expression at desired levels is critically needed in many contexts. For example, in gene therapy, overexpression of therapeutic transgenes can be toxic [1]. Similarly, in gene editing and imaging applications, overexpression can reduce specificity or increase background, respectively. However, popular expression systems, including DNA transfection and AAV vectors, as well as integrating systems such as lentivirus [2] or piggyBac transposons [3], typically generate a broad range of expression levels, due to variability in the number of gene copies taken up, integrated, and expressed by each cell, as well as gene expression noise [4, 5]. Selecting individual stable clones can reduce variability, but it is time-consuming and impossible for gene therapy.

What is needed is a simple gene regulation system that could compensate for unavoidable variation in delivery and expression (**Figure 3.1A**). The ideal system should have several key features: First, it should be genetically compact for compatibility with most delivery vectors. Second, it should allow predictive tuning of the expression setpoint. Third, it should permit construction of multiple independent (orthogonal) regulation systems for simultaneous control of multiple genes. Finally, it should operate across multiple cell types (portability) (**Figure 3.1B**).

The incoherent feed-forward loop (IFFL) circuit motif provides an ideal foundation for these capabilities [6]. A circuit in which a target gene and its negative regulator are both encoded in the same DNA construct represents an IFFL-like configuration in which gene dosage, considered as an input, modulates expression of both the target and its negative regulator. When these effects effectively cancel out, target expression can asymptotically approach a constant expression level independent of gene dosage (**Figure 3.1C**, **Supplementary Modeling Text**).

Previous studies have introduced synthetic circuits based on this principle. Bleris et al. showed that an IFFL based on microRNA (miRNA hereafter) as the negative regulator could achieve dosage compensation [7]. Strovas et al. introduced a similar design incorporating a natural miRNA and multiple repeats of its binding site within the target gene, and examined its expression dynamics over several days [8], achieving dosage compensation over a 20-fold range at the cost of potential crosstalk with endogenous genes. Yang et al. later introduced an "equalizer" architecture that combined transcriptional negative feedback through the TetR protein with feed-forward miRNA regulation [9]. This generated an extended regime of strong dosage compensation, but required expression of a bacterial protein. Tradeoffs between circuit complexity and efficiency are worth exploring further. Finally, while this work was under review, Love et al. demonstrated other configurations of miRNA-based IFFL circuits for dosage compensation [10]. Nevertheless, fundamental questions have remained unclear: What sequence and circuit design principles optimize dosage compensation? Can these systems allow expression tuning, multiplexing of independent regulatory systems in the same cell, and portability across cell types and delivery modalities? Further, can they provide durable control in vivo for gene therapy applications?

To address these goals, we combined mathematical modeling, synthetic design, and quantitative circuit analysis to create a set of miRNA-based dosage compensation systems termed DIMMERs (Dosage Invariant miRNA-Mediated Expression Regulators). These circuits take advantage of multivalent miRNA regulation through the natural TNRC6 scaffold system [11]. They allow systematic tuning of expression levels by modulating the number of miRNA cassettes, numbers of target binding sites, and miRNA-target site complementarity. Further, they can be used to orthogonally regulate multiple genes in the same cell, and operate similarly across different cell types. A toolkit of ten mutually orthogonal ready-to-use expression systems can be incorporated into

diverse systems. They facilitate biological imaging, improve CRISPR base editing, and function in vivo to allow AAV gene therapy applications. DIMMERs should thus allow routine research and biotechnology applications to operate with greater precision, control, and predictability.

3.3 Results

A minimal model shows that dosage compensation requires linear sensitivity to miRNA

To guide the design of DIMMERs, we first developed a minimal model of a miRNA-based IFFL circuit. This model makes several assumptions: (1) Primary miRNA (pri-miRNA) and target mRNA are each transcribed constitutively at a fixed ratio of rates, and in direct proportion to gene dosage. (2) There is a constant total rate, per gene copy, of RISC complex production, reflecting the combined process of pri-miRNA transcription, post-transcriptional processing, and binding to Argonaute proteins [12]. (3) RISC and its target mRNA bind reversibly to form a RISC-mRNA complex. Finally, (4) formation of this complex leads to degradation of the bound mRNA.

In certain regimes, this model exhibits dosage-invariant expression profiles, in which target protein expression levels increase linearly and then asymptotically approaches a dosage independent limiting expression level (**Figure 3.1D**). Several parameters modulate the limiting expression level while preserving asymptotic dosage compensation. These include the binding and unbinding rates of mRNA to miRNA, and the catalytic rate of mRNA degradation (**Figure 3.8A and B**). Accessing the dosage compensation regime requires that mRNA levels be linearly sensitive to RISC concentration (**Supplementary Modeling Text, Figure 3.8C**). Below, a more detailed model in which the total amounts of free and bound mRNA were also considered, revealed that miRNA-dependent catalytic degradation rates must exceed a minimal value for the total mRNA to show dosage invariance (**Supplementary Modeling Text, Figure 3.8D**). Together, these results suggest conditions in which miRNA-based IFFL circuits could produce gene dosage invariant expression.

Multimerization of weak target sites provides dosage compensation

Based on these results, we designed an initial set of regulatory circuits. Briefly, a synthetic miRNA and a target mRNA were placed in opposite orientations relative to a central divergent promoter [7] (**Figure 3.1E**). In one orientation, a previously characterized synthetic miRNA (miR-L, based on a Renilla luciferase sequence [13]) was co-expressed with the fluorescent protein mRuby3, serving as a dosage indicator. The miRNA expression cassette included the miR-E backbone for pri-miRNA expression [13] and was incorporated within a synthetic intron [14]. In the opposite orientation, we inserted a constitutively expressed EGFP target gene, with a single fully complementary 21-nt miRNA target site in the 3'-UTR. This format allowed analysis of multiple miRNA and target site configurations, and independent modulation of miRNA and target gene expression levels. We also systematically analyzed a broad variety of other architectures (**Figure 3.9**).

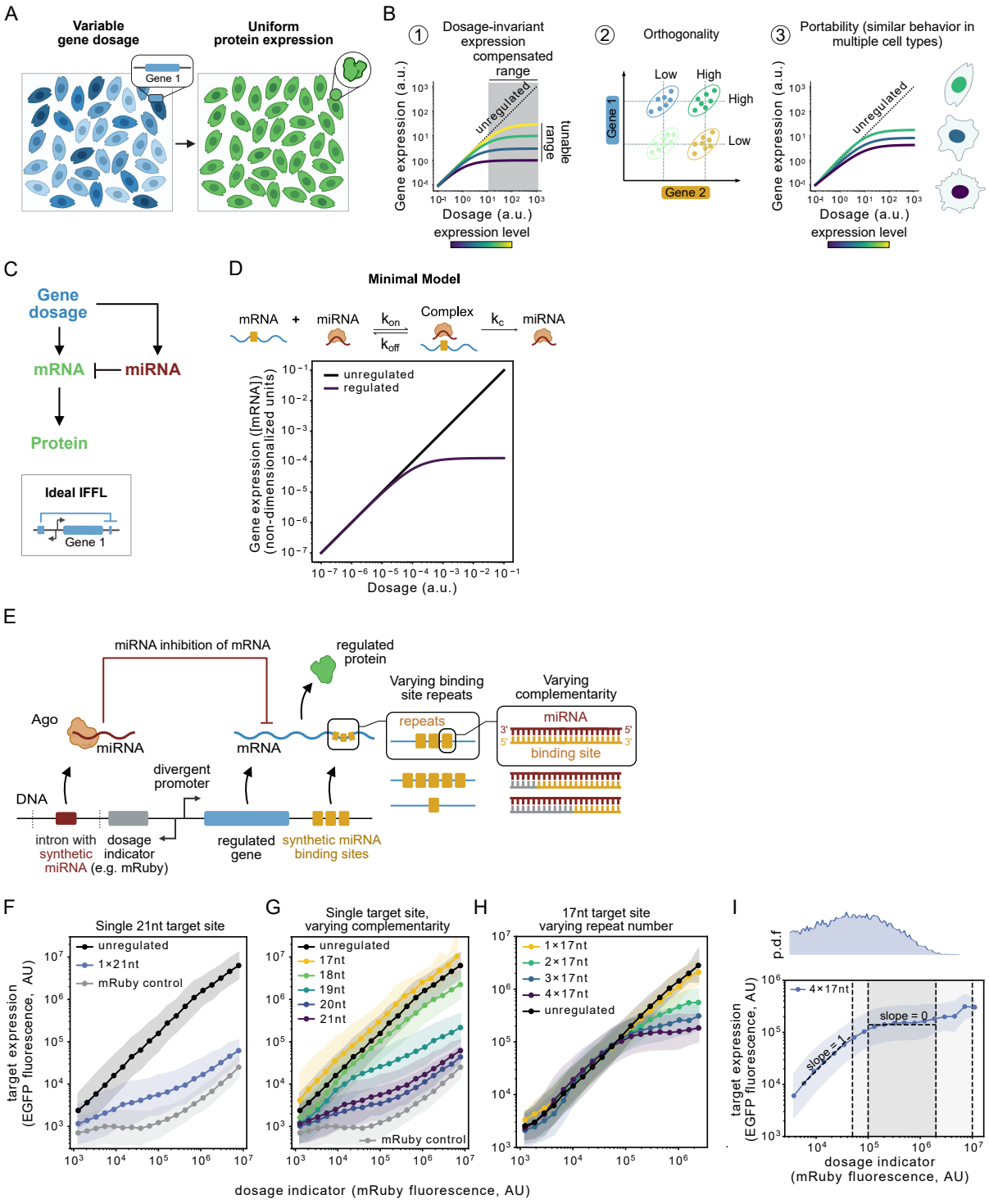

Figure 3.1: (Caption on next page)

Figure 3.1: **miRNA incoherent feedforward circuits enable dosage-invariant gene expression.** **(A)** An ideal gene expression system generates uniform protein expression (right panel, uniform green gradient) levels despite variable gene dosage delivered (left panel, variable blue gradient in the nucleus). **(B)** An ideal system would enable tunable (①), orthogonal (②) control of the target and operate in multiple cell contexts (③). (②) shows the protein expressions of two genes of interest regulated by four combinations (indicated by different colors) of two independently-controlled, tunable dosage compensators, with the ellipses indicating the majority of protein expression profile, and the dashed line indicating the centroids of expression. **(C)** The architecture of the incoherent feedforward loop (IFFL), top panel. The simplified miRNA-based IFFL circuit architecture, bottom panel. **(D)** The minimal model of miRNA inhibition. Upper panel, reactions in the miRNA-mediated regulation, see also **Supplementary Modeling Text**. Lower panel, the minimal model shows that miRNA-mediated regulation enables dosage compensation compared to the unregulated control. **(E)** The circuit configuration of the miRNA-mediated IFFL. Various target-miRNA complementarity and target site numbers were experimentally explored to implement dosage compensation. See also **Figure 3.9**. **(F)** A single, fully complementary miR-L site (1×21nt) shows strong repression compared to the unregulated but does not achieve dosage compensation. mRuby control shows the bleed-through signal from mRuby3 in the EGFP channel. Cells are gated and binned by mRuby3 intensities. Each dot corresponds to the geometric mean fluorescence intensity of mRuby3 bin breaks and median fluorescence intensity of EGFP in the bin. Shaded regions represent geometric standard deviation. **(G)** Reducing the complementarity (the pairing length starting from the seed region) of the single-site target weakens the repression but does not achieve dosage compensation. **(H)** Multimerizing the 17nt site restores the repression and achieves dosage compensation at four repeats. **(I)** A zoom-in of the 4 17nt in (H). p.d.f, probability distribution function, which measured the distribution of mRuby3 fluorescence intensity. The gray and light gray rectangles indicate the dosage range where the fluorescence intensity of EGFP does not change or change by 4-fold, respectively. The dashed lines with slope=0 or slope=1 indicate the linear dosage dependence at low dosage or dosage independence at high dosage, respectively.

To quantitatively measure the behavior of the circuit, we transfected U2OS cells with each circuit construct, and analyzed expression by flow cytometry 48h later. We then plotted target EGFP expression versus gene dosage, indicated by mRuby3 fluorescence (**Figure 3.1F**). Compared to an unregulated control with no miRNA target site, the circuit strongly reduced target EGFP expression by 1–2 orders of magnitude. However, it failed to achieve dosage compensation (**Figure 3.1F**). We also analyzed similar circuits in which the complementary region was systematically reduced in single nucleotide increments from 21nt to 17nt. Constructs with 19 or 20nt sites retained repression but failed to produce dosage compensation, while shorter sequences lost repression altogether (**Figure 3.1G**). Thus, single target sites with varying levels of complementarity did not provide dosage compensation.

Native miRNAs are known to use much shorter complementary regions, including central mismatches (bulges) [12], and multivalent interactions mediated by TNRC6 scaffold proteins [11,

15]. Therefore, we next considered designs with reduced complementarity and multimerized target binding sites. Tandem repeats of two to four copies of the 17nt target site, which was inactive in a single copy, progressively increased regulation (**Figure 3.1H**). More importantly, they successfully reduced dosage sensitivity, particularly at higher expression levels (**Figure 3.1H,I**). Thus, 4 tandem 17nt sites yielded only a 4-fold change in expression over a 200-fold range of dosages (**Figure 3.1I**). We also tested tandem repeats of two to four copies of the 18nt and 19nt target binding sites. Interestingly, the 18nt target site showed dosage compensation behavior at 3 copies, while the 19nt target site exhibited dosage compensation with only 2 copies (**Figure 3.10A, B**). "Bulged" target sites that provided no regulation individually nevertheless exhibited strong regulation and limited dosage compensation when multimerized (**Figure 3.10C, D**). These results indicate that multiple tandem copies of individually weak target sites can achieve dosage compensation over broad dosage regimes, at varying setpoints.

TNRC6 and Ago2 play key roles in regulation of multimerized 17nt targets

Why do multimerized weak sites produce better dosage compensation than stronger individual target sites? We reasoned that the TNRC6 scaffold protein could facilitate inhibition of multimerized weak target sites. To test whether regulation of multimeric weak sites requires TNRC6, we expressed T6B, a previously identified TNRC6B protein fragment that competitively inhibits TNRC6 activity (**Figure 3.2A**) [16]. T6B abolished regulation by a co-transfected 4×17nt DIMMER (**Figure 3.2B**), but had little effect on the single fully complementary 21nt construct (**Figure 3.2C**). As expected, negative controls using a T6B variant lacking the Ago2-binding domain failed to abolish regulation (**Figure 3.10E**). Interestingly, overexpression of wild-type Ago2 exhibited similar effects as ectopic T6B expression, nearly eliminating regulation in the 4×17nt case (**Figure 3.2D**), without affecting the 1×21nt configuration (**Figure 3.2E**), possibly due to sequestration of TNRC6. Thus, TNRC6 activity is required for regulation of the multimeric 4×17nt target, but not the single fully complementary site.

Although Ago2 predominantly directs repression through a slicing-independent mechanism, slicing occurs in at least a dozen complementary targets [17], and is required for maturation of two erythroid miRNAs that undergo non-canonical biogenesis [18]. To understand whether slicing activity is required for dosage compensation, we ectopically expressed a dominant negative Ago2 mutant (D669A) lacking slicing activity. This perturbation reduced regulation of both fully complementary and multimerized partially complementary targets (**Figure 3.2F,G**), suggesting that Ago2 slicer activity is required for repression of both targets. Further, to test for a role in cleavage of passenger strands, we incorporated a mismatch at 10-11nt in the original miRNA design (**Figure 3.2H**) in order to eliminate the requirement for Ago2 cleavage in miRNA maturation [18]. Inclusion of the mismatch slightly enhanced regulation of both the 4×17nt and 1×21nt (**Figure 3.10K, L**),

possibly by elevating the efficiency of miRNA biogenesis. Although passenger strand removal is not required in this case, dominant negative Ago2 expression nevertheless largely eliminated regulation of both targets, consistent with a role for Ago2-dependent slicing in target repression (**Figure 3.2I-J**). By contrast, wild-type Ago2 overexpression had a partial impact on the 4×17nt target but no impact on the 1×21nt target (**Figure 3.2I-J**). We also analyzed other target sequences. The 4×18nt and 4×19nt targets responded similarly to T6B and Ago2^{DN}, with Ago2^{DN} showing a slightly stronger de-repression effect. By contrast, wild-type Ago2 had weaker or no suppressive effect on these stronger target sites (**Figure 3.2K-L**), possibly because their longer complementary regions were less sensitive to indirect sequestration of TNRC6. Taken together, these results suggest that TNRC6 is required for repression of multimerized partially complementary targets but not fully complementary targets, and that Ago2-dependent slicing is required for repression of partially and fully complementary targets.

Taken together, these results suggest a potential explanation for why dosage compensation requires multimerized weak binding sites and TNRC6 (**Figure 3.2M**). Briefly, fully complementary binding sites are known to produce much more efficient slicing than partially complementary sites [19]. The fully complementary 21nt target could therefore produce dosage compensation at dosages and setpoints comparable to or lower than detection limits in the flow cytometry experiments. At the same time, the higher affinity of fully complementary binding sites would shift the "tail" of elevated expression due to bound-but-not-yet-degraded complexes to lower dosages. These combined effects could together explain the shape of the single 21nt fully complementary construct (**Figure 3.1F**). By contrast, constructs with multimerized weak binding sites (e.g., 4×17nt) would engage at higher gene dosages due to the lower intrinsic RNA affinity. This would lead to elevated setpoints and shift the "tail" of non-compensated expression to correspondingly higher dosages (**Figure 3.1I, Figure S1D**). Efficient mRNA degradation could be achieved by enhancing the intrinsically weaker Ago2-dependent slicing activity through stabilization of Ago2-mRNA complexes by TNRC6 and through TNRC6-dependent deadenylation, as described previously.

Dosage-invariant expression levels can be tuned

Having demonstrated dosage invariant circuit designs and gained insight into their design from the model, we next sought to identify experimentally tunable parameters that modulate expression level setpoints (**Figure 3.1E**). We first examined varying the length of target complementarity from 8 to 21nt, while maintaining 4 tandem repeats (**Figure 3.3A, Figure 3.11**). Repression was modest at 8nt, diminished with increased complementarity in the central region, and then strengthened again as more complementarity was added after the central region (**Figure 3.3A**, left panel). These results are consistent with previous observations that miRNA inhibition does not increase monotonically with complementarity [20, 21, 22]. For the miR-L target site, repression was most sensitive at

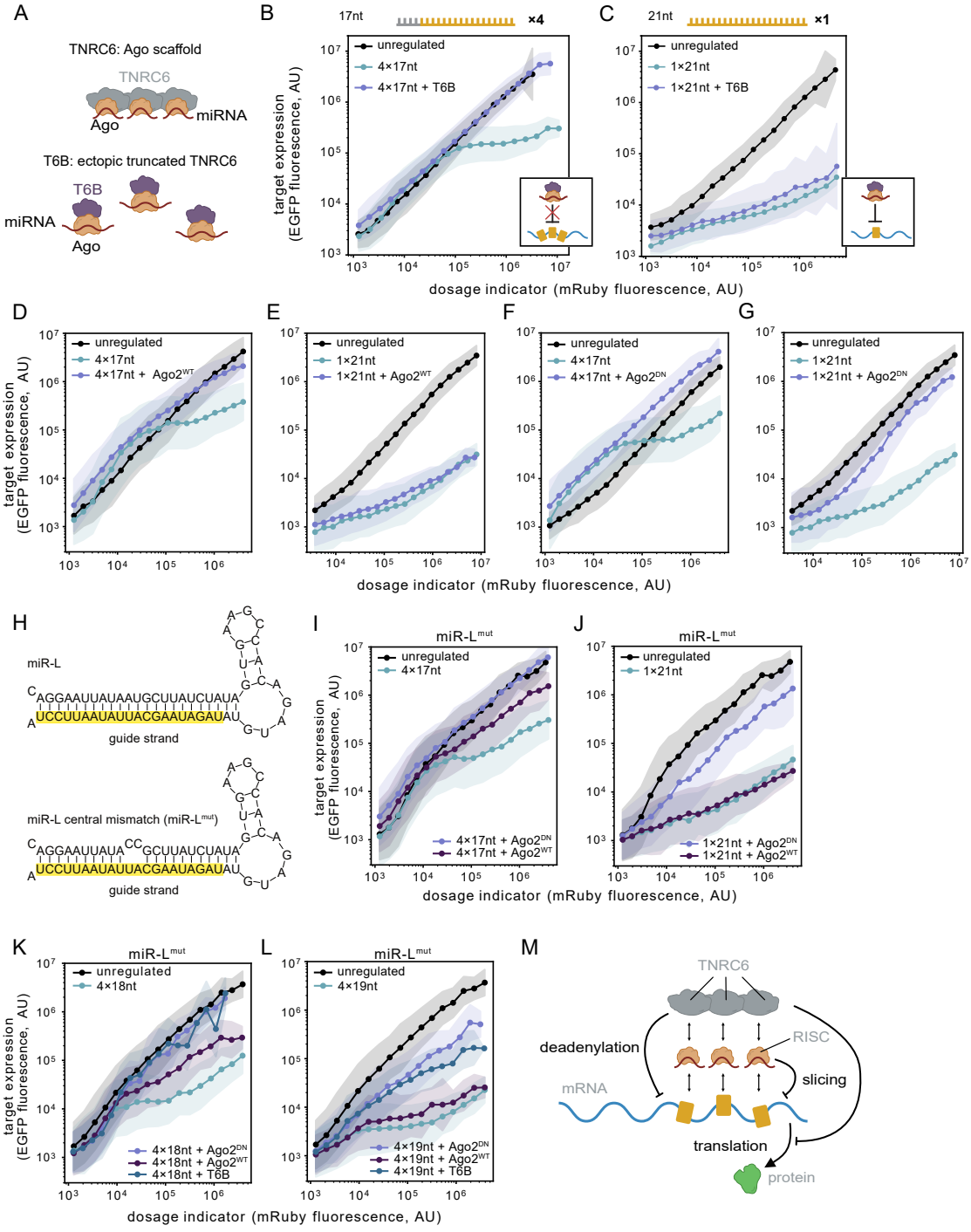

Figure 3.2: (Caption on next page)

Figure 3.2: **TNRC6 and Ago2 play key roles in regulation of multimerized weak targets. (A)** T6B peptide competitively inhibits TNRC6-dependent regulation (schematic). **(B, C)** T6B expression suppresses miR-L-mediated regulation of the 4×17nt construct (B) but not the 1×21nt (C) construct. Cells were co-transfected with the miR-L-mediated IFFL construct, either with the T6B-peptide or the iRFP as a control. **(D, E)** Ectopic wild type Ago2 ($Ago2^{WT}$) suppresses inhibition of 4×17nt (D) but not 1×21nt (E) constructs. **(F, G)** Dominant negative Ago2 ($Ago2^{DN}$) expression suppresses regulation of both 4×17nt (F) and 1×21nt (G) constructs. In D-G, cells were co-transfected with indicated miR-L-mediated IFFL constructs, with either $Ago2^{WT}$ (D,E), $Ago2^{DN}$ (F,G), or a filler plasmid. **(H)** Design of synthetic miR-L cassettes without (top) or with (bottom) central mismatches in the passenger strand. 9 **(I, J)** Ectopic $Ago2^{WT}$ or $Ago2^{DN}$ modulate miR-L^{mut}-mediated regulation of the 4×17nt (I) and 1×21nt (J) constructs. **(K, L)** The effect of the T6B/$Ago2^{WT}$/$Ago2^{DN}$ overexpression in the miR-L^{mut}-mediated regulation for the 4×18nt (K) and 4×19nt (L) constructs. **(M)** TNRC6 and Ago2 play important roles in dosage compensation (schematic).
See also **Figure 3.10**.

16-20 nucleotides of complementarity (**Figure 3.3A**). Three designs—4×17, 4×18, and 4×19—achieved dosage invariant expression, but did so at distinct setpoints spanning more than an order of magnitude in saturating expression level (**Figure 3.3B**). These results indicate that site length can tune setpoint.

We next varied the number of copies of the miRNA expression cassettes in the synthetic intron, effectively modulating the stoichiometric ratio of miRNA to mRNA (**Figure 3.3C**). Compared to a single copy, two or three copies of the miRNA reduced expression by 2-fold and 3-fold, respectively, while preserving dosage compensation, providing a means of fine-tuning expression control.

We also compared different promoters to vary transcription of the miRNA and target cassettes. For the miRNA the weaker PGK promoter allowed ~3.5-fold more target gene expression at a given dosage level compared to the stronger EF1α, but nevertheless preserved dosage compensation (**Figure 3.3D**). Finally, we incorporated a tetracycline-responsive system (**Figure 3.3E**). This allowed setpoint tuning over ~2 orders of magnitude while preserving dosage compensation (**Figure 3.3E, Figure 3.12**). Taken together, these results demonstrate that dosage invariance can be preserved while allowing multiple mechanisms of coarse (site length, target transcription rate) and fine (miRNA promoter and copy number) setpoint tuning.

Orthogonal dosage compensation circuits allow independent control of target genes

Engineered genetic systems increasingly require multiple genes and transcripts, provoking a need for multiple independent DIMMER systems based on orthogonal synthetic miRNA-target site pairs, which we term *synmiRs*. To design synmiRs, we started by 11 generating five random miRNA

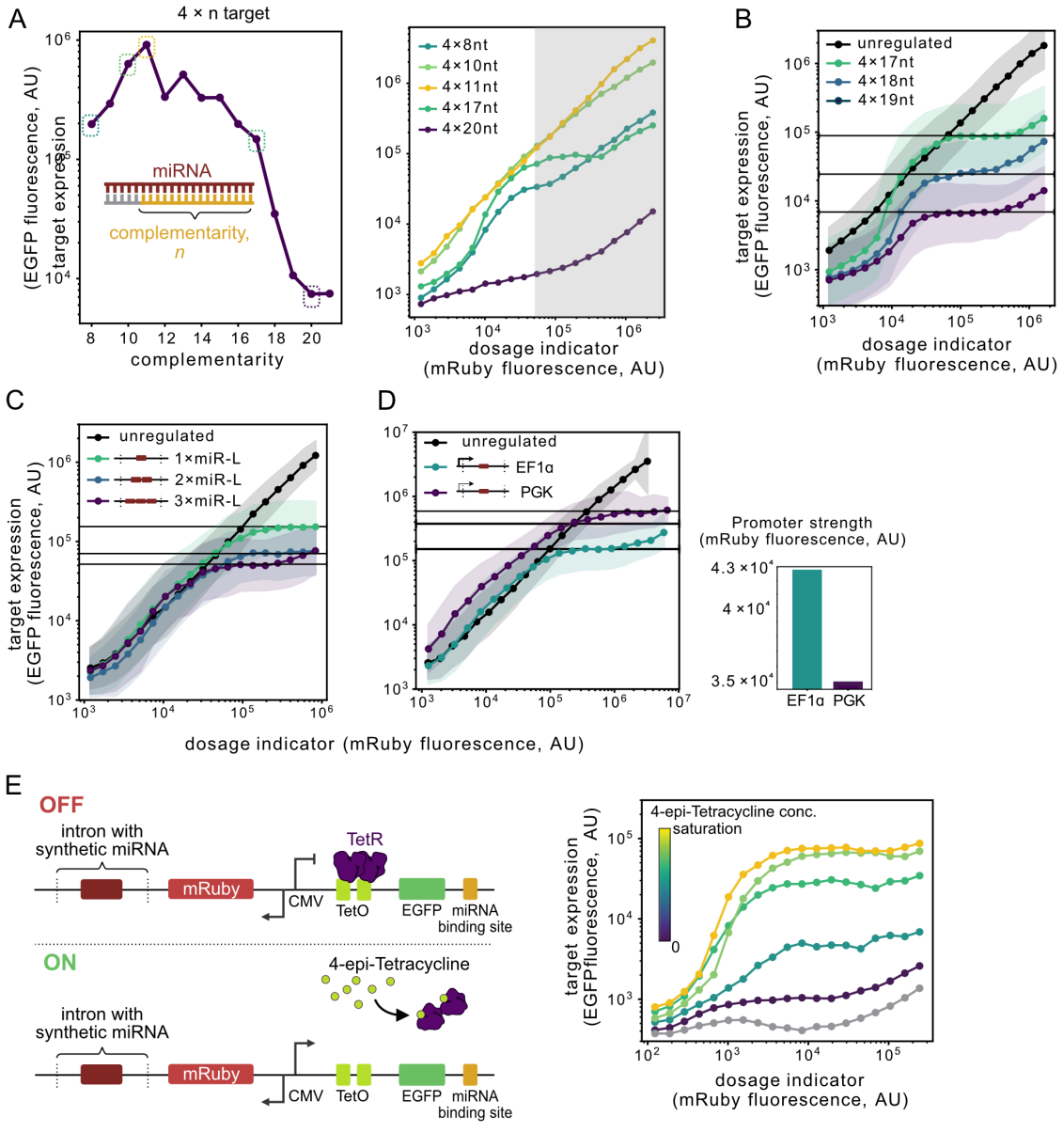

Figure 3.3: (Caption on next page)

Figure 3.3: **Dosage-invariant expression levels can be tuned.** **(A)** Expression levels exhibit biphasic dependence on complementarity. All circuits have 4 miR-L target sites. Expression (y-axis) represents median EGFP fluorescence intensity in cells with mRuby3 fluorescence intensity > 5×10^4, see also the shaded region in the right panel). Right panel, the dosage response curves of the five selected constructs (dashed boxes) in the left panel. **(B)** Complementarity of 17-19nt monotonically regulates setpoint (black horizontal line). See also **Figure 3.11. (C)** Number of miR cassettes fine-tunes setpoint (black horizontal line). **(D)** miR-L promoter strength regulates setpoint (black horizontal line). Bars indicate mRuby3 median fluorescence intensities of transfected cells. In C and D, the target is 4×17nt. **(E)** Tet system allows tunable dosage compensation. Left panel, design of the inducible DIMMER. A CMV promoter harboring two downstream TetO sites was utilized to drive 4-epi-Tc inducible expression in the TRex cell line. Right panel, the performance of the inducible 4×19nt DIMMER construct. The concentrations of the drug, from purple to yellow, were 0, 10, 33.3, 100, 333.3 ng/mL. The gray curve denotes the mRuby-only transfection control. See also **Figure 3.12**.

---

sequences, labeled synmiR1-5. Each sequence contained A at position 1 in the miRNA, a cognate U at the 3' end of a single target site and 25% GC content, similar to the structure of miR-L. We used an "open loop" system to analyze their behavior, allowing independent control of miRNA expression and measurement of its effect on a target miRNA reporter gene (**Figure 3.4A**). For a single fully complementary miR-L site, inhibition increased with miRNA dosage (**Figure 3.4B**), consistent with the earlier closed loop results (**Figure 3.1F**). synmiRs 1,4, and 5 repressed by at least an order of magnitude relative to a control lacking the miRNA (**Figure 3.4D, Figure 3.14A**). By contrast, synmiRs 2 and 3 achieved weaker repression, possibly due to subsequences containing two or more A/T pairs in the extensive region, which could destabilize the miRNA [23, 24] (**Figure 3.13A**). Consistent with this hypothesis, an A to G substitution at position 20 in synmiR-2 or at position 19 in synmiR-3 restored miRNA inhibition of target gene expression (**Figure 3.4D, Figure 3.14**).

Based on these results, we formulated an empirical algorithm for synmiR design. Briefly, we generate random 21nt candidate miRNA sequences in which the mature miRNA (1) includes a 5'-U, based on known requirements for miRNA loading [12, 25], and, (2) includes 5–8 G or C nucleotides, with 1–4 of them in the seed region, and 1–2 in the extensive region (**Figure 3.4C**). We synthesized five candidate sequences (synmiRs 6-10) based on this simple algorithm and analyzed their open loop behavior (**Figure 3.4D, Figure 3.13B**). All five sequences generated strong repression as single sites, comparable to that of miR-L. (One sequence, synmiR-6, was discarded due to its similarity to the human endogenous miRNA hsa-mir-5697 (**Methods**)). Altogether, these results produced ten miRNA sequences capable of strong repression in their fully complementary form (**Figure 3.4D, Figure 3.14A**). . Pairing each of these ten miRNAs with all ten of the target

sequences in the open loop system revealed strong orthogonality in regulation, as desired (**Figure 3.4E, Figure 3.13C**).

Using these synmiR sequences, we developed a set of ten orthogonal dosage compensation circuits, using the framework in **Figure 3.1E**. Each of these sequences could produce dosage compensation in some configurations and dosage regimes. For example, with synmiR-4 and synmiR-5, the 4×17nt configuration produced a strong inhibition profile more similar to the fully complementary target (**Figure 3.13D**). We reasoned that this could reflect higher GC content in the seed and supplementary regions of these two miRNAs compared to miR-L, allowing shorter 8 or 9 nt target sites, present in 4-8 repeats, to produce dosage compensating designs (**Figure 3.4F**). In a similar way, we identified dosage compensation regimes for the other sequences (**Figure 3.4F**). Eventually, we identified nine DIMMER circuits exhibiting different levels of dosage compensation, of which five showed at least a 30-fold dosage-invariant range (**Figure 3.4F, Figure 3.14B**). Further, expression levels in these circuits were sensitive to the number and complementarity of target sites, allowing setpoint tuning (**Figure 3.15A**). With one exception, like the circuits above, these designs relied TNRC6 for dosage compensation (**Figure 3.15B**). The exception was synmiR-2 8×8nt, which maintained strong repression even under the T6B perturbation (**Figure 3.15B**). Taken together, these results provide a toolkit of dosage compensating systems and, more generally, suggest that it should be relatively feasible to engineer many additional dosage compensating systems with varying expression setpoints.

With multiple dosage compensation systems, it should be possible to independently specify the expression of multiple target genes in the same cells (**Figure 3.1B, panel 2**). To test this 12 possibility, we constructed a second set of dosage compensation expression systems using distinct fluorescent reporters (**Figure 3.4G**). We transfected cells with pairs of systems that had different regulatory setpoints, and analyzed the resulting expression profiles of the two regulated target genes (**Figure 3.4H, Figure 3.13E**). Altogether, we analyzed four pairs of systems. Each produced a distinct two-dimensional expression distribution based on the setpoints for the two reporters (**Figure 3.4H**). By contrast, the unregulated group showed higher setpoints, and broader distributions of both reporters. The dosage indicators also exhibited the same distributions among all groups. Interestingly, the circuits allowed precise control of the stoichiometry of the regulated proteins (**Figure 3.13E**). The engineered dosage compensation systems thus make it possible to specify two-dimensional expression distributions, and suggest that control of higher dimensional distributions of more genes should also be accessible.

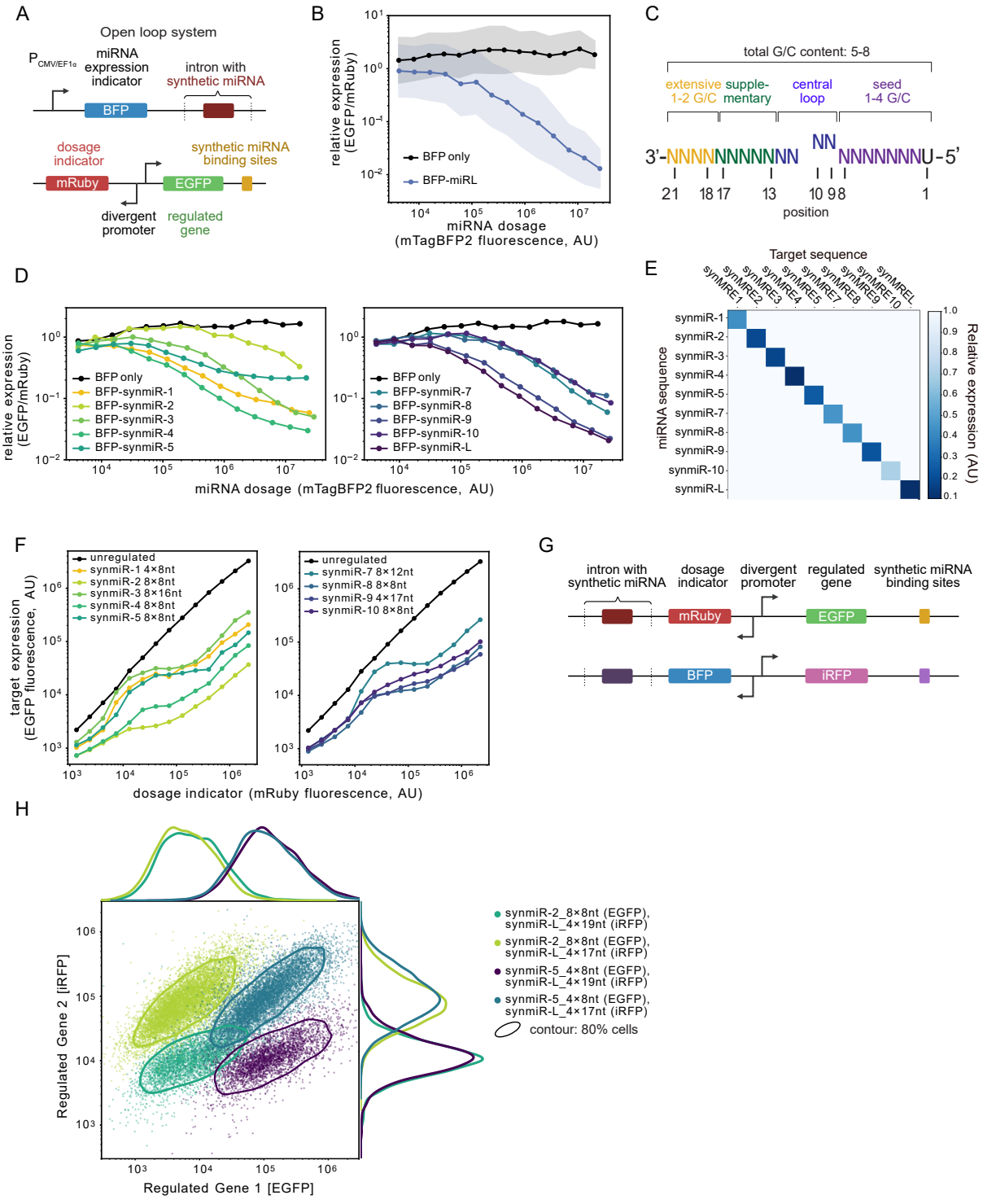

Figure 3.4: (Caption on next page)

Figure 3.4: **Orthogonal dosage compensation circuits allow independent control of target genes. (A)** Open loop system allows simultaneous readout of miRNA expression (BFP) and its effect on target gene expression (EGFP), at different target gene dosages (mRuby3). **(B)** Open loop circuits reveal quantitative miRNA input-output relationship. Cells were co-transfected with the 1×21nt target, either with the miR-L construct shown in (A), or mTagBFP2 as a control. Relative expression was defined as the median of the ratio of EGFP fluorescence intensity to mRuby3 fluorescence intensity in each mTagBFP2 bin. Dots and shaded regions are as described in **Figure 3.1F**. **(C)** synmiRs use a simple empirical design algorithm. **(D)** All synmiRs are capable of regulating cognate target genes. Open loop response curves are divided into plots for clarity. Each plot shows regulation of a target incorporating a single 21nt target site. **(E)** The 10 synmiRs function orthogonally. Cells were co-transfected with each combination of target and miRNA for open loop analysis. Relative expression levels were quantified as described in (B), and normalized by maximal expression of each construct across synmiRs. Biological replicates shown in **Figure 3.13C**. **(F)** synmiRs can generate dosage compensated regulation. Each construct has the indicated target repeat number and complementarity. **(G)** Two DIMMER constructs can be analyzed in the same cells using distinct fluorescent proteins. Diagrams depict double DIMMER reporter systems. **(H)** The double DIMMER reporter system allows independent regulation of two genes in four distinct expression configurations. Cells were poly-transfected with the double DIMMER reporter system for DIMMER combinations shown in (G). Data were gated by mRuby3 and mTagBFP2 intensities. Specific DIMMER circuits are indicated in legend (right). Solid contour lines indicate regions containing 80% of cells. Profile plots show the distributions of the corresponding fluorescent proteins. See also **Figure 3.13-3.15**.

Dosage comensation systems are portable and minimally perturbative

An ideal dosage compensation system would be portable, able to operate similarly across different cell types, function in both transient transfection and genomic integration, and minimally perturb the host cell. To examine these features, we transiently transfected several circuit variants, including the 4×17nt miR-L system (**Figure 3.1H**), in four mammalian cell lines: U2OS [26], CHO-K1 [27], HEK293 [28], and N2A [29]. In each cell line, we observed strong and qualitatively similar dosage compensation (**Figure 3.5A**). Cell lines varied in the threshold dosage at which expression saturated (**Figure 3.5A**, gray vertical line), and in the saturating expression level (**Figure 3.5A**, gray horizontal line), as measured in arbitrary fluorescence units. However, the ratio of these values was conserved (**Figure 3.5B**). We obtained similar results for other circuits as well, including synmiR-4, with 8 repeats of a 9nt target site, as well as both synmiR-L and synmiR-5, each with 8 repeats of an 8nt target site (**Figure 3.16A**). Again, the ratio of the saturating expression level to the threshold dosage was similar, for each construct, across cell lines (Figure 3.16B). This suggests a model in which the miRNA circuit functions equivalently in different cell types, but protein expression strengths vary, possibly due to differences in translational capacity or basal protein degradation rates [30, 31, 32, 33]. Together, these results indicate that the dosage compensation circuits can

function across different cell contexts.

Stable cell lines are important in research as well as applications like cell therapy. To find out whether dosage compensation circuits could also function in a stable integration context, we used PiggyBac transposition together with the iON system that allows expression only from constructs that have successfully integrated in the genome and undergone site-specific recombination [34]. We then selected mono-clones, and analyzed reporter expression by flow cytometry (**Figure 3.5C**). Integration copy numbers varied among clones by over two orders of magnitude, as indicated by mRuby3 fluorescence intensity (**Figure 3.5D left panel**, x-axis). Nevertheless, the cargo EGFP expression remained nearly constant (**Figure 3.5D left panel**, y-axis). By contrast, the unregulated mono-clones exhibited an apparent correlation between 15 the integrated copy numbers and the EGFP expression (**Figure 3.5D right panel**). Thus, dosage compensation circuits function in stable integration settings.

The expression of synthetic miRNAs could in principle perturb endogenous gene expression. To identify such effects, we performed bulk RNA sequencing on cells transfected with miR-L and each of the 9 orthogonal synmiRs, and compared them to a negative control transfection of a BFP expression vector. Only a few genes were significantly up- or down-regulated by the miRNA (**Figure 3.5E**). These were enriched for heat shock proteins such as HSPA6. Critically, the gene sets up-regulated by different miRNAs exhibited strong overlap (**Figure 3.5F, Figure 3.17**). Thus, for the synmiRs described here, off-target regulation appears to only reflect non-specific effects of miRNA expression, rather than sequence-specific perturbations.

Dosage compensation enhances biological imaging

Single-molecule imaging approaches such as DNA-PAINT (Point Accumulation for Imaging in Nanoscale Topography) often rely on ectopic expression of tagged proteins [35], which can dramatically exceed endogenous expression, distorting subcellular localization patterns. DIMMER circuits could potentially address this issue by limiting ectopic protein expression.

To test this, we transfected a DIMMER-regulated and unregulated EGFR-mEGFP membrane marker fusion protein expression-constructs into CHO-K1 cells, which lack endogenous EGFR expression [36] (**Figure 3.6A**). The DIMMER circuits successfully reduced expression, as measured by receptor density relative to an unregulated control, by at least 10-fold, as estimated from flow cytometry and confocal imaging (**Figure 3.18A-C**). To quantify this reduction more directly, we imaged transfected cells 48h post-transfection using the DNA-PAINT method, based on DNA-conjugated anti-GFP nanobodies targeting the intracellular mEGFP tag of the EGFR receptor (**Figure 3.6B-D**). Circuit-regulated receptor mean densities were ~15-47 times lower than unregulated receptors (**Figure 3.6E, Figure 3.18D**). Further, expression distributions were narrower for the regulated plas-

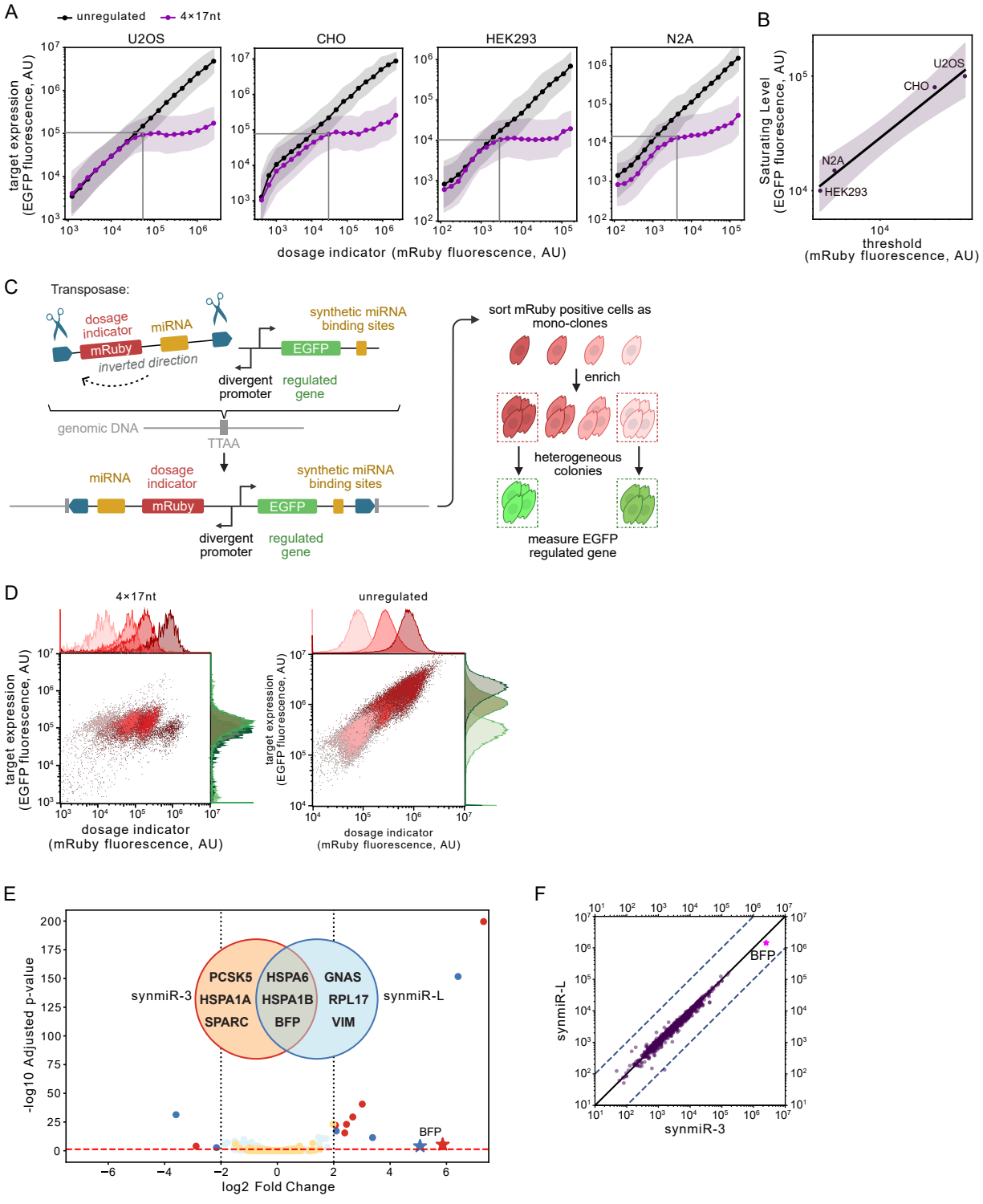

Figure 3.5: (Caption on next page)

Figure 3.5: **Dosage compensation systems are portable and minimally perturbative. (A)** DIMMERs function in multiple cell lines. 4×17nt DIMMER circuits were analyzed in indicated cell lines. Minimal dosage for the compensation regime (threshold level) and setpoint expression were indicated by gray vertical lines and horizontal lines, respectively. **(B)** The threshold and the saturating level co-vary across cell lines. The black solid line indicates a linear fit in the log-log space (**Methods**). **(C)** iON transposition was used to genomically integrate DIMMER circuits to prepare clonal stable lines (right). **(D)** DIMMER circuits enable uniform protein expression (y-axis) across a range of stable integrations and expression levels (x-axis). Each color shows the 4×17nt DIMMER (left panel) compared with the unregulated construct (right panel) in PiggyBac-integrated monoclonal U2OS cells. Profile histograms show the mRuby3 and the EGFP distributions of each mono-clone. **(E)** DIMMER circuits show limited perturbation to the endogenous transcriptome. Bulk RNAseq volcano plots of the cells transfected with either the synmiR-3 (orange and red dots) or synmiR-L (light blue and dark blue dots), compared to mTagBFP2-only transfected cells. Dashed lines indicate $|\log_2(\text{fold change})| = 2$. Red dashed line indicates adjusted p-value = 0.05. The Venn diagram shows differentially expressed genes between synmiR-3 and synmiR-L. **(F)** Global mRNA expression levels, measured in normalized transcripts per million (TPM), of the miR-L expressing cells plotted against those of the synmiR-3 expressing cells. Solid line indicates equal expression in both samples. Dashed lines indicate 10 fold expression differences. See also **Figures 3.16-3.17**.

---

mids compared to the unregulated ones (**Figure 3.6E**). Critically, DIMMER reduced expression more homogeneously than possible by simply reducing the concentration of unregulated plasmids in transfections (**Figure 3.6F, Figure 3.18D**). DIMMER-regulated expression with the 4×17nt circuit achieved EGFR expression levels comparable to but lower than those of wild-type U2OS cells. Thus, DIMMER circuits allow homogeneous reduction in ectopic expression to endogenous levels.

CRISPR imaging methods enable analysis of specific genomic loci in cell nuclei, but can be limited by high background fluorescence due to basal expression of dCas9 fluorescent protein fusions [37, 38]. DIMMER circuits also improved dCas9-EGFP imaging of telomeres. We transfected dCas9-EGFP with or without DIMMER circuits, along with a gRNA targeting repetitive telomeric sequences [38, 39, 40] (**Figure 3.6G**). The circuits reduced dCas9-EGFP expression, and its dosage sensitivity (**Figure 3.19A**). Critically, in the unregulated system, dCas9-EGFP formed bright aggregations in the nucleolus, but seldom labeled the telomeres, consistent with previous observations [41] (**Figure 3.6H left panel**). By contrast, the 4×17nt circuit restricted most fluorescence to puncta (dots), consistent with telomeric labeling38–40, and reduced labeling of the nucleolus (**Figure 3.6H middle panel**). Further, the stronger 18 4×19nt circuit removed nearly all labeling in the nucleolus, while maintaining apparent telomere labeling (**Figure 3.6H right panel**). DIMMER circuits also substantially improved signal-to-background ratio as well as contrast for individual dots (**Figure 3.19B**). Taken together, these results demonstrate that dosage compensation circuits can improve imaging of proteins and subcellular structures.

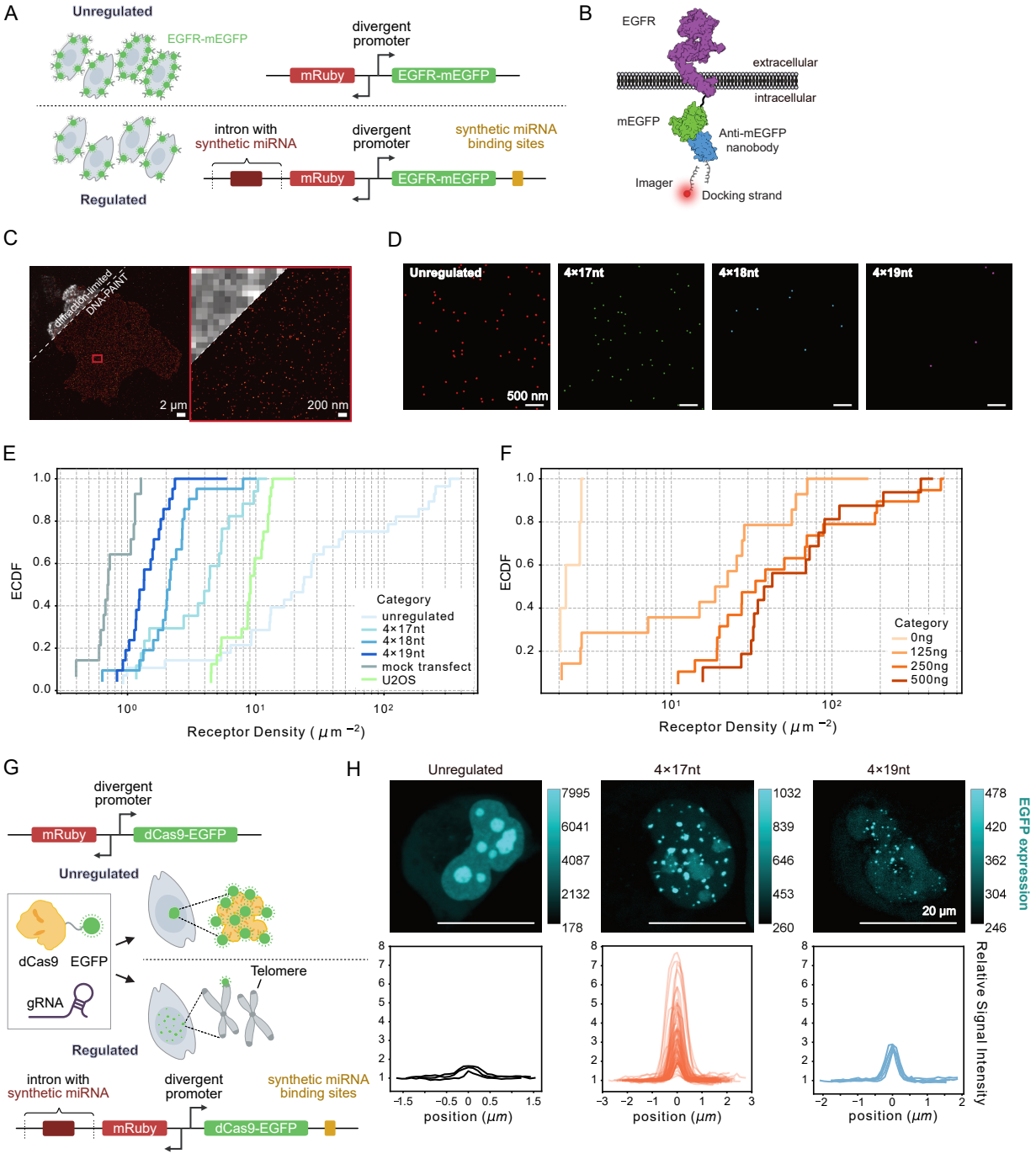

Figure 3.6: (Caption on next page)

Figure 3.6: **Dosage compensation circuits enhance biological imaging.** **(A)** EGFR receptor fusions to mEGFP were expressed with (lower panel) or without (upper) regulation. **(B)** Illustration of EGFR in the cell membrane (adapted from PDB 7SYE). EGFR is linked to an mEGFP labeled with a DNA-conjugated anti-GFP nanobody (PDB 6LR7). **(C)** Diffraction-limited and DNA-PAINT imaging of CHO-K1 cells expressing EGFR-mEGFP. The right panel shows the zoom-in of the boxed area. **(D)** Representative single-molecules images of EGFR-mEGFP using DNA-PAINT. Images show (L-R) unregulated or regulated with a series of DIMMER variants. **(E)** DIMMER circuits homogeneously reduce expression levels to near-background levels. The empirical cumulative distribution function (ECDF) of the receptor density for the CHO-K1 cells transfected with EGFR-mEGFP with or without the DIMMER module (blue curves), and the endogenous EGFR density in the U2OS cells (green curve), along with the mock-transfected CHO-K1 cells, measured by DNA-PAINT. Receptor densities were: unregulated: 79.0 ± 42.4 μm^{-2}, 4×17nt: 5.1 ± 1.7 μm^{-2}, 4×18nt: 2.8 ± 1.0 μm^{-2}, 4×19nt: 1.7 ± 0.5 μm^{-2} , mock transfection: 0.9 ± 0.2 μm^{-2} , U2OS: 9.8 ± 1.5 μm^{-2} (mean ± 95% confidence interval). **(F)** Reducing unregulated plasmid concentration does not provide a low and homogeneous receptor density, measured by DNA-PAINT. The density of the receptors are: 0 ng: 2.5 ± 0.4 μm^{-2}, 125 ng: 36.0 ± 25.0 μm^{-2}, 250 ng: 115.7 ± 74.4 μm^{-2}, 500 ng: 104.1 ± 64.5 μm^{-2}(mean ± 95% confidence interval). See also **Figure 3.18**. **(G)** dCas9 can be used to image telomeres (schematic). gRNA is driven by a U6 promoter, targeted to the telomeres, and expressed from the same vector as dCas9-EGFP. **(H)** DIMMER circuits enable telomere imaging. Upper panel: representative fluorescence images of dCas9-EGFP with or without the DIMMER module. Color bars indicate measured fluorescence intensities. Lower panel, quantitation of the relative signal intensity of the dots in the cells transfected with the dCas9-EGFP with or without the DIMMER module. Each line represents one dot. In the unregulated condition, high backgrounds made isolated dots rare. See also **Figure 3.19**.

DIMMER reduces off-target RNA base editing and transcriptome stress

The adenine base editor ABEMax is a powerful gene-editing tool. However, it can induce undesired transcriptome-wide off-target A-to-I editing in HEK293 cells [42] and cause detrimental transcriptional responses [43] raising concerns for gene therapy applications. Engineering of the TadA and/or TadA* domain has partly reduced these issues [44, 45]. However, the rate of off-target RNA editing rate increases with the deaminase level [46]. We therefore asked whether DIMMER regulation could further reduce off-target RNA editing while preserving high levels of on-target DNA editing without additional protein engineering.

We investigated the impact of ABEMax regulation on both on-target DNA, and off-target RNA editing. To evaluate the on-target editing rate, we designed a guide RNA targeting the widely used “site 3” in HEK293 [45, 47]. We then co-transfected HEK293 cells with the guide RNA construct along with the ABEMax base editor, either unregulated or regulated with the DIMMER circuit module (**Figure 3.7A**). We initially chose the 4×18nt miR-L design since it provides an intermediate expression level setpoint. We harvested the cells 72h post-transfection, and extracted

genomic DNA and RNA for further measurements.

The regulated ABEMax base editor performed on-target editing at a similar, if modestly decreased, rate compared to the unregulated control (**Figure 3.7B**). To measure the off-target RNA editing rate, we selected three transcripts (*CTNNB1, RSL1D1, IP90*) known to exhibit high off-target editing due to similarity with native TadA tRNA substrates [47, 44]. The DIMMER circuit successfully decreased the mean A-to-I editing rate by 4 to 8.4-fold across the three transcripts (**Figure 3.7C**). It also reduced the number of detectable A-to-I conversions within the transcript (using a cutoff of 0.1% based on the sequencing results) (**Figure 3.7D**). At specific adenosines that are highly edited by the unregulated ABEMax, the circuit exhibited dramatically lower (around 10-fold) editing rates (**Figure 3.7E, Figure 3.20A-B**). We did not detect off-target genomic DNA editing on the predicted potential endogenous off-target sites with or without DIMMER. These results show that DIMMER regulation can improve the ratio of on- to off-target edits.

The DIMMER circuit also reduced perturbations to the transcriptome. We performed bulk RNA sequencing on cells treated with ABEMax with or without DIMMER regulation. DIMMER reduced transcriptome changes compared with unregulated ABEMax expression (**Figure 3.7F, Figure 3.20C**). More specifically, unregulated ABEMax exhibited ~2-fold more differentially regulated genes (adjusted $p<0.05$) compared to the regulated construct. Gene ontology annotation showed that heat shock proteins, DNA damage, and repair-associated proteins were perturbed in both unregulated and regulated ABEMax groups. However, the unregulated ABEMax group produced 89 differentially expressed genes in the 'nucleoplasm and nucleus' category, suggesting a more pervasive impact (**Figure 3.7F, Figure 3.20C**). Together, these results indicate that DIMMER can improve the specificity of base editors by limiting their expression.

DIMMER circuits can operate in live mouse brains

Monogenic diseases of the central nervous system could potentially be cured by gene replacement therapies [48, 49] using recently developed adeno-associated virus (AAV) capsids that deliver genes across the blood-brain barrier [50, 51, 52]. However, many therapeutic genes are toxic or deleterious when overexpressed and expression must be maintained over long periods. These considerations underscore the need for controlled, durable cell type specific in vivo expression of transgenes.

To test whether DIMMER circuits could function in vivo in mouse brain, we designed a set of 6 AAV vectors that express H2B-EGFP with or without the broadly dosage compensating 4×19nt DIMMER circuit (**Figure 3.7G upper panel, Figure 3.21A**). For in vivo expression, we compared three promoter systems with distinct cell-type specificities: the CaMKIIα promoter (CaMKIIα hereafter), which drives expression in excitatory projection neurons, particularly in the cortex and

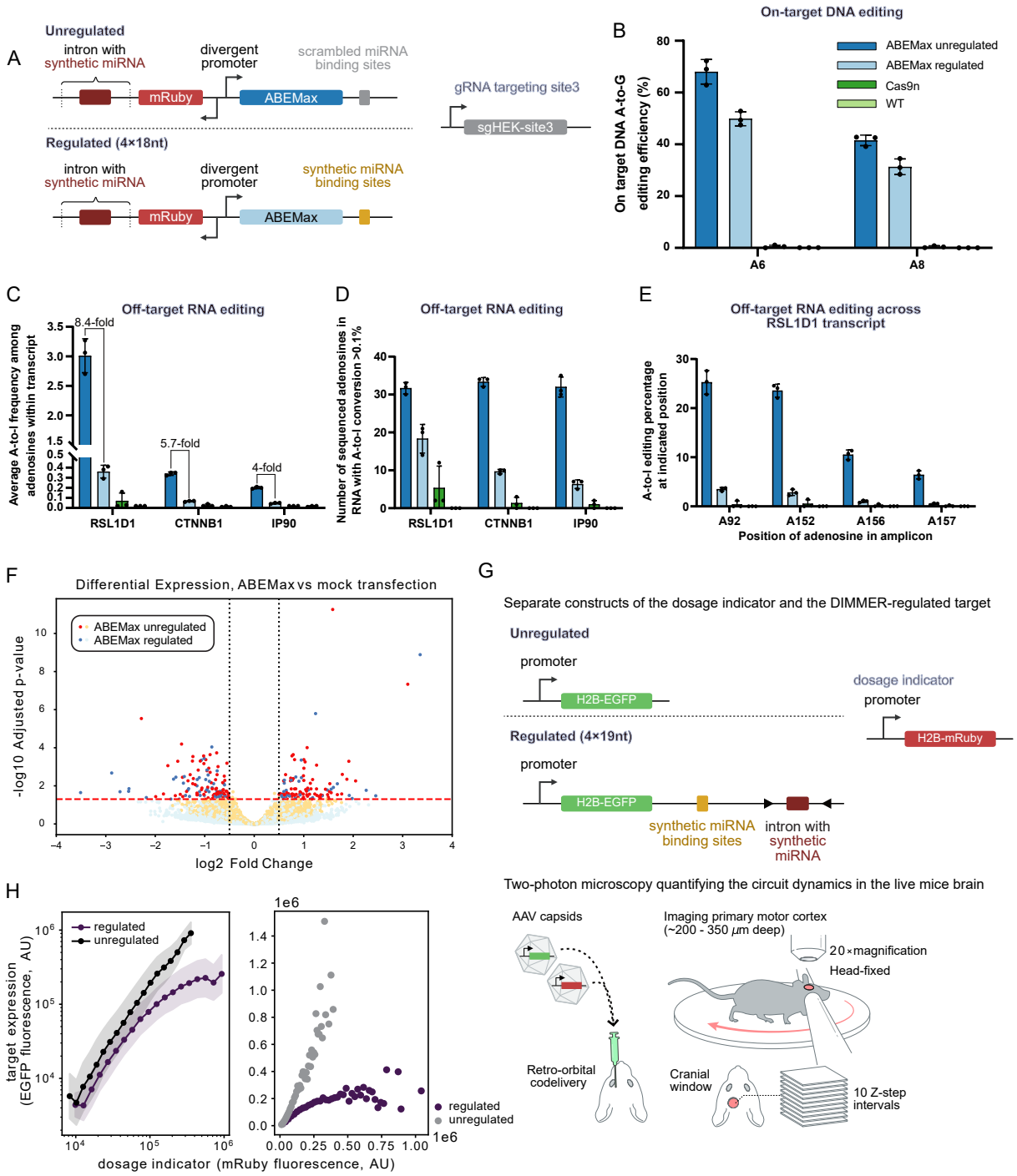

Figure 3.7: (Caption on next page)

Figure 3.7: **DIMMERs reduce ABEMax off-target RNA base editing and transcriptome stress, and operate in live mouse brains.** **(A)** The ABEMax base editor was expressed with or without the miR-L-based DIMMER. The gRNA targeting HEK293 site 3 is expressed on a separate plasmid driven by the U6 promoter. **(B)** DIMMER modestly reduces on-target A-to-G DNA editing efficiency, measured by the percentage of the A6 and A8 edited at the HEK293 site 3. **(C,D)** DIMMER strongly reduces off-target RNA editing. Average off-target A-to-I RNA editing percentage (C) and the total number of the sequenced adenosines in RNA with A-to-I conversion rate above 0.1% (D) for RSL1D1, CTNNB1, and IP90 across their transcripts. Fold changes show the decrease in off-target editing between the unregulated and the 4×18nt circuit-regulated group. **(E)** DIMMER strongly reduces off-target A-to-I RNA editing percentage of four specific A sites in the RSL1D1 transcript (cf. Fig. S13). In B-E, each dot is a biological replicate, and error bars denote standard deviation. **(F)** DIMMER reduces off-target transcriptome perturbation. Bulk RNAseq of cells containing regulated (4×18nt) and unregulated ABEMax compared to mock transfection. Black dashed lines indicate $|\log_2(\text{fold change})|=0.5$. The red dashed line indicates an adjusted p-value of 0.05. Dots of each group are shown with randomized z-order. Data represent three biological replicates. **(G)** AAV-delivered constructs enable comparison of DIMMER-regulated and unregulated expressions in the mouse brain. Upper panel: H2B-EGFP was expressed with or without the miR-L-based regulation. Both unregulated and regulated groups use the same tissue-specific promoters or enhancers (**Methods**). The H2B-mRuby3 uses the same promoter, and serves as the dosage indicator. The H2B-mRuby3 construct and the H2B-EGFP construct (either unregulated or regulated) were co-transfected in the in-vitro cultured cell lines, or were co-delivered to live mouse brains. Lower panel: Two-photon microscopy experiments were performed to quantify circuit dynamics in live mice brains. The detailed surgery and the imaging procedures are described in **Methods**. **(H)** In vivo brain expression in the CaMKIIα promoter cohort (unregulated and regulated, respectively) measured by microscope. Data were pooled, binned and plotted based on mRuby3 fluorescence intensities on logarithmic (left) or linear (right) scales. See also **Figure 3.21**.

---

hippocampus [53, 54]; a hybrid CMV enhancer–MeCP2 promoter (MeP hereafter), which enables broad neuronal expression throughout the brain [55]; and a minimal β-globin promoter driven by the mDLX2 enhancer (mDLX2 hereafter), which selectively targets GABAergic interneurons in the forebrain, including the motor cortex [56]. As controls, we also delivered a corresponding set of unregulated constructs lacking the DIMMER circuit. All constructs were codelivered with the mRuby3 construct for internal normalization. We injected each pair of constructs into one cohort of mice, and analyzed expression of both fluorescent reporters over time by two-photon head-fixed imaging at single-cell resolution through a previously described cranial implant [57]. over the left motor cortex (**Figure 3.7G lower 23 panel, Methods**). This approach allowed us to monitor the fluorescent dynamics of EGFP and mRuby3 expression in individual cortical cells over timescales of 56-84 days.

To better understand the relationship between promoter strength, expression dynamics, and regu-

latory compensation, we analyzed circuit performance over time (**Figure 3.21B-E**). For all three promoters, fluorescence gradually accumulated over multiple weeks. Notably, the mDLX2 promoter, which is relatively weak and interneuron-specific, showed no evidence of dosage compensation across all timepoints (**Figure 3.21D**). Similarly, the MeP promoter, which drives moderate, pan-neuronal expression, shifted the distribution downward in regulated conditions but did not reach expression levels sufficient to trigger compensation, even at later time points (**Figure 3.21E**). In contrast, the strong CaMKIIα promoter exhibited dosage compensation at later stages (28–56 days). By 28 days, expression from the regulated CaMKIIα construct plateaued and diverged from the unregulated condition, consistent with engagement of post-transcriptional attenuation by the DIMMER circuit (**Figure 3.7H, 3.21B**). These findings suggest that dosage compensation can occur in the brain for some construct designs, that promoter strength and time-dependent expression dynamics are critical for entering the dosage compensation regime, and that weaker promoters may be unable to reach high enough expression levels to activate compensatory repression with this DIMMER variant (**Figure 3.21F**). Modulation of circuit parameters such as the number of miRNA or target sites copies could allow dosage compensation with weaker promoters.

3.4 Discussion

Ectopic gene expression is a cornerstone of modern biology, and gene and cell therapy but precise control has remained elusive in most applications. The miRNA circuits described here achieve precise, sequence-tunable gene dosage-invariant control of protein expression (**Figures 3.2 and 3.3**), orthogonal control of multiple target genes (**Figure 3.4**), and portability across cell types and modes of delivery (**Figure 3.5**). We therefore anticipate that they could become standard systems for controlled gene expression in diverse areas of biomedical science and biotechnology, including imaging (**Figure 3.6**), CRISPR-based gene editing (**Figure 3.7**), and AAV-based gene therapy (**Figure 3.7, Figure 3.21**).

The process of engineering these circuits revealed unexpected design principles. Strong miRNA regulation, as obtained with the 1×21nt circuit, was not sufficient for dosage invariance within targeted expression levels (**Figure 3.1F**). Rather, simultaneously reducing complementarity and multimerizing target sites to engage multivalent TNRC6-dependent regulation was essential. In other regulatory systems, multispecific recognition is associated with ultrasensitivity. Here, however, it allowed linearly sensitive repression of mRNA required for dosage invariance but shifted responses to higher dosages. It will be interesting to learn whether natural miRNA regulatory systems use multispecific binding in similar ways [58, 45]. We also observed that separation of miRNA and target gene into divergently transcribed genes can be useful in allowing strong, independent control of miRNA expression relative to target mRNA (**Figure 3.3B-E**) but is not required for precise expression, facilitating applications like gene therapy where vector capacity

can be limiting (**Figure 3.9**).

The system may be extended or improved in different ways. First, better miRNA sequence-function models could potentially allow predictive design of setpoints. Second, 24 current expression distributions exhibit significant variability, or noise (**Figure 3.3E**). This could reflect transcriptional bursting of the target or the miRNA, and could also be exacerbated by potential differences in the time delays for miRNA production and processing compared to RNA splicing and nuclear export. Going forward, understanding these and other contributions to overall variability could help to reveal fundamental limits of expression precision in the cell [4, 5, 59, 60]. Finally, the ability to combine DIMMER regulation with inducible promoters or natural enhancers for cell type specificity could make these systems even more useful.

The circuits introduced here will be useful in diverse settings. They reduced background in imaging applications (**Figure 3.6**). They also reduced off-target RNA editing by CRISPR base editors, while maintaining on-target editing, suggesting they may be useful for gene editing applications (**Figure 3.7**). Analysis of stronger circuits with even lower expression setpoints could help to achieve even lower off-target edit rates. A major application category is gene therapy (**Figure 3.7G-H, Figure 3.21**). Many monogenic diseases that are gene therapy targets exhibit toxicity at high levels of the therapeutic gene, making it critical to suppress overexpression [61, 62, 63, 64, 65]. In the future, DIMMER circuits could also ensure fixed expression levels for receptors or other components in cell therapies [66] and therefore allow expression of transcription factors and other components at physiological expression levels for regenerative medicine and other applications [67, 68]. Thus, we anticipate these systems becoming useful components in a wide range of engineered research and therapeutic contexts.

Limitations of the study. The DIMMER circuits have limitations. First, we do not yet have a predictive model of how an arbitrary miRNA sequence will behave quantitatively in a dosage compensation circuit. Varying miRNA regulation strength by modulating complementarity was straightforward with miR-L, but more complex for other synmiRs, suggesting that additional factors besides base pairing are likely important for fully predicting the activity of a miRNA on its target. Second, some aspects of the relevant molecular mechanisms remain unclear, including the role of catalytic slicer-dependent regulation [69, 70].

3.5 Acknowledgement

We thank Phillip Zamore (UMass), Josh Mendell (UT Southwestern Medical Center), Acacia Mayfield, James Linton, Kaiwen Luo, Martin Tran, Duncan Chadly, Shiyu Xia, Yodai Takei, Felix Horns, Lucy Chong, Leah Santat, Sheng Wang for discussion and technical support; Inna-Marie Strazhnik (Caltech) for graphical design; Fei Chen (Broad Institute), Rui Malinowski, Evan Mun,

Judy Shon, Jacob Parres-Gold, and other members of the Elowitz lab for critical feedback, and administrative support. This work is supported by the National Institute Of Biomedical Imaging And Bioengineering of the National Institutes of Health under Award Number R01EB030015. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. M.B.E. is a Howard Hughes Medical Institute Investigator. We apologize for incomplete citations due to space limits.

3.6 Data and materials availability

Raw and analyzed data associated with this paper have been deposited at CaltechDATA (data.caltech.edu) and will also be shared by the corresponding author upon request. The link to the CaltechDATA page is https://data.caltech.edu/records/e2fjt-agg74.

3.7 Methods

Plasmids construction

Some constructs were generated using standard cloning procedures. The inserts were generated using PCR or gBlock synthesis (IDT) and were ligated either by T4 ligase (NEB #M0202M) or In-Fusion (Takara #102518) assembly with backbones that are linearized using restriction digestion. The rest of the constructs were designed by the authors and synthesized by GenScript. Selected constructs will be deposited at Addgene and the maps are available.

miRNA alignment to the database

Each synthetic miRNA sequence (mature miRNA, 22 nt) is aligned to the known miRNA sequence database [71, 72, 73, 74, 75, 76, 77, 78] (https://mirbase.org/) to identify if there are any similarities existing between the synthetic sequences and the natural sequences.

Tissue Culture

U2OS cells, T-Rex cells, CHO cells, and N2A cells were cultured at 37 ℃ in a humidity-controlled chamber with 5% CO2. The growth media consisted of DMEM (Dulbecco's Modified Eagle Medium, ThermoFisher #11960-069) supplemented with 10% FBS, 1 U/ml penicillin, 1 μg/ml streptomycin, 1 mM sodium pyruvate, 1×NEAA (ThermoFisher #11140-050), 1 mM L-glutamine, and 0.1 mg/mL Normocin (InvivoGen #ant-nr).

Transient Transfection

Cells were seeded at a density of 50,000 cells in each well of a 24 well plate (or at a density of 10,000 cells in each well of a 96 well plate), either standard for flow cytometry or glass-bottom for imaging experiments, and cultured under standard conditions overnight. The following day, the cells were transiently transfected using Fugene HD (Promega #E2311), according to the manufacturer's

protocol.

Flow Cytometry

Cells were incubated 2 days after transient transfection, and the culture media was replaced 24 hours post-transfection. Cells were trypsinized with 75 μL of 0.25% trypsin for 5 minutes at 37 ℃. After digestion, cells were resuspended with 125 μL of HBSS containing 2.5 mg/ml BSA and 1 mM EDTA. Cells were then filtered through a 40 μm cell strainer and analyzed using a CytoFLEX S instrument (Beckman Coulter). We used the FlowJo V10 and self-build python code to analyze the flow data. The regression curves and the confidence intervals in **Figure 3.5B** and **Figure 3.16** were computed using the statsmodels.regression.linear_model.OLS package in logarithm spaces.

Cell sorting

To prepare the mono-clones that expressed the genomic-integrated DIMMER circuit, cells were harvested and resuspended in sorting buffer (BD FACS Pre-Sort Buffer) supplemented with 1 U/ml DNAse I by the cell sorter (Sony MA900) as mono-clones. Cells were sorted into 96 well plates in the normal U2OS culture media. Cells were expanded in the 24 well plate before flow cytometry measurement.

In-vitro image analysis

The transiently transfected U2OS cells were imaged using a Nikon confocal microscope at 60× magnification, such that each image was spaced by 0.5 microns in the z-direction. Images were processed by the Fiji software [79].

To analyze the relative signal intensity in the dCas9 imaging experiment, maximum intensity projection of 11 slices of the z-stacks were applied. To determine the signal intensity of the dots, freehand lines were drawn to select the dot regions. To determine the signal intensity in the background, ~5 micron-long straight lines centering the dots were drawn. The signal intensities were generated by the Fiji ROI mean intensity function. The relative signal intensity is calculated by normalizing the background intensity to be 1. To analyze the signal to noise ratio (SNR), freehand lines were drawn to select the dot regions and the nucleus regions. The noise intensity was calculated by the intensity in the nucleus excluding the dots area. The SNR was calculated dividing the mean intensities of each dot by the noise.

Bulk RNA sequencing to identify the off-target effects of the synthetic miRNA on the transcriptome

Sample preparation and sequencing

To verify the off-target effect of all the synmiRs, U2OS cells were plated on 6-well plates with 300,000 cells per well. Cells were transfected the following day with 1,000 ng of either the control plasmid or the BFP-miRNA plasmid using Fugene HD (Promega #E2311) according to the manufacturer's instructions. Media was replaced with 2 mL of fresh media 24 hours post-transfection. Cells were harvested 48 hours post-transfection by digestion with 0.25% Trypsin-EDTA, centrifugation at 300g for 5 minutes, and removal of the supernatant by aspiration. The cell's pellet was stored in -80 ℃ prior to the purification. RNA was extracted using the RNeasy kit (Qiagen #74106) according to the manufacturer's instructions. RNA was treated with Turbo DNase (ThermoFisher #AM2238) and purified using the RNeasy kit RNA cleanup protocol. mRNA sequencing libraries were prepared by Novogene.

Preprocessing of sequencing data

Reads from the RNA sequencing were aligned to a custom reference genome using kallisto (0.48.0) [80]. This reference consisted of the human genome GRCh38 cDNA [1] and mTagBFP2 coding sequences. Weakly expressed genes were filtered out if they exhibited fewer than 3 samples expressing at least 10 transcripts per million (TPM), or if the maximum TPM among all samples was less than 105. Then, filtered counts were input to DEseq to eliminate the impact of size factors. As the BFP-only cells were used as a reference to evaluate the off-target effect of the miRNAs, genes that showed fluctuating expressions among the three biological replicates of BFP-only cells should be removed from analysis. To achieve this, we computed log(1+x), where x denotes the normalized TPM among the three biological replicates of BFP-only cells. The Fano factors [81, 82] of these logarithmic expressions were determined and ranked. Transcripts that ranked as the largest 2.5% in logarithmic Fano factors were eliminated from further analysis. Finally, we computed log(1+x), where x denotes the normalized TPM among all the samples. A difference function was defined to compute the absolute value of the log(1+x) difference between each sample and the untransfected sample. The medians of the difference function of the BFP-only groups and the experimental groups were calculated and used for comparison. The difference between those two difference functions were ranked and similarly, transcripts that ranked as the largest 3% were removed from further analysis.

[1] https://github.com/pachterlab/kallisto-transcriptome-indices/releases

The Fano factor is defined as

$$\text{Fano factor} = \frac{\text{variance}(\log(1+x))}{\text{mean}(\log(1+x))}. \tag{3.1}$$

The equation of the difference function is defined as:

$$\Delta(\text{BFP, untransfected}) = |\log(1+x_{\text{BFP}}) - \log(1+x_{\text{untransfected}})| \tag{3.2}$$

$$\Delta(\text{experimental, untransfected}) = |\log(1+x_{\text{experimental}}) - \log(1+x_{\text{untransfected}})| \tag{3.3}$$

$$\Delta_{\text{ranked}} = \text{median}(\Delta(\text{BFP, untransfected})) - \text{median}(\Delta(\text{experimental, untransfected})). \tag{3.4}$$

Differential gene expression analysis

To characterize the perturbations that synthetic miRNA brought to the endogenous transcriptome, differential expression analysis was performed using DESeq2 (1.40.1) [83] in R (4.3.1) comparing transcript counts in miRNA transfected cells and BFP-only cells.

DNA-PAINT

Buffers

The following buffers were used:

- NH_4Cl solution: NH_4Cl (Roth, no. K298.1) was dissolved in ddH_2O for a 2 M stock solution, filtered with 0.2 μm filter.
- Blocking buffer: 1×PBS, 1 mM EDTA (Thermo Fisher, no. AM9260G), 0.02% Tween-20 (Life Science, no. P7949), 0.05% NaN_3 (Serva, no. 30175.01), 2% BSA (Sigma-Aldrich, no. A9647-100G), 0.05 mg/ml sheared salmon sperm DNA (Life Technologies, no. 15632011), filtered with 0.2 μm filter.
- Imaging buffer: 1×PBS, 1 mM EDTA, 0.02% Tween-20, 500 mM NaCl (Thermo Fisher, no. AM9760G), supplemented with PCA, PCD, trolox, filtered with 0.2 μm filter.

PCA, PCD, and Trolox

Trolox (100×) was prepared by dissolving 100 mg (±)-6-hydroxy-2,5,7,8-tetra-methylchromane-2-carboxylic acid (trolox; Sigma Aldrich, 238813-5G) in 430μL of 100% methanol (Sigma-Aldrich, 32213-2.5L), 345μL of 1M NaOH (VWR, 31627.290) and 3.2mL of water.

PCD (40×) was made by mixing 154 mg of 3,4-dihydroxybenzoic acid (PCA; Sigma-Aldrich, 37580-25G-F) in 10 ml of water and NaOH and adjusting the pH to 9.0. PCD (100×) was prepared

by adding 9.3 mg of protocatechuate 3,4-dioxygenase pseudomonas (PCD; Sigma-Aldrich, P8279) to 13.3 ml of buffer containing 100 mM Tris-HCl pH 8.0 (Thermo Fisher Scientific, AM9855G), 50 mM KCl (Thermo Fisher Scientific, AM9640G), 1 mM EDTA, and 50% glycerol (Sigma-Aldrich, 65516-500ml)).

Cloning

An mEGFP gBlock (obtained from IDT) was inserted into a pcDNA3.1(+) backbone (Thermo Fisher, no.V79020) via Gibson assembly. Two codon-optimized fragments of human EGFR (obtained from IDT) were fused to the mEGFP-pcDNA3.1(+) backbone via Gibson assembly. The plasmid concentration was measured with the NanoDrop One (Thermo Scientific).

Cell Culture

CHO-K1 cells (ATCC: CCL-61) were cultured in Ham's F-12K (Kaighn's) medium (Gibco, no. 21127022) supplemented with 10% FBS (Gibco, no. 11573397). U2OS-CRISPR-Nup96-mEGFP cells (a gift from the Ries and Ellenberg laboratories) were cultured in McCoy's 5A medium (Gibco, no. 16600082) supplemented with 10% FBS. All cells were cultured at 37 ℃ and 5% CO_2 and split every 2-3 days via trypsinization using trypsin-EDTA (Gibco, no. 25300096).

Nanobody-DNA conjugation

First, the anti-GFP nanobody (clone 1H1, Nanotag Biotechnologies, N0305) and anti-rabbit IgG nanobody (Nanotag Biotechnologies, N2405) were conjugated to a DBCO-PEG4-Maleimide linker (Jena Bioscience, no. CLK-A108P). After removing the unreacted linker with Amicon centrifugal filters (10,000 MWCO), the DBCO-nanobody was conjugated via DBCO-azide click chemistry to the docking strand (Metabion, see sequence in Table 1). A detailed description of the conjugation can be found in the former work [84].

Fixation of cells

The cells were fixed with 37 ℃ pre-warmed methanol-free 4% PFA (Thermofisher, no. 043368.9M) in 1×PBS for 15 min. Then, the cells were washed 3 times with 1×PBS and then permeabilized with 0.125% TritonX-100 (Sigma Aldrich, no. 93443) in 1×PBS for 2 min. After washing 3 times with 1×PBS, the cells were blocked with the blocking buffer either overnight or for at least 3h at 4 ℃.

Sample preparation of U2OS cells

10,000 cm^{-2} U2OS cells were seeded on an ibidi eight-well high glass-bottom chambers (no. 80807). On the next day, the cells were fixed as stated in the fixation protocol and blocked for 3h at 4 ℃. After washing 3 times with 1×PBS, the primary EGFR antibody (Cell Signaling, clone D38B1, no. 4267) with a dilution of 1:200 in blocking buffer was incubated overnight at 4 ℃. The next day, the sample was washed 3 times with 1×PBS and 25nM of R2 anti-rabbit NB in the blocking buffer was incubated at RT for 1h. After washing 3 times with 1×PBS, the sample was post-fixed with 4% PFA in 1×PBS for 5 min at RT. The cells were then quenched with 200mM NH_4Cl (Roth, no. K298.1) for 5 min and washed 3 times with 1×PBS. 90nm gold-nanoparticles (Absource, no. G-90-100) in 1:1 in 1×PBS were incubated for 5 min at RT. After washing 3 times with 1×PBS, the cells were washed once with the imaging buffer.

Sample preparation for plasmid dosage measurement

5,000 cm^{-2} CHO-K1 cells were seeded on an ibidi eight-well high glass-bottom chambers (no. 80807) one day before transfection. The cells were transfected with EGFR-meGFP plasmids with a Thermo Fisher Lipofectamine 3,000 reagent (no. L3000008) with the lower Lipofectamine concentration as indicated by the manufacturer and different plasmid concentrations (0 ng, 125 ng, 250 ng and 500 ng of plasmid per well). After 48h of transfection, the cells were fixed as indicated in the fixation protocol and blocked with blocking buffer for 3h at 4 ℃. The cells were then post-fixed with 4% PFA and 0.2% glutaraldehyde (Serva, no. 23115.01) in 1×PBS for 10 min. After quenching the sample with 200mM NH_4Cl for 5 min and washing 3 times with 1×PBS. 90 nm gold-nanoparticles in 1:1 in 1xPBS were incubated for 5 min at RT. After washing 3 times with 1×PBS, the cells were washed once with the imaging buffer.

Sample preparation of DIMMER plasmids

5,000 cm^{-2} CHO-K1 cells were seeded on an ibidi eight-well high glass-bottom chambers (no. 80807) one day before transfection. The cells were transfected with EGFR-mEGFP plasmids with a Thermo Fisher Lipofectamine 3,000 reagent (no. L3000008) with the lower Lipofectamine concentration as indicated by the manufacturer and 250 ng plasmid per well (200μL solution per well and 25μL transfection solution). After 48h of transfection, the cells were fixed as indicated in the fixation protocol and blocked with the blocking buffer for 3h at room temperature (RT). 25nM R3 anti-GFP nanobodies were incubated in the blocking buffer for 1h at RT. After washing 3 times with 1×PBS, the nanobodies were post-fixed with 4% PFA and 0.2% glutaraldehyde in 1×PBS for 10 min. The cells were then quenched with 200mM NH_4Cl for 5 min and washed 3 times with 1×PBS. 90nm gold-nanoparticles (Absource, no. G-90-100) in 1:1 in 1×PBS were incubated for 5

min at RT. After washing 3 times with 1×PBS, the cells were washed once with the imaging buffer and imaged in the imaging buffer.

DNA-PAINT imaging

The samples were imaged in the imaging buffer with the corresponding imager strand (obtained from Metabion, see table 1 for imager strand sequences) for 40k frames with 100ms exposure time per frame and a readout rate of 200 MHz.

Microscope setup

The samples were measured on inverted total internal reflection fluorescence (TIRF) microscopes (Nikon Instruments, Eclipse Ti2) which are equipped with an oil-immersion objective (Nikon Instruments, Apo SR TIRF ×100/numerical aperture 1.49, oil) and a perfect focusing system. The mRuby3 signal was bleached by the 560 nm laser (MPB Communications, 1 W) by using Highly inclined and laminated optical sheet (HILO) illumination. Afterwards, the TIRF mode was established. The Cy3B-conjugated imagers were excited with the 560 nm laser. The laser beam was cleaned with a filter (Chroma Technology, no. ZET561/10) and coupled into the microscope with a beam splitter (Chroma Technology, no. ZT561rdc). The fluorescent signal was filtered with an emission filter (Chroma Technology, nos. ET600/50m and ET575lp) and projected onto a sCMOS camera (Hamamatsu Fusion BT) without further magnification. The camera's central 1152×1152 pixels (576×576 pixels after binning) were used as the region of interest, with a resulting effective pixel size of 130nm. The raw microscopy data was acquired via μManager (Version 2.0.1).

DNA-PAINT analysis

Obtained fluorescent data was reconstructed with Picasso software [35]. The data was first drift-corrected with redundant cross-correlation, after that with picked gold particles as fiducials.

In order to determine the receptor density, a homogeneous area of the cells was picked and the DNA-PAINT data was clustered with the SMLMS clustering algorithm of Picasso [35, 85].

The identified cluster centers were used to calculate the measured receptor density per μm^2 (number of cluster centers per area). Given that the labeling efficiency (LE) of the binders to their targets was less than 100%, a correction factor was applied to account for incomplete labeling. The EGFR receptor density was calculated by multiplying the measured receptor density with the LE of the respective binders. The LE values of the binders were determined as previously described in the former work [86]. Specifically, the LE of the anti-GFP nanobody was determined to be 37%, while the LE of the EGFR antibody was 71%.

Statistical testing

To test the statistical significance of the determined receptor densities, we performed the Mood's median test. Mood's median test is a non-parametric statistical test which tests whether the median of two groups are statistically different. Our statistical tests were performed with a custom python script using functions from the scipy.stats module [87].

Table 3.1: DNA sequence of docking strands and imager strands

Sequence name	Docking strand sequence (5' to 3')	Imager strand sequence (5' to 3')
R2	ACCACCACCACCACCACCA	TGGTGGT-Cy3B
R3	CTCTCTCTCTCTCTCTCTC	GAGAGAG-Cy3B

Base editor experiment

Transfection and sample collection

HEK293 cells were seeded at the density of 150k/well in the 24-well plate one day before the transfection. For each well, 750 ng ABEMax base editor plasmid and 250 ng sgRNA plasmid were co-transfected using Fugene HD (Promega #E2311), according to the manufacturer's protocol. Cells were harvested 72 h post-transfection.

Amplicon sequencing and data analysis

To quantify the on-target base editor editing rate, cells were sorted based on the constitutively-expressing mRuby on the base editor plasmid and EGFP on the sgRNA plasmid. Genomic DNA was extracted using Qiagen DNeasy kit (cat. nos. 69504) according to the manufacturer's protocol. The genomic DNA amplicons were amplified using the primer in the reference [47], with the adaptor sequence on each end. Amplicons were size-verified by DNA electrophoresis, purified by the Qiagen gel purification kit, and sequencing was performed by Genewiz Amplicon-EZ (150-500bp) service. The on-target editing rates were analyzed using the online tool CRISPResso2 [88].

To quantify the off-target base editor editing rate on the genome, genomic DNA was extracted, and amplicons were obtained using the potential genomic DNA off-target loci-targeted primers reported previously [47]. To quantify the off-target editing on the RNA level, the total RNA was extracted from the cell using the Zymo Direct-zol RNA miniprep kit, according to the manufacturer's protocol. cDNA samples were prepared using the Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Scientific). Amplicons were obtained using the potential RNA off-target-loci-targeted primers reported previously [47]. The amplicons were then size-verified by DNA electrophoresis, purified with the Qiagen gel purification kit, barcoded with the Illumina Miseq 16S Metagenomic

sequencing index primers, and further purified with the magnetic NGS beads (Omega Bio-Tek). Library was quantified and normalized with the Qubit fluorometer. Library was then denatured and sequenced with Element AVITI System Sequencing Instrument using AVITI 2×150 Sequencing Kit Cloudbreak (Catalog # 860-00013). The off-target editing rates were analyzed using the online tool CRISPResso2 [88].

Bulk RNA sequencing

- To analyze the transcriptome-wide perturbation of the base editor, total RNA was extracted from the cells with the Zymo Direct-zol RNA miniprep kit (lot # R2050). 50 ng of extracted mRNA from each sample were used as inputs for downstream NGS library preparation.

- mRNA-seq libraries were prepared in 96-well format with a modified 3'Pool-seq protocol [89]. In brief, reverse transcription reaction were prepared by mixing input RNA with 1 μl Indexed RT Primer (10 μM), 1 μl 10 mM dNTP Mix (New England Biolabs Cat# N0447S),1 μl diluted ERCC Spike-In Mix 1 (0.004 μL stock ERCC per μg RNA, ThermoFisher Cat# 4456740), 3.6 μl of 5×RT buffer (ThermoFisher Cat# EP0752), 0.5 μl of RNAse inhibitor (Thermofisher Cat# EO0381), 1 μl Maxima RT H minus (Thermofisher Cat# EP0752), 2.5 μl 10 uM Template Switching Oligo into a 18 μl reaction. Reverse transcription was carried out in a thermocycler with a program described in 3'Pool-seq protocol.

- Samples from each row of 96-well plate were pooled (column pooling) by mixing an equal volume of each Reverse Transcription reaction into a new well at a total volume of 20 μl. Residual primers were then degraded with the addition of 1 μl Exonuclease I (New England Biolabs) and incubated at 37 ℃ for 45 min followed by denaturation at 92 ℃ for 15 min. Subsequent cDNA amplification, tagmentation, and row pooling was performed following 3'Pool-seq protocol.

- Finally, 20 μl of pooled NGS library were subject to Gel-based size selection using E-Gel EX Agarose Gel (ThermoFisher Cat# G401001) to enrich for fragments with size range between 200-1000 bp and eluted in 15 μl.

- Eluted pooled NGS libraries were examined in an Agilent TapeStation 4200 (Agilent Technologies) to determine average fragment sizes. Library concentration was quantified in a Qubit 3.0 Fluorometer (Life Technologies). NGS library molarity was then calculated using 660 g/mol per base-pair as a molecular weight. NGS libraries were diluted to 2 nM, denatured in 0.2N NaOH, and loaded onto Element AVITI sequencer following Element Biosciences Cloudbreak Sequencing user guide.

Data analysis

- Weakly expressed genes were filtered out if they exhibited fewer than 3 samples expressing at least 10 transcripts per million (TPM). Differential expression analysis was performed using DESeq2 (1.40.1) [83] in R (4.3.1). Identified significantly differentially expressed genes were eliminated from further analysis if they were identified among the unregulated, regulated, and untransfected groups with the same trend, which indicates that these genes tend to be expressed as more similar profiles among the base editor-transfected groups and the WT group, thus very likely to be the artifact of the filler-plasmid expression. GO annotation was performed using the DAVID web server (https://david.ncifcrf.gov/).

DNA-FISH imaging

To cross-validate the success of DIMMER-regulating dCas9 labeling the telomeres, Nikon 2 Multi-point acquisition was used to record the absolute coordinates of the fields of view. After the live cell imaging, the plate was removed from the scope and DNA-FISH was performed. Dye-conjugated probes designed to bind to the distal telomeric regions of the chromosome were used following the previously reported protocol [90, 91]. The plate was then loaded on the scope, and the recorded coordinates were retrieved. The exactly same fields of view were matched by eye around the recorded coordinates.

DIMMER circuit dynamical quantitation in mice cranial window

AAV viral production

All constructs were cloned using the cloning strategy described above. The viral titers of the viruses used in this study:

Group	Cargo	Viral Titers
CaMKIIα_unregulated	CAP-B10.CaMKIIα.H2B.EGFP.sv40pA	1.06×10^{14}
	CAP-B10.CaMKIIα.H2B.mRuby3.sv40pA	2.12×10^{14}
CaMKIIα_regulated	CAP-B10.CaMKIIα.H2B.EGFP(4×19nt regulated).sv40pA	8.11×10^{13}
	CAP-B10.CaMKIIα.H2B.mRuby3.sv40pA	2.12×10^{14}
MeP_unregulated	CAP-B22.CMVenhancer.MEP229.H2B.EGFP.sv40pA	1.22×10^{14}
	CAP-B22.CMVenhancer.MEP229.H2B.mRuby3.sv40pA	4.04×10^{13}
MeP_regulated	CAP-B22.CMVenhancer.MEP229.H2B.EGFP(4×19nt regulated).sv40pA	4.14×10^{14}
	CAP-B22.CMVenhancer.MEP229.H2B.mRuby3.sv40pA	1.94×10^{14}
mDLX2_unregulated	CAP-B10.mDLX2enhancer.H2B.EGFP.sv40pA	1.53×10^{14}
	CAP-B10.mDLX2enhancer.H2B.mRuby3.sv40pA	1.64×10^{14}
mDLX2_regulated	CAP-B10.mDLX2enhancer.H2B.EGFP(4×19nt regulated).sv40pA	7.79×10^{13}
	CAP-B10.mDLX2enhancer.H2B.mRuby3.sv40pA	1.64×10^{14}

CAP-B22 and CAP-B10 are AAV capsids developed in the previous study [51]. The unit of the viral titers is viral genome copies per mL.

Animals

All animals used in this study were approved by the Institutional Animal Care and Use Committee (IACUC) at the California Institute of Technology. Wildtype C57BL female mice (1.5-4 months of age) were ordered from Jackson Laboratories (Bar Harbor, ME, USA) and were used for surgeries, viral injection, and head-fixed imaging. All mice were singly-housed post-operation and for the remainder of the data acquisition in a room with a reverse light cycle (12h – 12h). For all imaging sessions, mice were imaged during the dark phase of their light cycle.

Cranial Window Surgeries

Mice were deeply anesthetized under 1-2% isoflurane and given subcutaneous injections of ketofen (5 mg/kg), buprenorphine (3.25 mg/kg), dexamethasone (2 mg/kg) and saline for analgesia and hydration. A subcutaneous injection of a few drops of bupivacaine was given under the incision site for local analgesia. An incision was performed over the skull and a piece of skin near the left motor cortex was removed to make room for the implant. A circular craniotomy of approximately 3-mm in diameter was drilled over the left motor cortex and centered 1.6-mm anterior and 1.6-mm lateral from bregma. Saline was flushed periodically throughout the drilling over the bone to hydrate and prevent heating. After extraction of the cranial bone, a gelatin sponge soaked in saline was applied over the exposed dura to remove debris and stop any residual bleeding. A small drop of Kwik-sil silicone elastomer was applied over the dura and a 3-mm diameter glass coverslip (#1.5) was implanted over the craniotomy and sealed with dental cement. A custom-made titanium headbar was fixed using cement over the window to allow for later head-fixation during 2-photon imaging. Following surgery, mice were allowed to recover for at least 4-days before beginning habituation to a rotary platter setup.

AAV Vectors and Administration in Mice

AAV packaging, purification, and delivery were performed as described in the earlier study [92]. Following surgical implants of cranial windows and recovery, mice were anesthetized under 1-2% isoflurane. Intravenous administration of AAVs was performed by injection into the retro-orbital sinus of mice. All vectors were each delivered at a total of 5×10^{12} viral genomes (VG) in a volume of 50 μL per mouse from 2-4 months of age. A total of three cohorts were designed to test for the performance of the DIMMER circuits with different gene-regulatory elements: CMV-enhancer paired with a minimal MeCP2 promoter (CMV enhancer-MEP229), an mDLX2 enhancer paired

with a minimal beta-globin promoter (mDLX2-minBglobin), and a CaMKIIα promoter. Each cohort had two groups: One group coexpressing an unregulated H2B-EGFP transgene and an unregulated H2B-mRuby3 transgene, and a regulated group co-expressing a 4×19nt DIMMER regulated H2B-EGFP transgene and an unregulated H2B-mRuby3 transgene. These transgenes were co-delivered separately in their own AAV-capsids. The mRuby3 signal was later used as an internal reference for the EGFP signal for each mouse.

Two-Photon Head-Fixed Imaging and Acquisition

Mice were habituated to the handler's hand for 3 days. Following this, they were head-fixed onto a rotary platter for an additional 3 days to further habituate to the imaging setup. After habituation, mice were imaged over a variable number of days and weeks to capture the progression of EGFP and mRuby3 expression. We used a custom home built 2-photon microscope equipped with a galvo-galvo scanner for imaging. Each imaging session involved acquiring two high-resolution Z-stacks per mouse using a 20X/1.0 NA objective (Zeiss W Plan-Apochromat, Cat. No. 421452-9700-000) with water immersion. Each of these Z-stacks consisted of 10 Z-steps, ranging from 200-250 μm to 300-350 μm below the dura, with each Z-step separated by 10 μm intervals, for a final 100 μm thick Z-stack. Each Z-step was acquired at a scan rate of 1 Hz with a resolution of 256 by 256 pixels and a pixel dwell time of 13 μs for a total of 180 seconds, over an area of 400 μm × 400 μm. Z-steps in each Z-stack were later processed as mean-intensity projected TIF files to generate a single Z-step image. Each Z-stack was first acquired at 940 nm wavelength (EGFP), followed by 1100 nm (mRuby3). The average laser power beneath the objective was maintained across imaging sessions at 36-37 mW for the 940 nm wavelength and 46-47 mW for the 1100 nm wavelength.

Two-Photon Data Analysis

Two-photon Z-stacks were acquired separately for EGFP and mRuby3 channels and processed independently. Raw data files were segmented into individual z-step movies based on predefined frame intervals using a custom MATLAB script, with each segment saved as a multi-page TIFF file. Motion correction was applied to each TIFF using rigid registration, and motion-corrected data were used to generate mean- and maximum-intensity projection images for each z-step. Following motion correction, mean-intensity projection images were flatfield corrected using MATLAB's imflatfield function with a specified radius of 20 pixels to remove uneven illumination artifacts. Corrected images were saved in a dedicated output directory for subsequent analysis. To ensure spatial correspondence between channels, each flatfield-corrected mRuby3 image was aligned to its corresponding EGFP image using normalized cross-correlation, correcting for lateral drift and maximizing overlap between channels.

Cell segmentation was performed on the mRuby3 channel using the Cellpose deep learning-based segmentation algorithm (model type = "cyto", diameter = 9.0 pixels), generating individual cell masks for each z-step image. These masks were saved in both TIFF and NumPy formats and served as a consistent reference for fluorescence quantification across channels. For each z-step, the corresponding mRuby3-derived cell mask was applied to the EGFP image to extract the total EGFP fluorescence per cell by summing the pixel intensities within each ROI. Separately, a dedicated script was used to quantify total mRuby3 fluorescence per cell by applying the same cell mask to the original mRuby3 image for that z-step. Results were saved as per-z-step .csv files, with all individual pixel values archived in cumulative text files for downstream statistical analysis and visualization.

To evaluate single-cell expression and dynamics across time, total EGFP and mRuby3 fluorescence values per cell were aggregated across multiple mice and imaging timepoints. For the CaMKIIα and CMVe-MeCP2 cohorts, z-steps 1, 5, and 10 were selected from each 10-step Z-stack to reduce the likelihood of overcounting the same cells across adjacent planes. For the mDLX2-minβglobin cohort, only z-step 5 was selected.

3.8 Supplementary figures

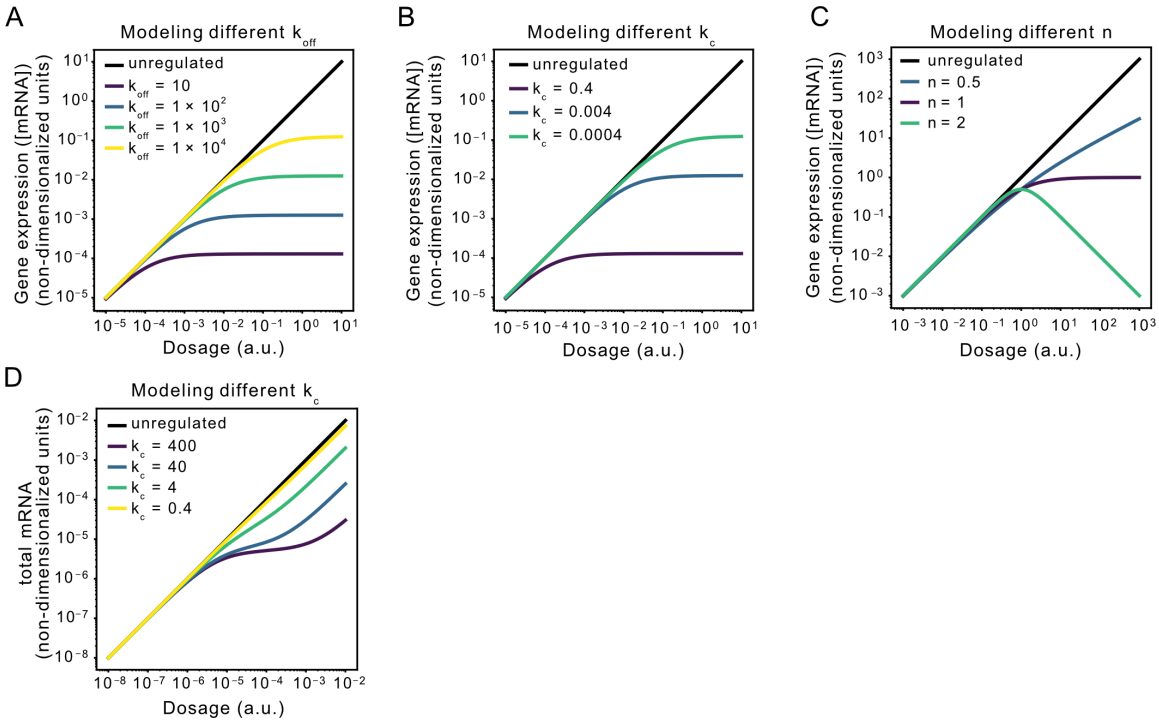

Figure 3.8: **The minimal model of the miRNA-mediated IFFL. (A)** Modeling the steady-state, dimensionless mRNA concentration under different k_{off} while maintaining $k_{on} = 2 \times 10^5$, $k_c = 0.4$. **(B)** Modeling the steady-state, dimensionless mRNA concentration under different k_C while maintaining $k_{on} = 2 \times 10^5$, $k_{off} = 10$. **(C)** Modeling the steady-state, dimensionless mRNA concentration under different HIll coefficient n while maintaining $K = 1$. **(D)** Modeling the steady-state, dimensionless total mRNA amount (including the free mRNA and the bounded complex) under different k_c while maintaining $k_{on} = 2 \times 10^5$, $k_{off} = 10$ by considering the complex degradation caused by natural degradation of either mRNA or miRNA, and the translational contribution of the bounded complex.

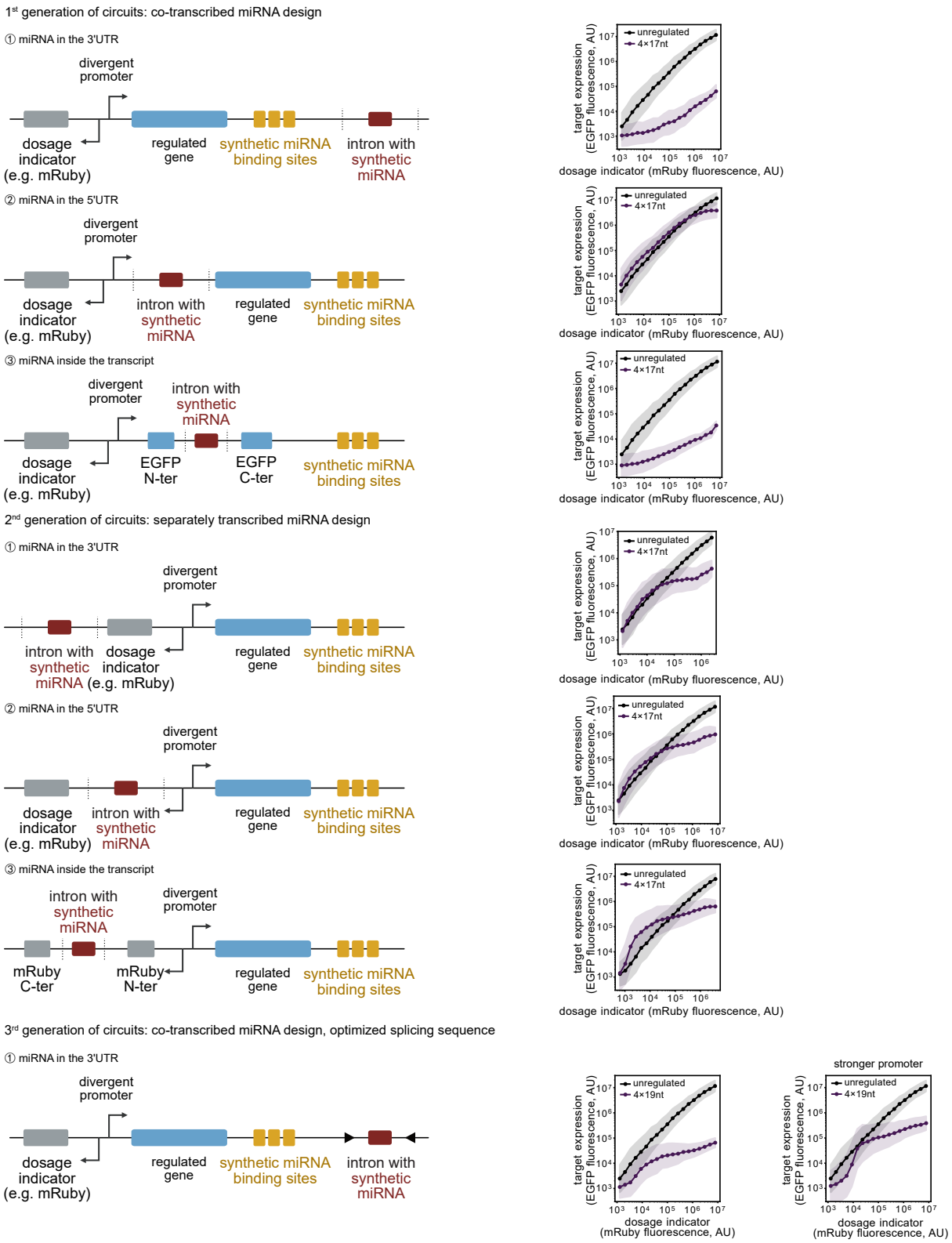

Figure 3.9: (Caption on next page)

Figure 3.9: **The iterative engineering process of different circuit architectures.** Left panel, various circuit configurations; right panel, corresponding circuit quantitative behaviors measured by flow cytometry. In the first generation of the circuit, a co-transcribed miRNA design was used. Placing the intronic synthetic miRNA in the 3'-UTR or in the middle of the transcript both lead to the strong repression effects that might be caused by the destabilization of the transcript that is independent of the miRNA inhibition. Placing the intronic synthetic miRNA in the 5'-UTR of the transcript does not produce enough miRNA.
In the second generation of the circuit, the intronic synthetic miRNA sequence was placed in the separated transcript (which is mRuby3). No matter where the miRNA sequence is inserted, the dosage compensation behavior is maintained. However, when the miRNA sequence is placed into the 3'-UTR, the amount of the mature miRNA seems to be the greatest, since the circuit has the lowest setpoint.
In the third generation of the circuit, the intronic synthetic miRNA sequence was placed in the same transcript as the regulated EGFP. The key difference between the third generation design and the first generation design is that the splicing signal sequence is optimized (indicated by the black triangles). The dosage compensation behavior is achieved, and the utilization of a stronger promoter changes the setpoint.

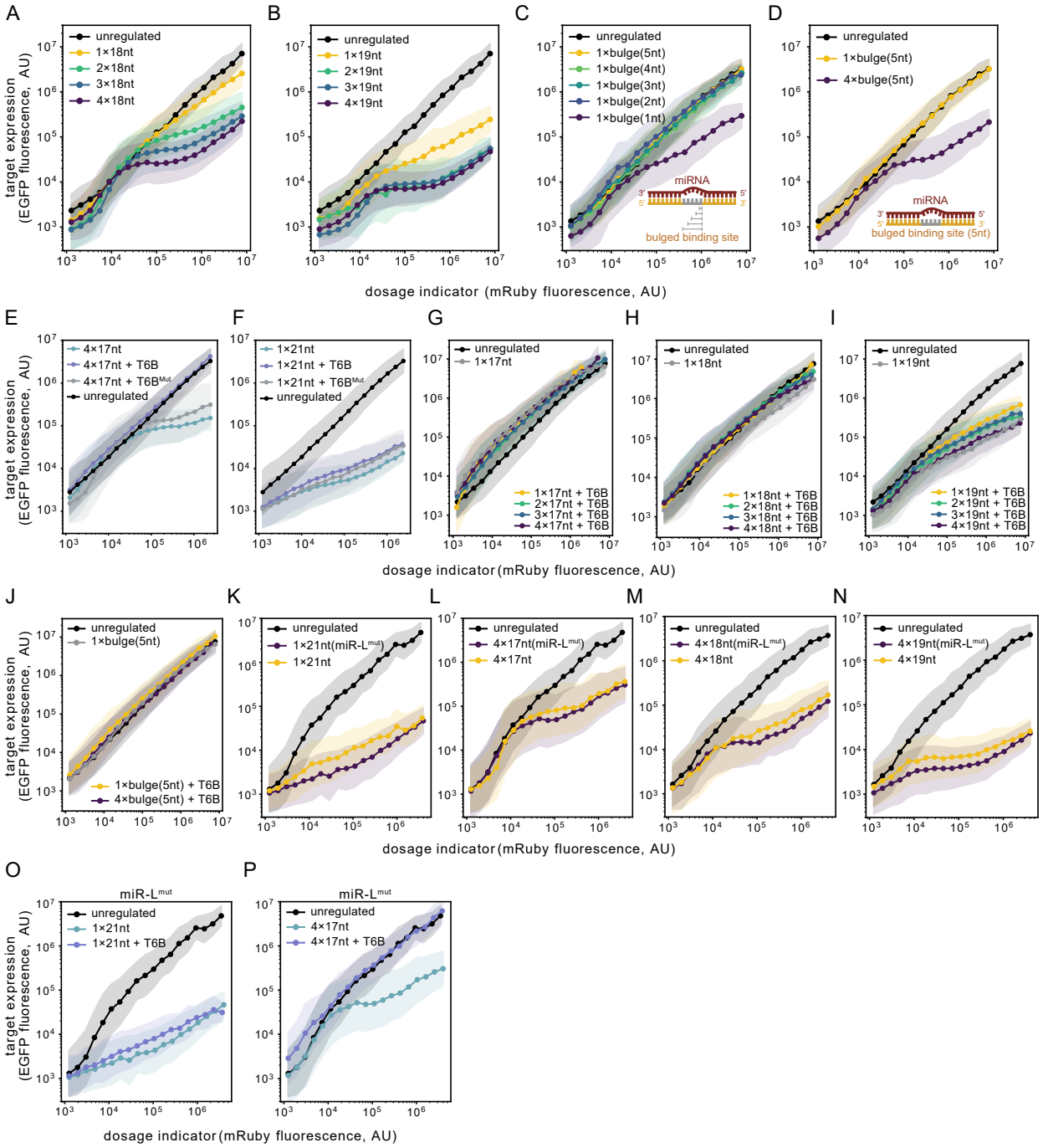

Figure 3.10: (Caption on next page)

Figure 3.10: **The dosage response curves of the miR-L and miR-L^{mut}-based targets and their dependencies of the TNRC6. (A-B)** The performance of the miR-L-mediated 1-4×18nt constructs (A) and 1-4×19nt constructs (B), measured by flow cytometry. **(C-D)** The performance of the bulged target constructs either in the single copy (C) or in the multiple copy (D) configurations. The bulge designs were shown as the insets in the center of the plots. The bulge region starts from the 9th nucleotide on the target, and the lengths of which range from 1 nt to 5 nt. **(E-F)** We performed flow cytometry on the cells transfected with the miR-L 4×17nt construct (E) or miR-L 1×21nt circuit (F) described in **Figure 3.2B-C** along with the fluorescent protein-only negative control, the T6B peptide, or the catalytically dead T6B peptide (denoted as T6B^{Mut}), respectively. The design of the catalytically dead T6B peptide was consistent with the previous literature [16]. **(G-J)** We performed flow cytometry on the cells co-transfected with the labeled constructs and the T6B peptide or the control described in **Figure 3.2B-C** to identify the TNRC6 dependence of different constructs. T6B restored the multiple-site targets to the similar level of the corresponding single-site targets. **(K-N)** We performed flow cytometry on the cells transfected with the constructs regulated by the miR-L^{mut}, and compared them with the corresponding constructs regulated by the perfectly matched miRNA. The inclusion of the central mismatch in the miRNA slightly enhanced the regulation. **(O-P)** We performed flow cytometry to measure the TNRC6 dependence of the 1×21nt (O) and the 4×17nt (P) constructs, respectively. Similar TNRC6 dependence was observed for these miR-L^{mut}-regulated constructs compared to that of the perfectly matched miR-L-regulated constructs shown in **Figure 3.2B-C**.

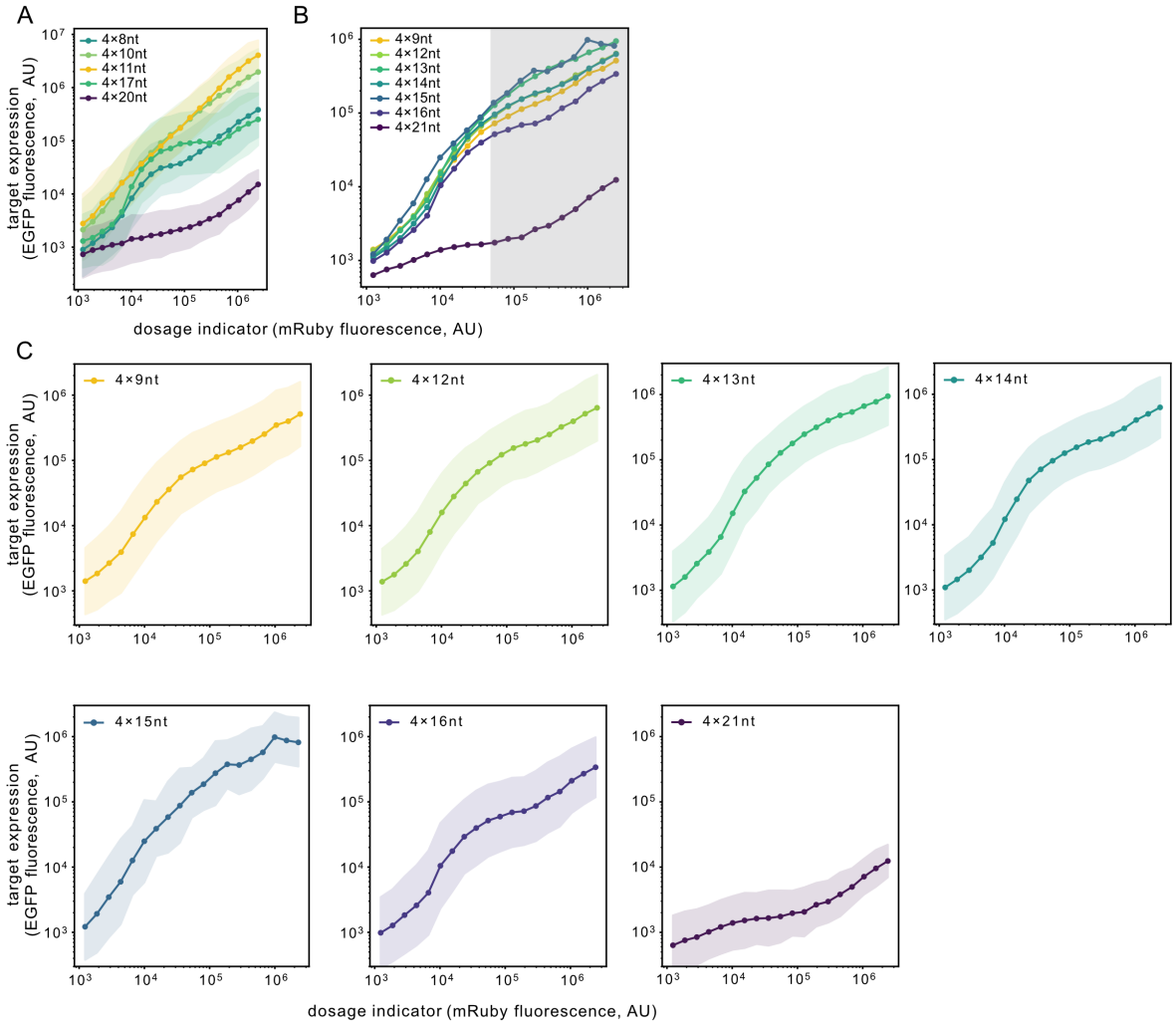

Figure 3.11: **The dosage response curves of multiple miR-L 4×n circuits (A)** The curves in **Figure 3.3A** right panels with the geometric variances plotted. **(B)** We performed flow cytometry on the cells transfected with the miR-L ing 4×n circuits described in **Figure 3.3A**. Gray rectangle indicates the gated region of the **Figure 3.3A** left panel. **(C)** The curves in (B) with the geometric variances were plotted separately to avoid being cluttered together.

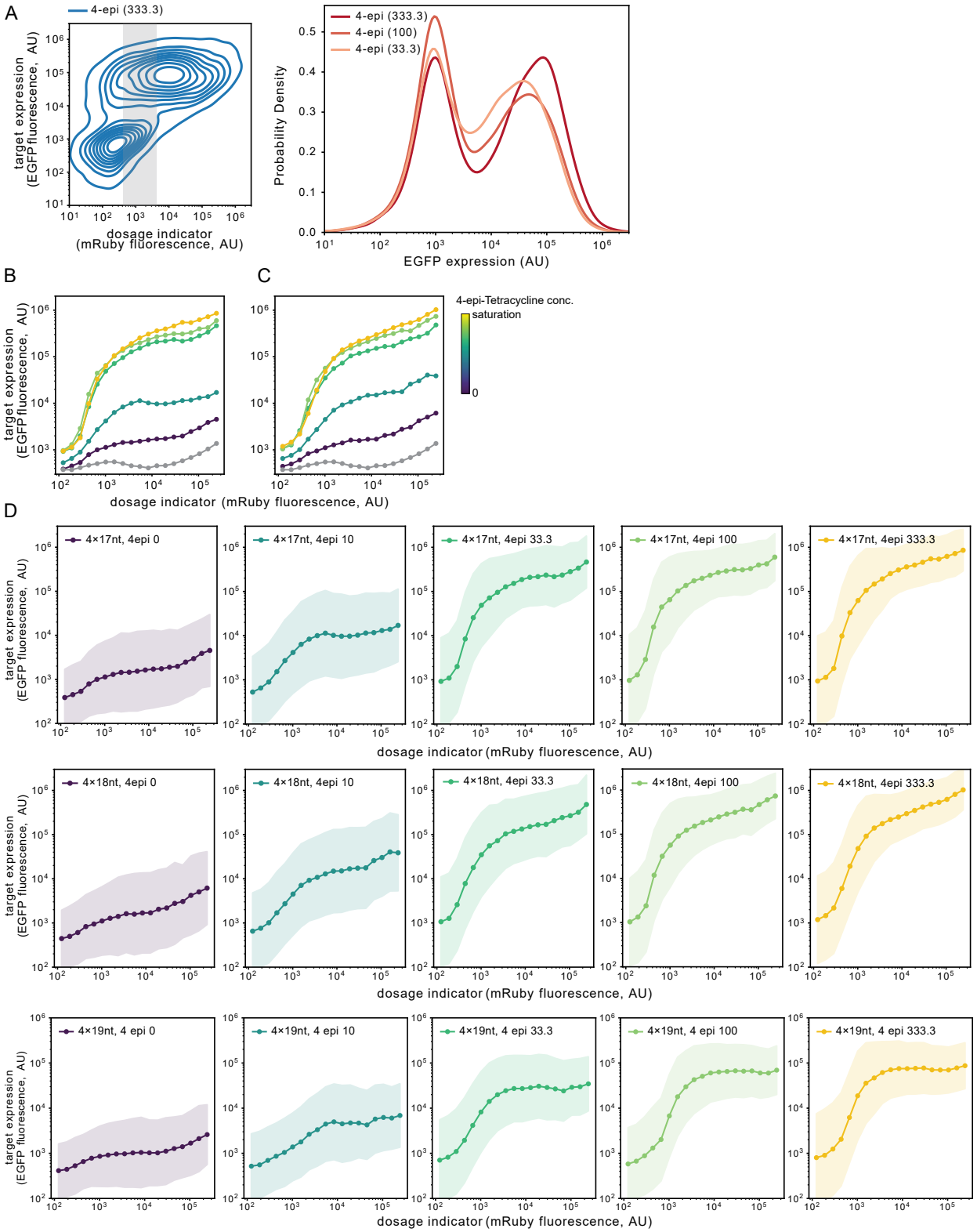

Figure 3.12: (Caption on next page)

Figure 3.12: **The dosage response curves of the inducible DIMMER circuits in TRex cells. (A)** The left panel shows the contour plot of the 4×19nt construct described in **Figure ??**. The concentration of 4-epi was 333.3 ng/mL. Cells falling into the shaded grey region were selected to plot the EGFP distribution in the right panel. Cells in **Figure 3.3E** falling into the same mRuby3 range were also selected to plot the EGFP distribution at the 4-epi concentration of 100 and 33.3 ng/mL. The bimodal distribution of EGFP indicates the bursty nature of the Tet-On promoter used here, and explains the seemingly ultrasensitive behavior observed in **Figure 3.3E** when the dosage is low. **(B-C)** We performed flow cytometry on the TRex cell line which was transfected with the 4×17nt (B) or 4×18nt (C) DIMMER construct described in **Figure 3.3E**. The concentrations of the 4-epi-Tetracycline, from purple to yellow, were 0, 10, 33.3, 100, 333.3 ng/mL. The gray curve denotes the mRuby-only transfection control. **(D)** The curves in **Figure 3.3E** and **3.12B-C** with the geometric variances plotted separately to avoid being cluttered together.

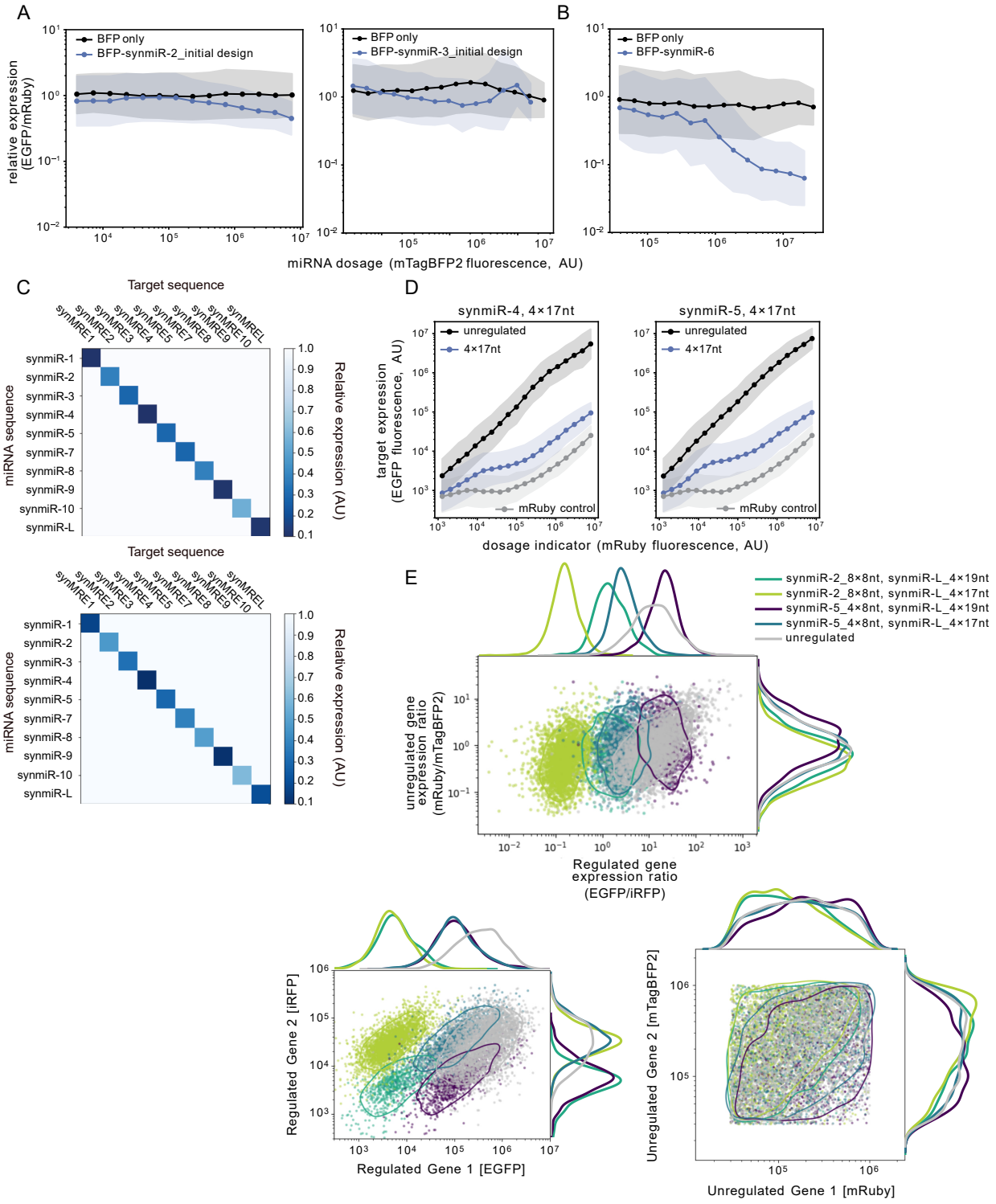

Figure 3.13: (Caption on next page)

Figure 3.13: **The initial synmiR-2, 3, and 6 performances, the biological replicates of the orthogonality test, the 4×17nt designs of synmiR-4 and synmiR-5, and the measurements of the fluorescent protein expression of the dual-reporter system. (A-B)** We performed flow cytometry on the cells co-transfected with the circuit shown in **Figure 3.4A** using the initial sequence of synmiR-2 (A, left panel), synmiR-3 (A, right panel), synmiR-6 (B), and their corresponding targets with a single, fully complementary target site, respectively. The BFP only control does not have the 3'UTR miRNA. Relative expression levels were quantified as described in **Figure 3.4B**. Initial designs of synmiR-2 and synmiR-3 didn't function. synmiR-6 functioned but showed a sequence similarity to the endogenous miRNA. **(C)** The other two biological replicates of the experiment described by **Figure 3.4E**. **(D)** The dosage response curves of 4×17nt designs of synmiR-4 and synmiR-5, measured by flow cytometry. **(E)** The biological replicate of the experiment described in **Figure 3.4H**, along with the unregulated control group (grey). The unregulated group contains the dual reporter system, both without the regulation element. Other experimental settings are the same as described in **Figure 3.4H**.
Upper panel, the distributions of the ratio of both unregulated proteins ([mRuby]/[mTagBFP2]) against the ratio of both regulated proteins ([EGFP]/[iRFP]).
The bottom left panel shows the regulated proteins' distributions of each group. The bottom right panel shows the dosage indicators' distributions of each group.
The dosage indicators exhibit the same distribution among different groups, by contrast, different combinations of DIMMERs allow a clear separation of the populations by the regulated proteins. Additionally, DIMMERs tightly control the stoichiometry of the regulated proteins.

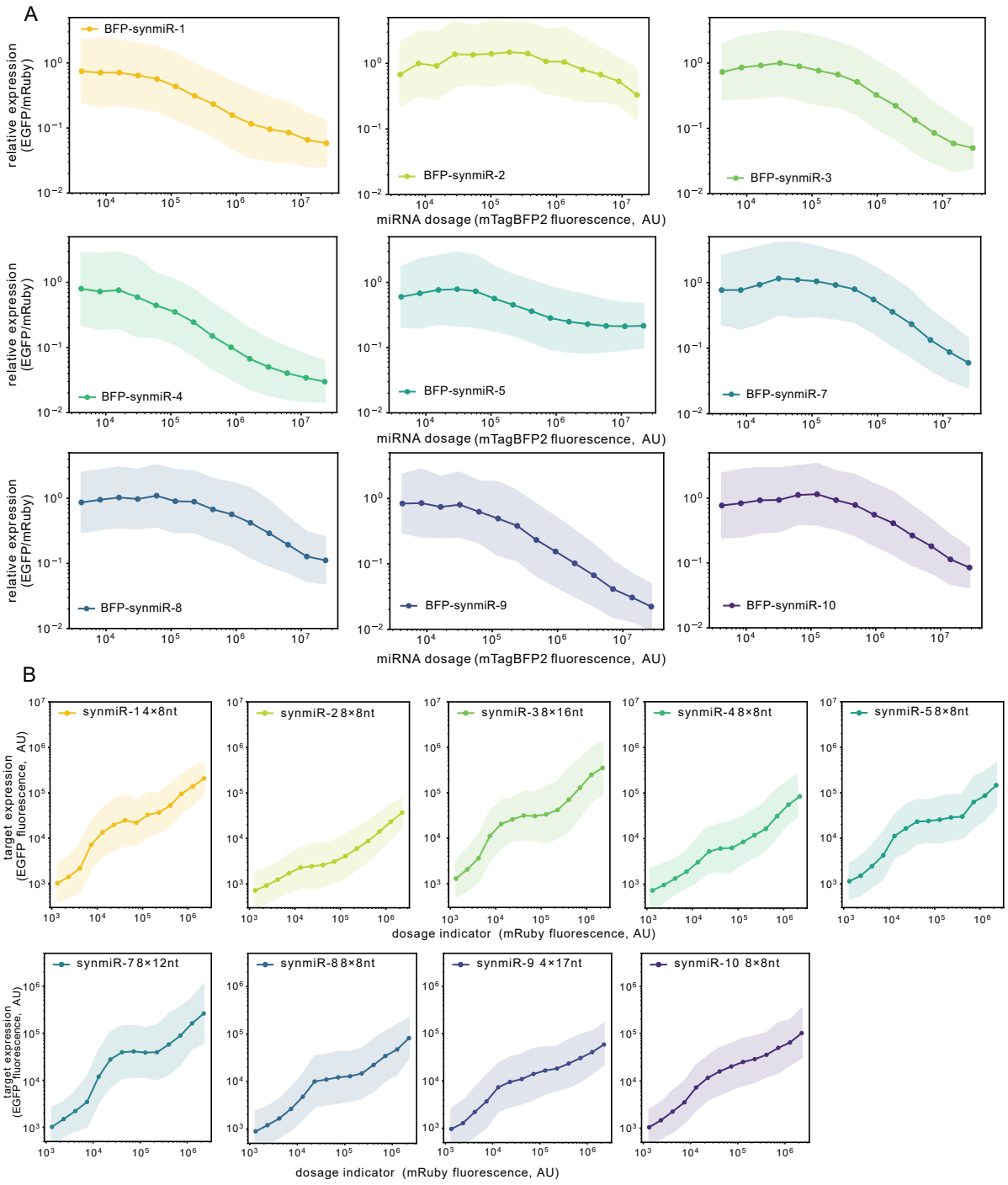

Figure 3.14: **The geometric variance of the dosage response curves in Figure 3.4D (A) and Figure 3.4F (B) to avoid curve cluttering.**

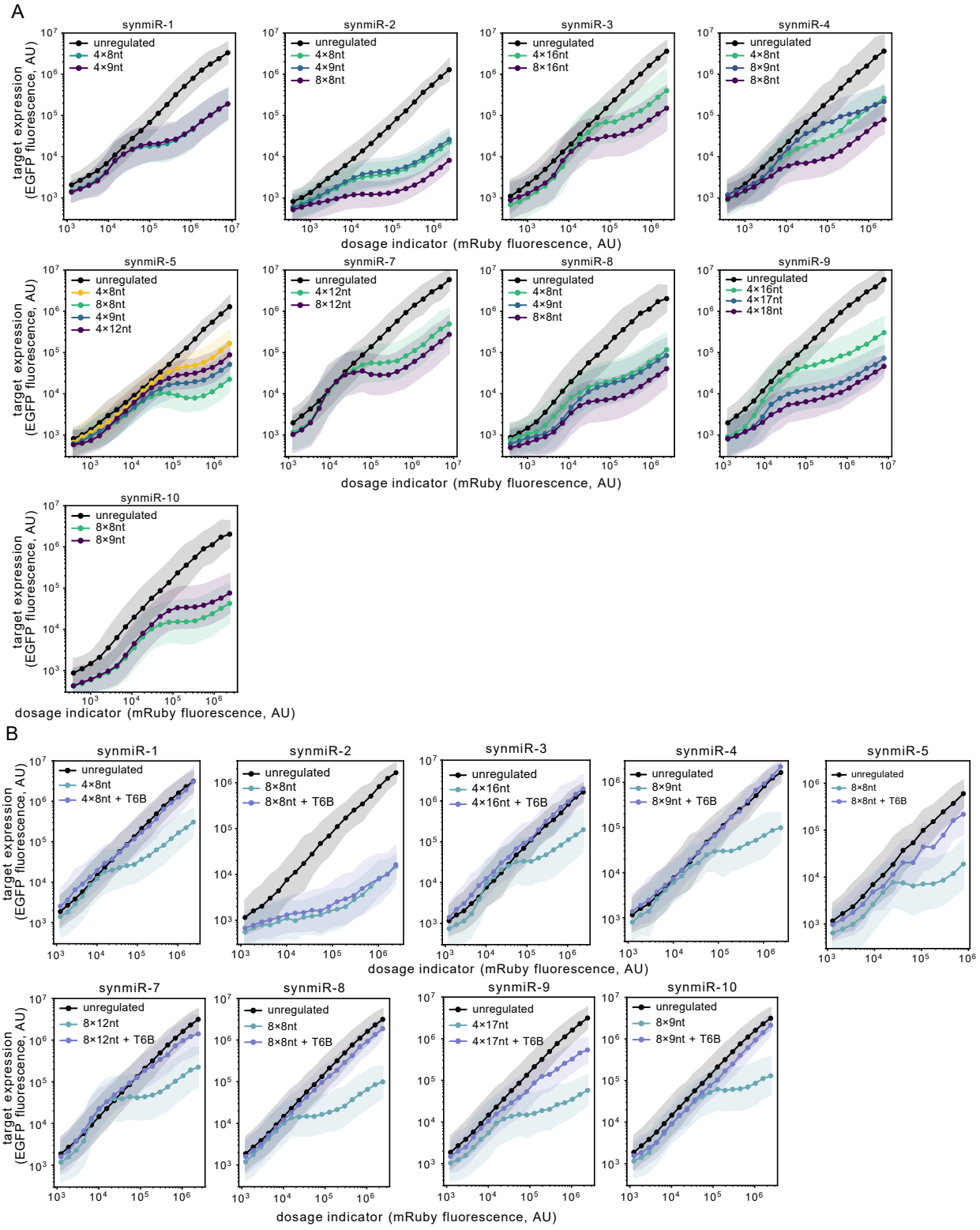

Figure 3.15: **A gallery of all the DIMMER designs based on different miRNAs and targets (A) and the dependence of TNRC6-based regulation (B).** Experimental settings are the same as described in **Figure 3.2B**. Almost all circuits used here relied on the TNRC6 to implement the inhibition, except synmiR-2 8×8nt, which might already be strong enough to achieve the strong regulation.

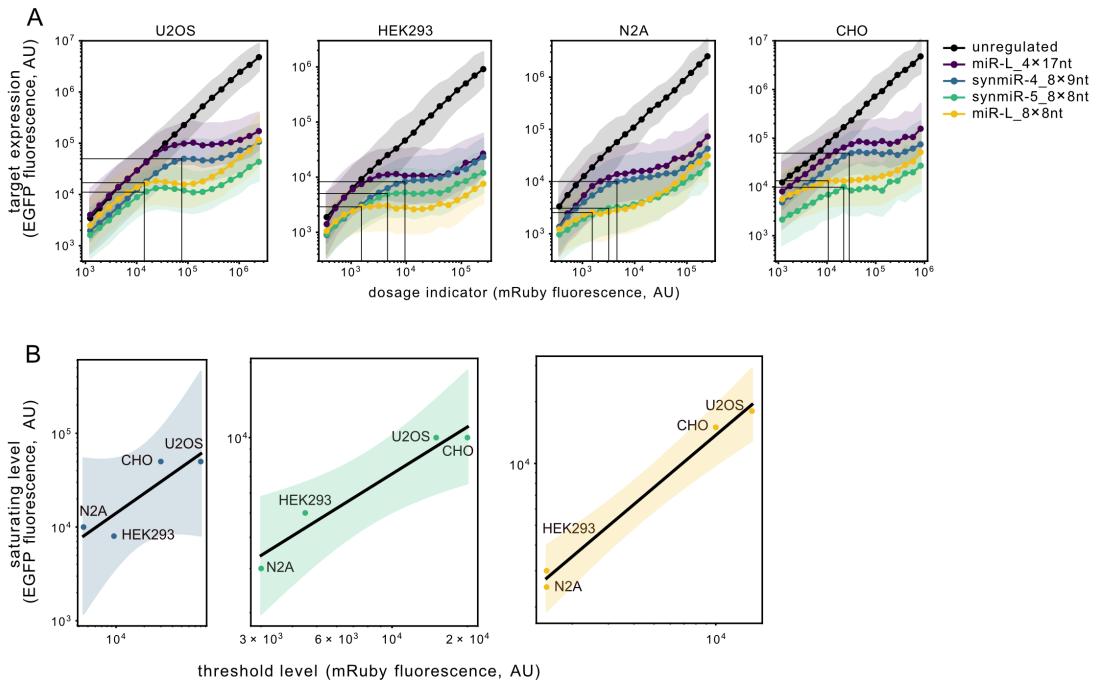

Figure 3.16: **IFFL works across different cell lines.** **(A)** The performance of different DIMMERs among various cell lines. The starting dosage of the dosage compensation behavior and the setpoint expression were indicated by the gray vertical lines and the horizontal lines, respectively. **(B)** The threshold and the saturating level co-vary across cell lines. The black solid line indicates a linear fit in the log-log space (**Methods**). See also **Figure 3.5B**.

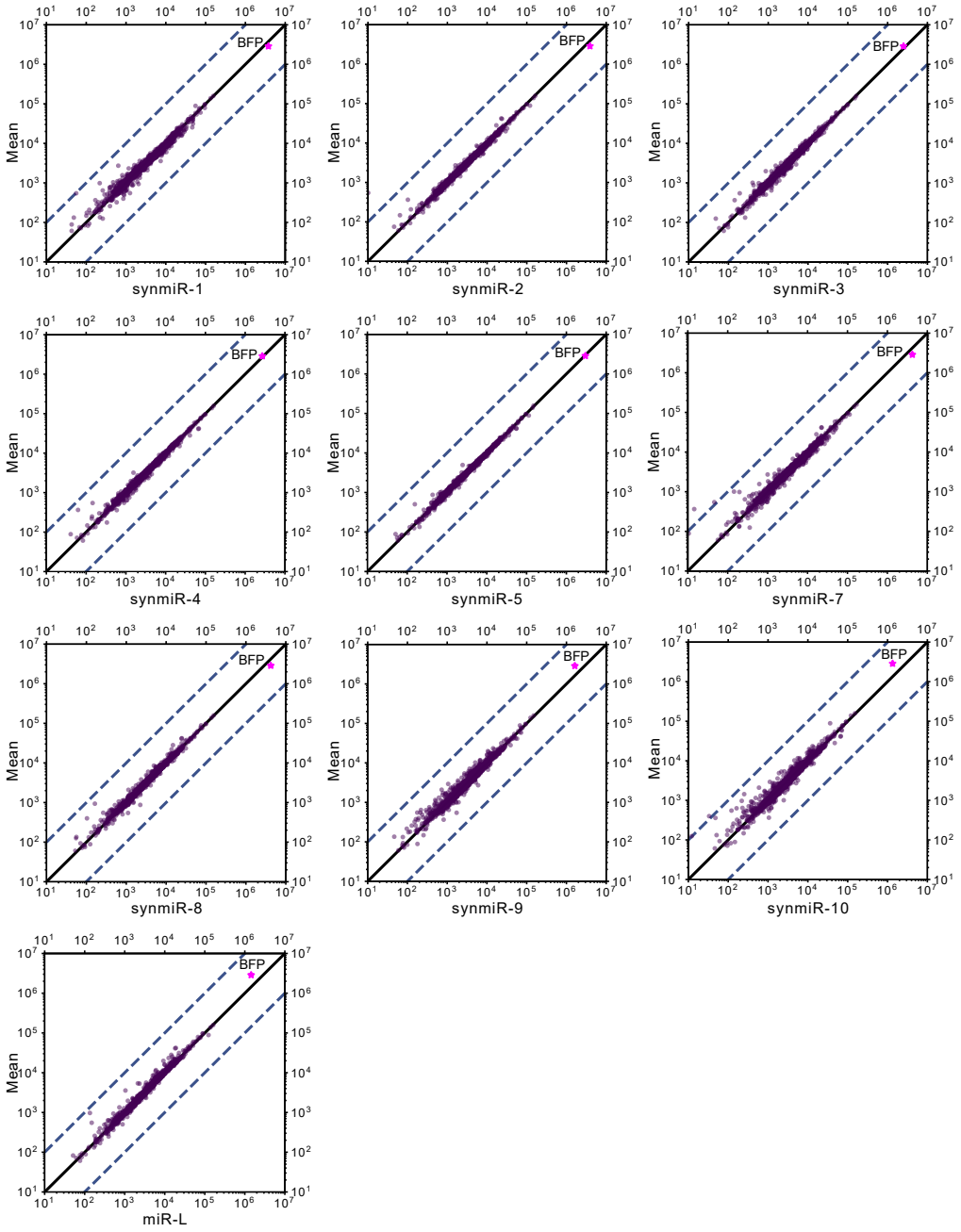

Figure 3.17: **Plots of the normalized transcripts per million (TPM) of the synthetic miRNA expressing cells versus the mean TPM.** The mean TPM was calculated by averaging all the TPM of all the synmiR-expressing cell samples. Solid line indicates where the TPM of the sample is equal to the mean. Dashed lines indicate 10 fold expression differences between the mean and the synmiR-expressing cells.

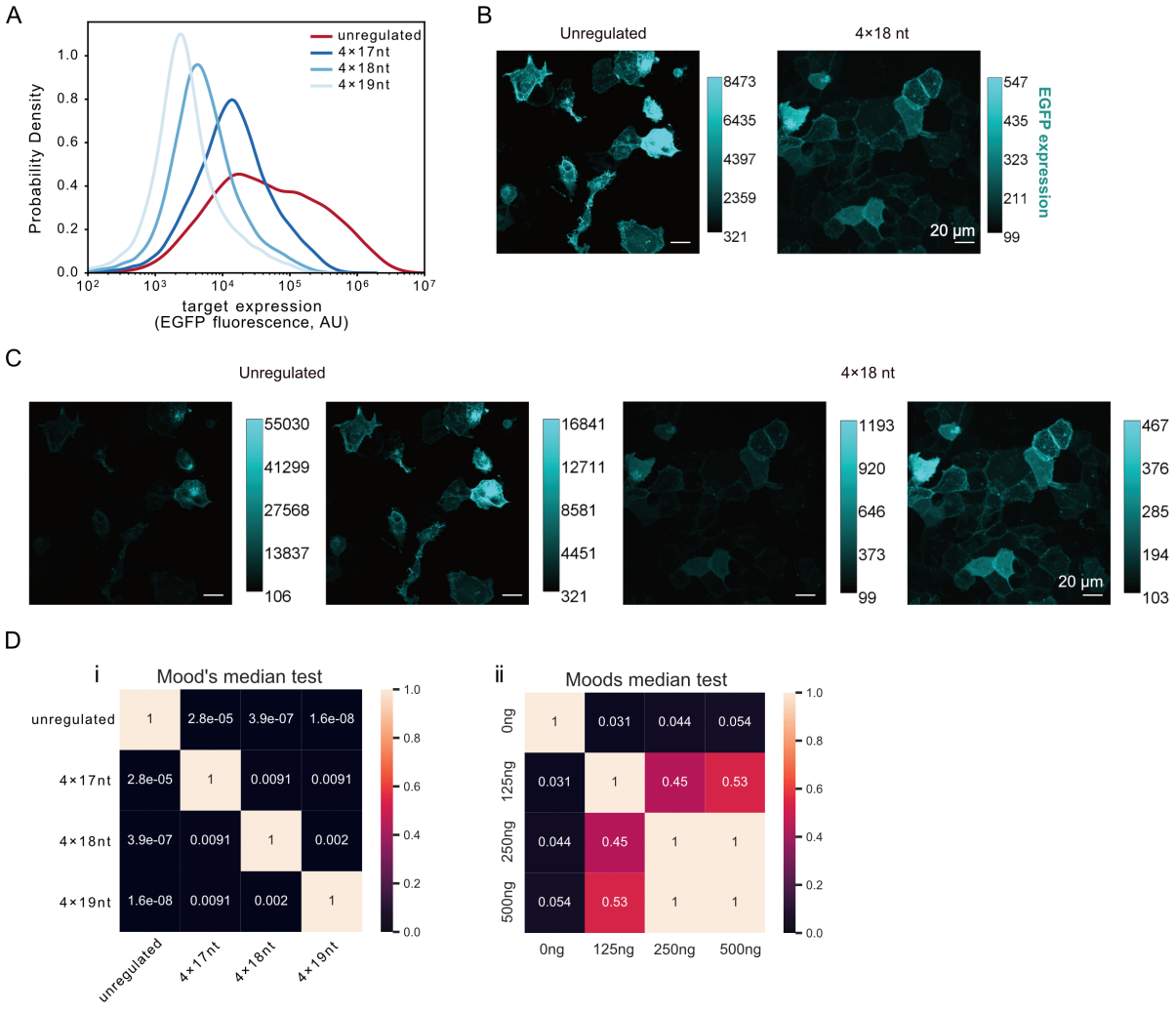

Figure 3.18: **DIMMER circuit improves the DNA-PAINT experiment.** **(A)** We performed flow cytometry on cells that were transfected with the EGFR-mEGFP with or without the DIMMER module. Cells were gated and binned by mRuby3 intensities. The expression of EGFR-mEGFP was plotted. **(B-C)** Representative confocal microscopy images of the U2OS cells transfected with or without the DIMMER circuits (B) with different contrasts (C, indicated by the side colorbars). The images were taken using the 60× magnification objective. Numbers on the color bars indicate the fluorescence intensities measured by imageJ. Scale bar, 20 μm. DIMMERs allow a more uniform expression at a lower setpoint. **(D)** The statistical test of the DNA-PAINT experiment described in **Figure 3.6E** (i) and **Figure 3.6F** (ii). The pipeline of the statistical test is described in the **Methods** section.

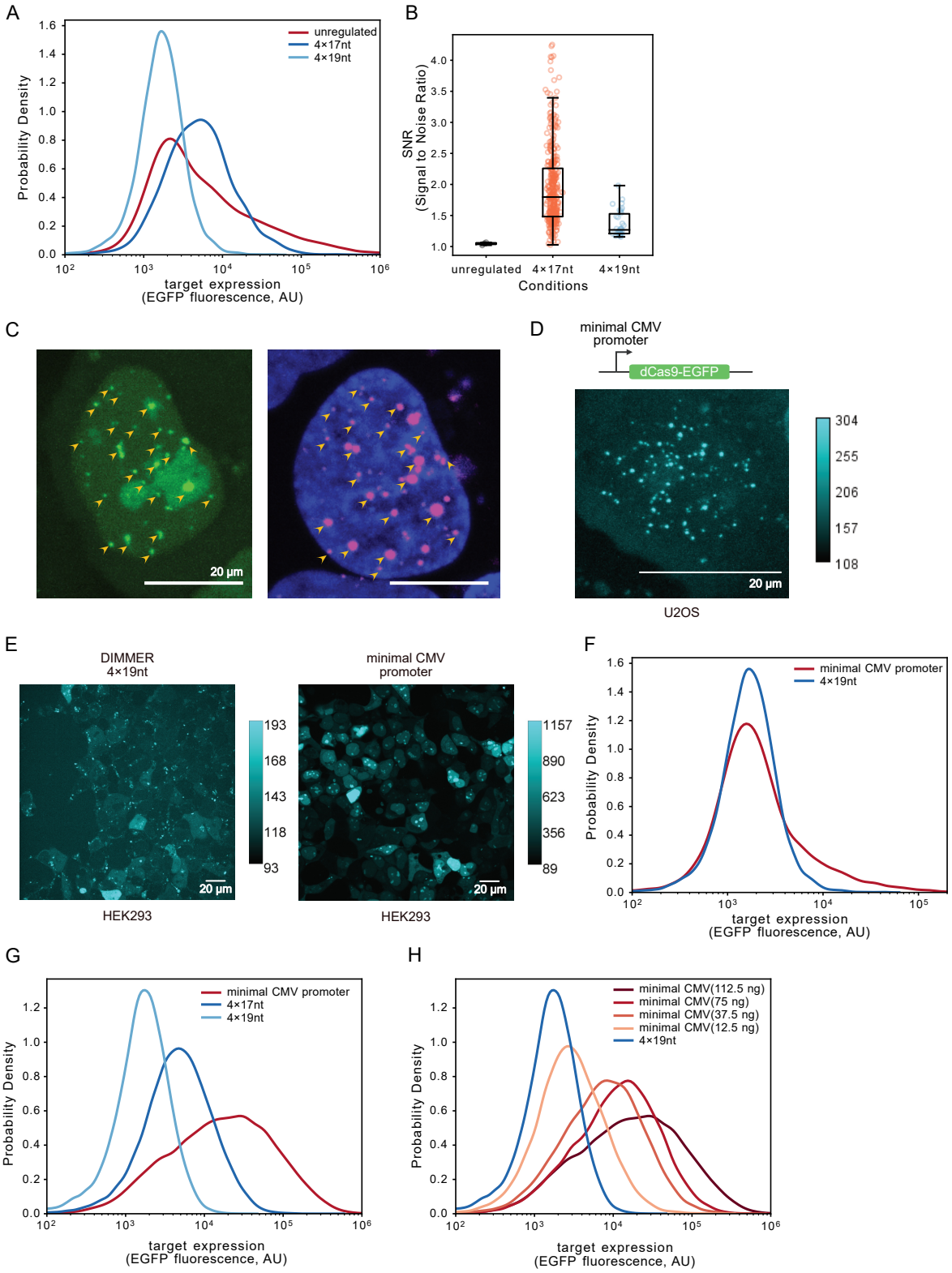

Figure 3.19: (Caption on next page)

Figure 3.19: **DIMMER circuit improves the CRISPR-based imaging.** **(A)** We performed flow cytometry on the cells co-transfected with the dCas9-EGFP and the guide RNA plasmid, with or without the DIMMER circuit regulating the dCas9-EGFP. The plot shows the distributions of the EGFP fluorescence signal. **(B)** The boxplot shows the quantification of the signal-to-noise ratio (SNR) of the dots in the cells transfected with the dCas9-EGFP with or without the DIMMER module. Each scatter represents one dot inside a cell. **(C)** The confocal microscopy images of one live U2OS cell expressing the dCas9-EGFP regulated by the 4×17nt circuit (left) and the same cell labeled by the DNA-FISH (right) probes targeting the telomeres post-fixation. The yellow arrowheads indicate the locations of the corresponding telomeres in the live cell and the fixed cell. Scale bar, 20 μm. Images were taken using the 60× objective. Detailed experimental procedures are described in **Methods**. **(D)** The confocal microscopy image of one live U2OS cell expressing the dCas9-EGFP driven by the minimal CMV promoter. Numbers on the color bars indicate the fluorescence intensities measured by imageJ. Scale bar, 20 μm. **(E)** The confocal microscopy images of live HEK293T cells expressing the dCas9-EGFP either driven by the constitutive promoter and regulated by the 4×19nt DIMMER circuit (left) or driven by the minimal CMV promoter (right). Numbers on the color bars indicate the fluorescence intensities measured by imageJ. Scale bar, 20 μm. **(F)** We performed flow cytometry on the U2OS cells expressing the dCas9-EGFP either driven by the constitutive promoter and regulated by the 4×19nt DIMMER circuit or driven by the minimal CMV promoter. The plot shows the distributions of the EGFP fluorescence signal. **(G)** We performed flow cytometry on the HEK293T cells expressing the dCas9-EGFP either driven by the constitutive promoter and regulated by the 4×17ng and 4×19nt DIMMER circuits or driven by the minimal CMV promoter. The minimal CMV promoter generates a broader EGFP distribution. **(H)** We performed flow cytometry on the HEK293T cells expressing the dCas9-EGFP either driven by the constitutive promoter and regulated by the 4×19nt DIMMER circuit or driven by the minimal CMV promoter with different transfection doses. The lowest transfection amount condition (12.5 ng) still cannot generate a distribution matching that of the 4×19nt DIMMER circuit.

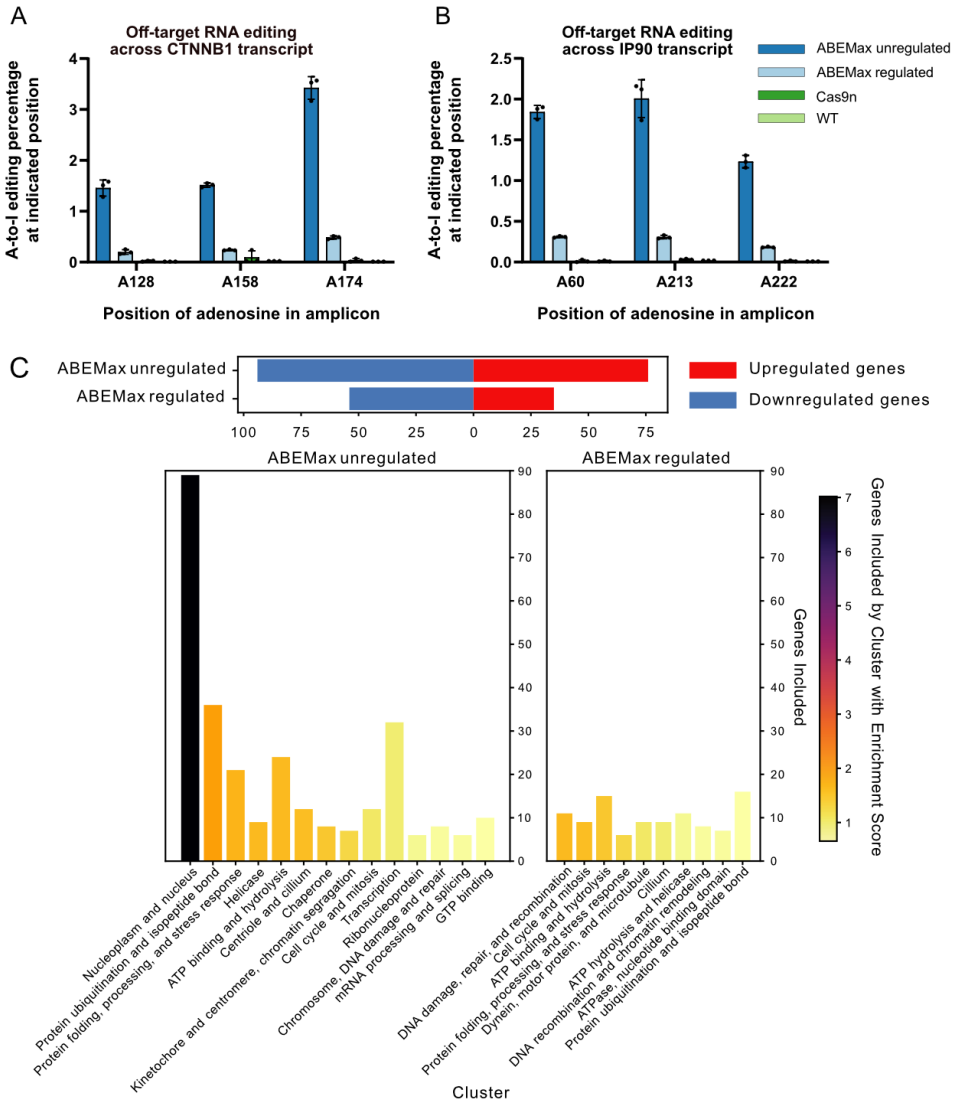

Figure 3.20: **DIMMER circuits reduce the off-target RNA editing of the ABEMax base editor.** The off-target A-to-I RNA editing percentage of the four specific A sites in the CTNNB1 **(A)** and the IP90 **(B)** transcript. Each dot is a biological replicate. Error bars show the standard deviation. **(C)** Gene ontology annotation of significantly affected genes identified by bulk RNA sequencing (as in **Figure 3.7F**)(**Methods**). Upper panel, the number of significantly affected genes in the unregulated ABEMax and 4×18nt circuit-regulated ABEMax groups. Lower panel, gene ontology annotation of significantly upregulated/downregulated genes in the unregulated ABEMax group (left) and the 4×18nt circuit-regulated ABEMax group (right). Color bar denotes enrichment score.

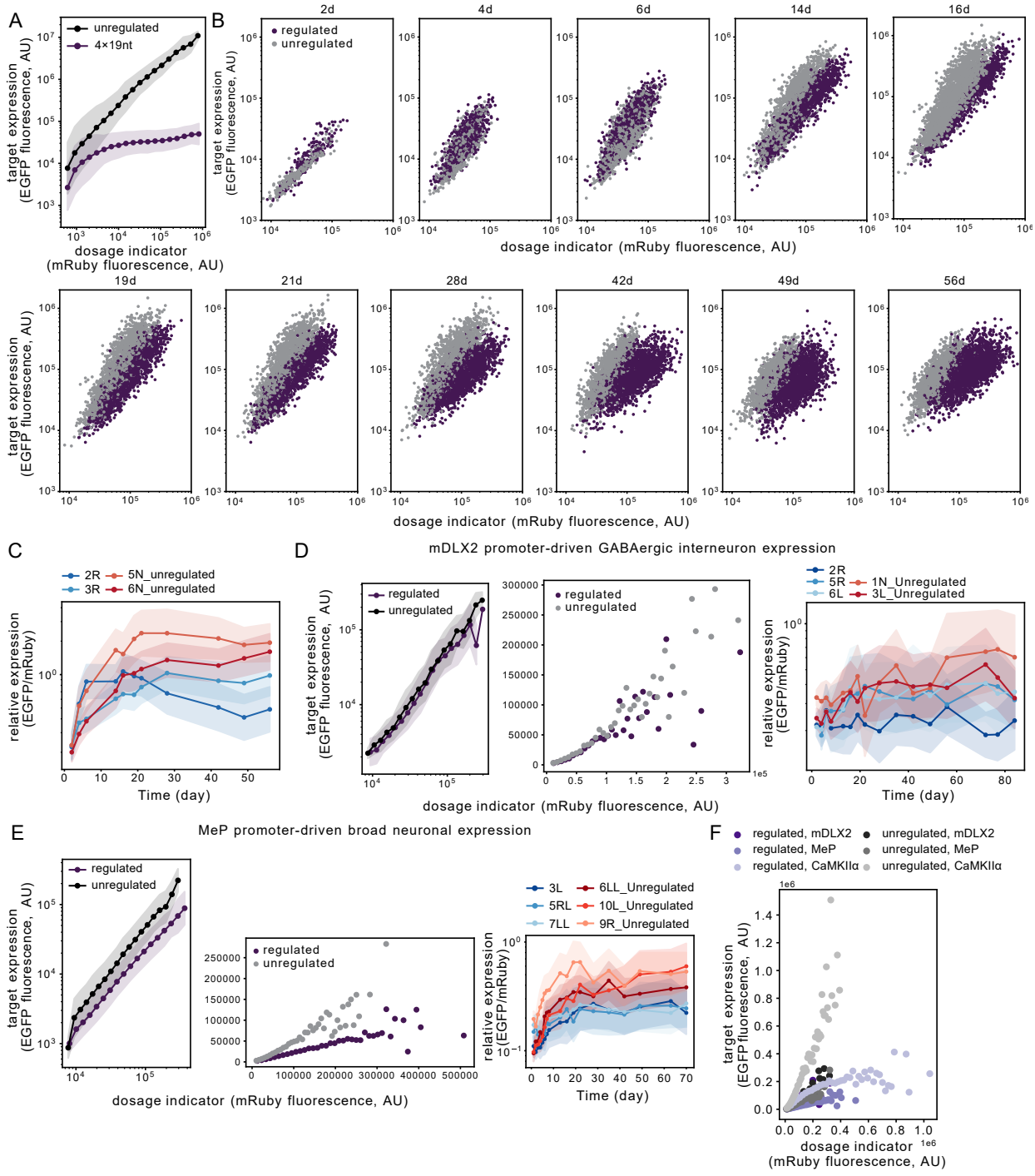

Figure 3.21: (Caption on next page)

Figure 3.21: **DIMMER circuits operate in live mouse brains. (A)** We performed flow cytometry on the cells that were co-transfected with the circuits described in **Figure 3.7G**, upper panel. **(B)** We quantified the expressions of H2B-mRuby3 and H2B-EGFP driven by the CaMKIIα promoter which labels the excitatory neurons over time. Each dot represents the protein expressions of one single cell. Expressions of the unregulated and the regulated groups both accumulated with time, and gradually showed separation from each other. The time shown on the top of each plot indicates the time post injection. The image analysis procedure is described in detail in **Methods**. **(C)** We plotted the relative expression (the median single-cell EGFP/mRuby value) of each mouse at each single time point in the CaMKIIα promoter cohort. The orange curve and the red curve shows the EGFP/mRuby dynamics of two mice in the unregulated group, and the light blue and the dark blue curves show the EGFP/mRuby dynamics of two mice in the 4×19nt DIMMER-regulated group. The shaded region was calculated as described in **Figure 3.1F**. **(D)(E)** We used the mDLX2 promoter to drive the GABAergic interneuronal expression (D), and used the MeP promoter to drive pan-neuronal expression (E) in the mouse brains. The left and the middle panels show the quantification of the pooled cells under each experimental condition, as described in **Figure 3.7H**, with the corresponding tissue-specific promoter, respectively. The right panels show the relative expression of each mouse at each single time point, as described in (C), with the corresponding tissue-specific promoter, respectively. **(F)** We pooled all the data of different cohorts together and plotted the EGFP intensities against mRuby intensities in the linear scale. The dots were calculated by the binned mRuby intensities.

References

[1] M. J. Flynn et al. "Synthetic dosage-compensating miRNA circuits for quantitative gene therapy." *bioRxiv* (Mar. 2024). DOI: 10.1101/2024.03.13.584179. URL: http://dx.doi.org/10.1101/2024.03.13.584179.

[2] K. Franz, A. Singh, and L. S. Weinberger. "Chapter twenty-six - Lentiviral vectors to study stochastic noise in gene expression." *Synthetic Biology, Part A*. Ed. by C. Voigt. Vol. 497. Methods in Enzymology. Academic Press, 2011, pp. 603–622. DOI: https://doi.org/10.1016/B978-0-12-385075-1.00026-3. URL: https://www.sciencedirect.com/science/article/pii/B9780123850751000263.

[3] S. Balasubramanian et al. "Rapid recombinant protein production from piggyBac transposon-mediated stable CHO cell pools." *Journal of Biotechnology* 200 (2015), pp. 61–69. ISSN: 0168-1656. DOI: https://doi.org/10.1016/j.jbiotec.2015.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0168165615000954.

[4] Z. S. Singer et al. "Dynamic heterogeneity and DNA methylation in embryonic stem cells." *Molecular Cell* 55.2 (2014), pp. 319–331. ISSN: 1097-2765. DOI: https://doi.org/10.1016/j.molcel.2014.06.029. URL: https://www.sciencedirect.com/science/article/pii/S1097276514005632.

[5] H. Maamar, A. Raj, and D. Dubnau. "Noise in gene expression determines cell fate in Bacillus subtilis." *Science* 317.5837 (2007), pp. 526–529. DOI: 10.1126/science.1140818. eprint: https://www.science.org/doi/pdf/10.1126/science.1140818. URL: https://www.science.org/doi/abs/10.1126/science.1140818.

[6] U. Alon. "Network motifs: Theory and experimental approaches." *Nature Reviews Genetics* 8 (2007), pp. 450–461. DOI: 10.1038/nrg2102.

[7] L. Bleris et al. "Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template." *Molecular Systems Biology* 7.1 (Jan. 2011). ISSN: 1744-4292. DOI: 10.1038/msb.2011.49. URL: http://dx.doi.org/10.1038/msb.2011.49.

[8] T. J. Strovas et al. "MicroRNA-based single-gene circuits buffer protein synthesis rates against perturbations." en. *ACS Synthetic Biology* 3.5 (May 2014), pp. 324–331.

[9] J. Yang et al. "A synthetic circuit for buffering gene dosage variation between individual mammalian cells." *Nature Communications* 12.1 (July 2021). ISSN: 2041-1723. DOI: 10.1038/s41467-021-23889-0. URL: http://dx.doi.org/10.1038/s41467-021-23889-0.

[10] K. S. Love et al. "Model-guided design of microRNA-based gene circuits supports precise dosage of transgenic cargoes into diverse primary cells." *bioRxiv* (June 2024). DOI: 10.1101/2024.06.25.600629. URL: http://dx.doi.org/10.1101/2024.06.25.600629.

[11] J. Sheu-Gruttadauria and I. J. MacRae. "Phase transitions in the assembly and function of human miRISC." *Cell* 173.4 (May 2018), 946–957.e16. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.02.051. URL: http://dx.doi.org/10.1016/j.cell.2018.02.051.

[12] D. P. Bartel. "Metazoan microRNAs." *Cell* 173.1 (Mar. 2018), pp. 20–51. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.03.006. URL: http://dx.doi.org/10.1016/j.cell.2018.03.006.

[13] C. Fellmann et al. "An optimized microRNA backbone for effective single-copy RNAi." *Cell Reports* 5.6 (Dec. 2013), pp. 1704–1713. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2013.11.020. URL: http://dx.doi.org/10.1016/j.celrep.2013.11.020.

[14] L. Qiu et al. "A construct with fluorescent indicators for conditional expression of miRNA." *BMC Biotechnology* 8.1 (Oct. 2008). ISSN: 1472-6750. DOI: 10.1186/1472-6750-8-77. URL: http://dx.doi.org/10.1186/1472-6750-8-77.

[15] E. Elkayam et al. "Multivalent recruitment of human argonaute by GW182." *Molecular Cell* 67.4 (Aug. 2017), 646–658.e3. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2017.07.007. URL: http://dx.doi.org/10.1016/j.molcel.2017.07.007.

[16] G. La Rocca et al. "Inducible and reversible inhibition of miRNA-mediated gene repression in vivo." *eLife* 10 (Aug. 2021). ISSN: 2050-084X. DOI: 10.7554/elife.70948. URL: http://dx.doi.org/10.7554/eLife.70948.

[17] C. Shin et al. "Expanding the microRNA targeting code: Functional sites with centered pairing." *Molecular Cell* 38.6 (June 2010), pp. 789–802. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2010.06.005. URL: http://dx.doi.org/10.1016/j.molcel.2010.06.005.

[18] D. Jee et al. "Dual strategies for Argonaute2-mediated biogenesis of erythroid miRNAs underlie conserved requirements for slicing in mammals." *Molecular Cell* 69.2 (Jan. 2018), 265–278.e6. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2017.12.027. URL: http://dx.doi.org/10.1016/j.molcel.2017.12.027.

[19] P. Y. Wang and D. P. Bartel. "The guide-RNA sequence dictates the slicing kinetics and conformational dynamics of the Argonaute silencing complex." *Molecular Cell* 84.15 (Aug. 2024), 2918–2934.e11. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2024.06.026. URL: http://dx.doi.org/10.1016/j.molcel.2024.06.026.

[20] S. E. McGeary et al. "MicroRNA 3'-compensatory pairing occurs through two binding modes, with affinity shaped by nucleotide identity and position." *eLife* 11 (Feb. 2022). Ed. by T. W. Nilsen et al., e69803. ISSN: 2050-084X. DOI: 10.7554/eLife.69803. URL: https://doi.org/10.7554/eLife.69803.

[21] W. R. Becker et al. "High-throughput analysis reveals rules for target RNA binding and cleavage by AGO2." *Molecular Cell* 75.4 (Aug. 2019), 741–755.e11. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2019.06.012. URL: http://dx.doi.org/10.1016/j.molcel.2019.06.012.

[22] Y. S. Michaels et al. "Precise tuning of gene expression levels in mammalian cells." *Nature Communications* 10.1 (Feb. 2019). ISSN: 2041-1723. DOI: 10.1038/s41467-019-08777-y. URL: http://dx.doi.org/10.1038/s41467-019-08777-y.

[23] S. L. Ameres et al. "Target RNA-directed trimming and tailing of small silencing RNAs." *Science* 328.5985 (June 2010), pp. 1534–1539. ISSN: 1095-9203. DOI: 10.1126/science.1187058. URL: http://dx.doi.org/10.1126/science.1187058.

[24] M. Ha and V. N. Kim. "Regulation of microRNA biogenesis." *Nature Reviews Molecular Cell Biology* 15.8 (July 2014), pp. 509–524. ISSN: 1471-0080. DOI: 10.1038/nrm3838. URL: http://dx.doi.org/10.1038/nrm3838.

[25] S. E. McGeary et al. "The biochemical basis of microRNA targeting efficacy." *Science* 366.6472 (Dec. 2019). ISSN: 1095-9203. DOI: 10.1126/science.aav1741. URL: http://dx.doi.org/10.1126/science.aav1741.

[26] J. Ponten and E. Saksela. "Two established in vitro cell lines from human mesenchymal tumours." *International Journal of Cancer* 2.5 (Sept. 1967), pp. 434–447. ISSN: 1097-0215. DOI: 10.1002/ijc.2910020505. URL: http://dx.doi.org/10.1002/ijc.2910020505.

[27] T. T. Puck, S. J. Cieciura, and A. Robinson. "Genetics of somatic mammalian cells." *The Journal of Experimental Medicine* 108.6 (Dec. 1958), pp. 945–956. ISSN: 0022-1007. DOI: 10.1084/jem.108.6.945. URL: http://dx.doi.org/10.1084/jem.108.6.945.

[28] F. L. Graham et al. "Characteristics of a human cell line transformed by DNA from human adenovirus type 5." *Journal of General Virology* 36.1 (July 1977), pp. 59–72. ISSN: 1465-2099. DOI: 10.1099/0022-1317-36-1-59. URL: http://dx.doi.org/10.1099/0022-1317-36-1-59.

[29] J. B. Olmsted et al. "Isolation of microtubule protein from cultured mouse neuroblastoma cells." *Proceedings of the National Academy of Sciences* 65.1 (Jan. 1970), pp. 129–136. ISSN: 1091-6490. DOI: 10.1073/pnas.65.1.129. URL: http://dx.doi.org/10.1073/pnas.65.1.129.

[30] J. M. Marchingo and D. A. Cantrell. "Protein synthesis, degradation, and energy metabolism in T cell immunity." *Cellular and Molecular Immunology* 19.3 (Jan. 2022), pp. 303–315. ISSN: 2042-0226. DOI: 10.1038/s41423-021-00792-8. URL: http://dx.doi.org/10.1038/s41423-021-00792-8.

[31] E. Lakatos et al. "Protein degradation rate is the dominant mechanism accounting for the differences in protein abundance of basal p53 in a human breast and colorectal cancer cell line." *PLoS One* 12.5 (May 2017). Ed. by S. Deb, e0177336. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0177336. URL: http://dx.doi.org/10.1371/journal.pone.0177336.

[32] S. Liu et al. "Oversized cells activate global proteasome-mediated protein degradation to maintain cell size homeostasis." *eLife* 14 (Jan. 2025). ISSN: 2050-084X. DOI: 10.7554/elife.75393. URL: http://dx.doi.org/10.7554/eLife.75393.

[33] K. Swovick et al. "Interspecies differences in proteome turnover kinetics are correlated with life spans and energetic demands." *Molecular and Cellular Proteomics* 20 (2021), p. 100041. ISSN: 1535-9476. DOI: 10.1074/mcp.ra120.002301. URL: http://dx.doi.org/10.1074/mcp.RA120.002301.

[34] T. Kumamoto et al. "Direct readout of neural stem cell transgenesis with an integration-coupled gene expression switch." *Neuron* 107.4 (Aug. 2020), 617–630.e6. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2020.05.038. URL: http://dx.doi.org/10.1016/j.neuron.2020.05.038.

[35] J. Schnitzbauer et al. "Super-resolution microscopy with DNA-PAINT." *Nature Protocols* 12.6 (May 2017), pp. 1198–1228. ISSN: 1750-2799. DOI: 10.1038/nprot.2017.024. URL: http://dx.doi.org/10.1038/nprot.2017.024.

[36] M. Jo et al. "Epidermal growth factor receptor-dependent and -independent cell-signaling pathways originating from the urokinase receptor." *Journal of Biological Chemistry* 278.3 (Jan. 2003), pp. 1642–1646. ISSN: 0021-9258. DOI: 10.1074/jbc.m210877200. URL: http://dx.doi.org/10.1074/jbc.M210877200.

[37] B. Andrews et al. "Imaging cell biology." *Nature Cell Biology* 24.8 (July 2022), pp. 1180–1185. ISSN: 1476-4679. DOI: 10.1038/s41556-022-00960-6. URL: http://dx.doi.org/10.1038/s41556-022-00960-6.

[38] B. Chen et al. "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system." *Cell* 155.7 (Dec. 2013), pp. 1479–1491. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.12.001. URL: http://dx.doi.org/10.1016/j.cell.2013.12.001.

[39] H. Ma et al. "Multicolor CRISPR labeling of chromosomal loci in human cells." *Proceedings of the National Academy of Sciences* 112.10 (Feb. 2015), pp. 3002–3007. ISSN: 1091-6490. DOI: 10.1073/pnas.1420024112. URL: http://dx.doi.org/10.1073/pnas.1420024112.

[40] H. Ma et al. "Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow." *Nature Biotechnology* 34.5 (Apr. 2016), pp. 528–530. ISSN: 1546-1696. DOI: 10.1038/nbt.3526. URL: http://dx.doi.org/10.1038/nbt.3526.

[41] H. Wang et al. "CRISPR-mediated live imaging of genome editing and transcription." *Science* 365.6459 (Sept. 2019), pp. 1301–1305. ISSN: 1095-9203. DOI: 10.1126/science.aax7852. URL: http://dx.doi.org/10.1126/science.aax7852.

[42] J. Grünewald et al. "Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors." *Nature* 569.7756 (Apr. 2019), pp. 433–437. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1161-z. URL: http://dx.doi.org/10.1038/s41586-019-1161-z.

[43] M. Fiumara et al. "Genotoxic effects of base and prime editing in human hematopoietic stem cells." *Nature Biotechnology* 42.6 (Sept. 2023), pp. 877–891. ISSN: 1546-1696. DOI: 10.1038/s41587-023-01915-4. URL: http://dx.doi.org/10.1038/s41587-023-01915-4.

[44] H. A. Rees et al. "Analysis and minimization of cellular RNA editing by DNA adenine base editors." *Science Advances* 5.5 (May 2019). ISSN: 2375-2548. DOI: 10.1126/sciadv.aax5717. URL: http://dx.doi.org/10.1126/sciadv.aax5717.

[45] J. Li et al. "Structure-guided engineering of adenine base editor with minimized RNA off-targeting activity." *Nature Communications* 12.1 (Apr. 2021). ISSN: 2041-1723. DOI: 10.1038/s41467-021-22519-z. URL: http://dx.doi.org/10.1038/s41467-021-22519-z.

[46] C. Zhou et al. "Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis." *Nature* 571.7764 (June 2019), pp. 275–278. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1314-0. URL: http://dx.doi.org/10.1038/s41586-019-1314-0.

[47] M. E. Neugebauer et al. "Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity." *Nature Biotechnology* 41.5 (Nov. 2022), pp. 673–685. ISSN: 1546-1696. DOI: 10.1038/s41587-022-01533-6. URL: http://dx.doi.org/10.1038/s41587-022-01533-6.

[48] L. Zhou et al. "Advances in AAV-mediated gene replacement therapy for pediatric monogenic neurological disorders." *Molecular Therapy - Methods and Clinical Development* 32.4 (Dec. 2024), p. 101357. ISSN: 2329-0501. DOI: 10.1016/j.omtm.2024.101357. URL: http://dx.doi.org/10.1016/j.omtm.2024.101357.

[49] J. Gao et al. "Gene therapy for CNS disorders: Modalities, delivery and translational challenges." *Nature Reviews Neuroscience* 25.8 (June 2024), pp. 553–572. ISSN: 1471-0048. DOI: 10.1038/s41583-024-00829-7. URL: http://dx.doi.org/10.1038/s41583-024-00829-7.

[50] M. R. Chuapoco et al. "Adeno-associated viral vectors for functional intravenous gene transfer throughout the non-human primate brain." *Nature Nanotechnology* 18.10 (July 2023), pp. 1241–1251. ISSN: 1748-3395. DOI: 10.1038/s41565-023-01419-x. URL: http://dx.doi.org/10.1038/s41565-023-01419-x.

[51] D. Goertsen et al. "AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset." *Nature Neuroscience* 25.1 (Dec. 2021), pp. 106–115. ISSN: 1546-1726. DOI: 10.1038/s41593-021-00969-4. URL: http://dx.doi.org/10.1038/s41593-021-00969-4.

[52] Q. Huang et al. "An AAV capsid reprogrammed to bind human transferrin receptor mediates brain-wide gene delivery." *Science* 384.6701 (June 2024), pp. 1220–1227. ISSN: 1095-9203. DOI: 10.1126/science.adm8386. URL: http://dx.doi.org/10.1126/science.adm8386.

[53] X. B. Liu and E. G. Jones. "Localization of alpha type II calcium calmodulin-dependent protein kinase at glutamatergic but not gamma-aminobutyric acid (GABAergic) synapses in thalamus and cerebral cortex." *Proceedings of the National Academy of Sciences* 93.14 (July 1996), pp. 7332–7336. ISSN: 1091-6490. DOI: 10.1073/pnas.93.14.7332. URL: http://dx.doi.org/10.1073/pnas.93.14.7332.

[54] A. Sík et al. "The absence of a major Ca 2+ signaling pathway in GABAergic neurons of the hippocampus." *Proceedings of the National Academy of Sciences* 95.6 (Mar. 1998), pp. 3245–3250. ISSN: 1091-6490. DOI: 10.1073/pnas.95.6.3245. URL: http://dx.doi.org/10.1073/pnas.95.6.3245.

[55] M. Rastegar et al. "MECP2 isoform-specific vectors with regulated expression for Rett syndrome gene therapy." *PLoS ONE* 4.8 (Aug. 2009). Ed. by R. Linden, e6810. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0006810. URL: http://dx.doi.org/10.1371/journal.pone.0006810.

[56] J. Dimidschstein et al. “A viral strategy for targeting and manipulating interneurons across vertebrate species.” *Nature Neuroscience* 19.12 (Oct. 2016), pp. 1743–1749. ISSN: 1546-1726. DOI: 10.1038/nn.4430. URL: http://dx.doi.org/10.1038/nn.4430.

[57] J. R. Cho et al. “Dorsal raphe dopamine neurons signal motivational salience dependent on internal state, expectation, and behavioral Context.” *The Journal of Neuroscience* 41.12 (Feb. 2021), pp. 2645–2655. ISSN: 1529-2401. DOI: 10.1523/jneurosci.2690-20.2021. URL: http://dx.doi.org/10.1523/JNEUROSCI.2690-20.2021.

[58] J. Cursons et al. “Combinatorial targeting by microRNAs co-ordinates post-transcriptional control of EMT.” *Cell Systems* 7.1 (July 2018), 77–91.e7. ISSN: 2405-4712. DOI: 10.1016/j.cels.2018.05.019. URL: http://dx.doi.org/10.1016/j.cels.2018.05.019.

[59] J. M. Raser and E. K. O’Shea. “Noise in gene expression: Origins, consequences, and control.” *Science* 309.5743 (Sept. 2005), pp. 2010–2013. ISSN: 1095-9203. DOI: 10.1126/science.1105891. URL: http://dx.doi.org/10.1126/science.1105891.

[60] J. M. Pedraza and A. van Oudenaarden. “Noise propagation in gene networks.” *Science* 307.5717 (Mar. 2005), pp. 1965–1969. ISSN: 1095-9203. DOI: 10.1126/science.1109090. URL: http://dx.doi.org/10.1126/science.1109090.

[61] M. Van Alstyne et al. “Gain of toxic function by long-term AAV9-mediated SMN overexpression in the sensorimotor circuit.” *Nature Neuroscience* 24.7 (Apr. 2021), pp. 930–940. ISSN: 1546-1726. DOI: 10.1038/s41593-021-00827-3. URL: http://dx.doi.org/10.1038/s41593-021-00827-3.

[62] C. Hinderer et al. “Severe toxicity in nonhuman primates and piglets following high-Dose intravenous administration of an adeno-associated virus vector expressing human SMN.” *Human Gene Therapy* 29.3 (Mar. 2018), pp. 285–298. ISSN: 1557-7422. DOI: 10.1089/hum.2018.015. URL: http://dx.doi.org/10.1089/hum.2018.015.

[63] S. E. P. Smith et al. “Increased gene dosage of Ube3a results in autism traits and decreased glutamate synaptic transmission in mice.” *Science Translational Medicine* 3.103 (Oct. 2011). ISSN: 1946-6242. DOI: 10.1126/scitranslmed.3002627. URL: http://dx.doi.org/10.1126/scitranslmed.3002627.

[64] A. L. Collins et al. “Mild overexpression of MeCP2 causes a progressive neurological disorder in mice.” *Human Molecular Genetics* 13.21 (Sept. 2004), pp. 2679–2689. ISSN: 0964-6906. DOI: 10.1093/hmg/ddh282. URL: http://dx.doi.org/10.1093/hmg/ddh282.

[65] V. Matagne et al. “Severe offtarget effects following intravenous delivery of AAV9-MECP2 in a female mouse model of Rett syndrome.” *Neurobiology of Disease* 149 (Feb. 2021), p. 105235. ISSN: 0969-9961. DOI: 10.1016/j.nbd.2020.105235. URL: http://dx.doi.org/10.1016/j.nbd.2020.105235.

[66] C. J. Bashor et al. “Engineering the next generation of cell-based therapeutics.” *Nature Reviews Drug Discovery* 21.9 (May 2022), pp. 655–675. ISSN: 1474-1784. DOI: 10.1038/s41573-022-00476-6. URL: http://dx.doi.org/10.1038/s41573-022-00476-6.

[67] Y.-H. Loh et al. "Reprogramming of T-cells from human peripheral blood." *Cell Stem Cell* 7.1 (July 2010), pp. 15–19. ISSN: 1934-5909. DOI: 10.1016/j.stem.2010.06.004. URL: http://dx.doi.org/10.1016/j.stem.2010.06.004.

[68] H. Wang et al. "Direct cell reprogramming: Approaches, mechanisms and progress." *Nature Reviews Molecular Cell Biology* 22.6 (Feb. 2021), pp. 410–424. ISSN: 1471-0080. DOI: 10.1038/s41580-021-00335-z. URL: http://dx.doi.org/10.1038/s41580-021-00335-z.

[69] Z. Liu et al. "Expression of TNRC6 (GW182) proteins is not necessary for gene silencing by fully complementary RNA duplexes." *Nucleic Acid Therapeutics* 29.6 (Dec. 2019), pp. 323–334. ISSN: 2159-3345. DOI: 10.1089/nat.2019.0815. URL: http://dx.doi.org/10.1089/nat.2019.0815.

[70] A. Kilikevicius, G. Meister, and D. R. Corey. "Reexamining assumptions about miRNA-guided gene silencing." *Nucleic Acids Research* 50.2 (Dec. 2021), pp. 617–634. ISSN: 1362-4962. DOI: 10.1093/nar/gkab1256. URL: http://dx.doi.org/10.1093/nar/gkab1256.

[71] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones. "miRBase: From microRNA sequences to function." *Nucleic Acids Research* 47.D1 (Nov. 2018), pp. D155–D162. ISSN: 1362-4962. DOI: 10.1093/nar/gky1141. URL: http://dx.doi.org/10.1093/nar/gky1141.

[72] A. Kozomara and S. Griffiths-Jones. "miRBase: Annotating high confidence microRNAs using deep sequencing data." *Nucleic Acids Research* 42.D1 (Nov. 2013), pp. D68–D73. ISSN: 1362-4962. DOI: 10.1093/nar/gkt1181. URL: http://dx.doi.org/10.1093/nar/gkt1181.

[73] A. Kozomara and S. Griffiths-Jones. "miRBase: Integrating microRNA annotation and deep-sequencing data." *Nucleic Acids Research* 39.Database (Oct. 2010), pp. D152–D157. ISSN: 1362-4962. DOI: 10.1093/nar/gkq1027. URL: http://dx.doi.org/10.1093/nar/gkq1027.

[74] S. Griffiths-Jones et al. "miRBase: Tools for microRNA genomics." *Nucleic Acids Research* 36.Database (Dec. 2007), pp. D154–D158. ISSN: 1362-4962. DOI: 10.1093/nar/gkm952. URL: http://dx.doi.org/10.1093/nar/gkm952.

[75] S. Griffiths-Jones. "miRBase: MicroRNA sequences, targets and gene nomenclature." *Nucleic Acids Research* 34.90001 (Jan. 2006), pp. D140–D144. ISSN: 1362-4962. DOI: 10.1093/nar/gkj112. URL: http://dx.doi.org/10.1093/nar/gkj112.

[76] S. Griffiths-Jones. "The microRNA registry." *Nucleic Acids Research* 32.90001 (Jan. 2004), pp. 109D–111. ISSN: 1362-4962. DOI: 10.1093/nar/gkh023. URL: http://dx.doi.org/10.1093/nar/gkh023.

[77] V. Ambros et al. "A uniform system for microRNA annotation." *RNA* 9.3 (Mar. 2003), pp. 277–279. ISSN: 1469-9001. DOI: 10.1261/rna.2183803. URL: http://dx.doi.org/10.1261/rna.2183803.

[78] B. C. Meyers et al. "Criteria for annotation of plant microRNAs." *The Plant Cell* 20.12 (Dec. 2008), pp. 3186–3190. ISSN: 1532-298X. DOI: 10.1105/tpc.108.064311. URL: http://dx.doi.org/10.1105/tpc.108.064311.

[79] J. Schindelin et al. "Fiji: an open-source platform for biological-image analysis." *Nature Methods* 9.7 (June 2012), pp. 676–682. ISSN: 1548-7105. DOI: 10.1038/nmeth.2019. URL: http://dx.doi.org/10.1038/nmeth.2019.

[80] N. L. Bray et al. "Near-optimal probabilistic RNA-seq quantification." *Nature Biotechnology* 34.5 (Apr. 2016), pp. 525–527. ISSN: 1546-1696. DOI: 10.1038/nbt.3519. URL: http://dx.doi.org/10.1038/nbt.3519.

[81] B. Munsky, G. Neuert, and A. van Oudenaarden. "Using gene expression noise to understand gene regulation." *Science* 336.6078 (Apr. 2012), pp. 183–187. ISSN: 1095-9203. DOI: 10.1126/science.1216379. URL: http://dx.doi.org/10.1126/science.1216379.

[82] A. Raj and A. van Oudenaarden. "Single-molecule approaches to stochastic gene expression." *Annual Review of Biophysics* 38.1 (June 2009), pp. 255–270. ISSN: 1936-1238. DOI: 10.1146/annurev.biophys.37.032807.125928. URL: http://dx.doi.org/10.1146/annurev.biophys.37.032807.125928.

[83] M. I. Love, W. Huber, and S. Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* 15.12 (Dec. 2014). ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8. URL: http://dx.doi.org/10.1186/s13059-014-0550-8.

[84] S. Strauss and R. Jungmann. "Up to 100-fold speed-up and multiplexing in optimized DNA-PAINT." *Nature Methods* 17.8 (June 2020), pp. 789–791. ISSN: 1548-7105. DOI: 10.1038/s41592-020-0869-x. URL: http://dx.doi.org/10.1038/s41592-020-0869-x.

[85] S. C. M. Reinhardt et al. "Ångström-resolution fluorescence microscopy." *Nature* 617.7962 (May 2023), pp. 711–716. ISSN: 1476-4687. DOI: 10.1038/s41586-023-05925-9. URL: http://dx.doi.org/10.1038/s41586-023-05925-9.

[86] J. Hellmeier et al. "Quantification of absolute labeling efficiency at the single-protein level." *Nature Methods* 21.9 (Apr. 2024), pp. 1702–1707. ISSN: 1548-7105. DOI: 10.1038/s41592-024-02242-5. URL: http://dx.doi.org/10.1038/s41592-024-02242-5.

[87] P. Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature Methods* 17.3 (Feb. 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. URL: http://dx.doi.org/10.1038/s41592-019-0686-2.

[88] K. Clement et al. "CRISPResso2 provides accurate and rapid genome editing sequence analysis." *Nature Biotechnology* 37.3 (Feb. 2019), pp. 224–226. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0032-3. URL: http://dx.doi.org/10.1038/s41587-019-0032-3.

[89] G. Sholder et al. "3'Pool-seq: an optimized cost-efficient and scalable method of whole-transcriptome gene expression profiling." *BMC Genomics* 21.1 (Jan. 2020). ISSN: 1471-2164. DOI: 10.1186/s12864-020-6478-3. URL: http://dx.doi.org/10.1186/s12864-020-6478-3.

[90] Y. Takei et al. “Multiplexed dynamic imaging of genomic Loci by combined CRISPR imaging and DNA sequential FISH.” *Biophysical Journal* 112.9 (May 2017), pp. 1773–1776. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2017.03.024. URL: http://dx.doi.org/10.1016/j.bpj.2017.03.024.

[91] Y. Takei et al. “Integrated spatial genomics reveals global architecture of single nuclei.” *Nature* 590.7845 (Jan. 2021), pp. 344–350. ISSN: 1476-4687. DOI: 10.1038/s41586-020-03126-2. URL: http://dx.doi.org/10.1038/s41586-020-03126-2.

[92] R. C. Challis et al. “Systemic AAV vectors for widespread and targeted gene delivery in rodents.” *Nature Protocols* 14.2 (2019), pp. 379–414.

Chapter 4

OBSERVATIONS ON MIRNA

Through the completion of these projects, several key principles of miRNA regulation have emerged:

1. No miRNA was able to regulate a single target below a certain complementarity limit of 14-17 nucleotides.
2. Cooperativity enables regulation of targets with low complementarity to the miRNA.
3. When reduced complementarity, multimerized targets are used in an IFFL, they produce regions of constant expression.
4. MicroRNAs can mediate strong repression (>10-fold) via 8 nucleotide, canonical target sites if they are repeated three or four times.

The model most consistent with the data is the "cooperative-linear" model. The "additive" model is inconsistent: reducing the number of cooperative target sites eliminates regulation, while the additive model predicted it would only weaken it by a few fold. The "cooperative-cubic" doesn't fit because it predicted that an IFFL would have a peak of gene expression and then decline proportional to the inverse square of the dosage. Instead, synthetic miRNA IFFLs produced regions of constant gene expression. This was consistent with the "cooperative-linear model," which also correctly predicts that single targets will be ineffective.

These observations could have important consequences in miRNA biology. MicroRNAs have been thought to be a gentle sculptor of gene expression. This thesis demonstrates that they can be powerful regulators, but only when acting in groups of at least three. This important concept has consequences for data analysis of miRNA gene expression perturbations, because the expected targets of miRNA regulation are not simply mRNAs that contain 8 nucleotide targets of that miRNA, but those that contain at least 2 targets of other miRNAs that are also expressed. Focusing on this set of mRNAs could greatly increase the signal to noise ratio of an experiment.

This thesis also produced useful devices that had important applications. Incoherent feedforward loops could play an important role in future gene therapies, to regulate their cargo and prevent overexpression toxicity. DIMMER circuits have the potential to become one the fundamental building blocks of synthetic circuits in the future—every project needs an IFFL. All together, these

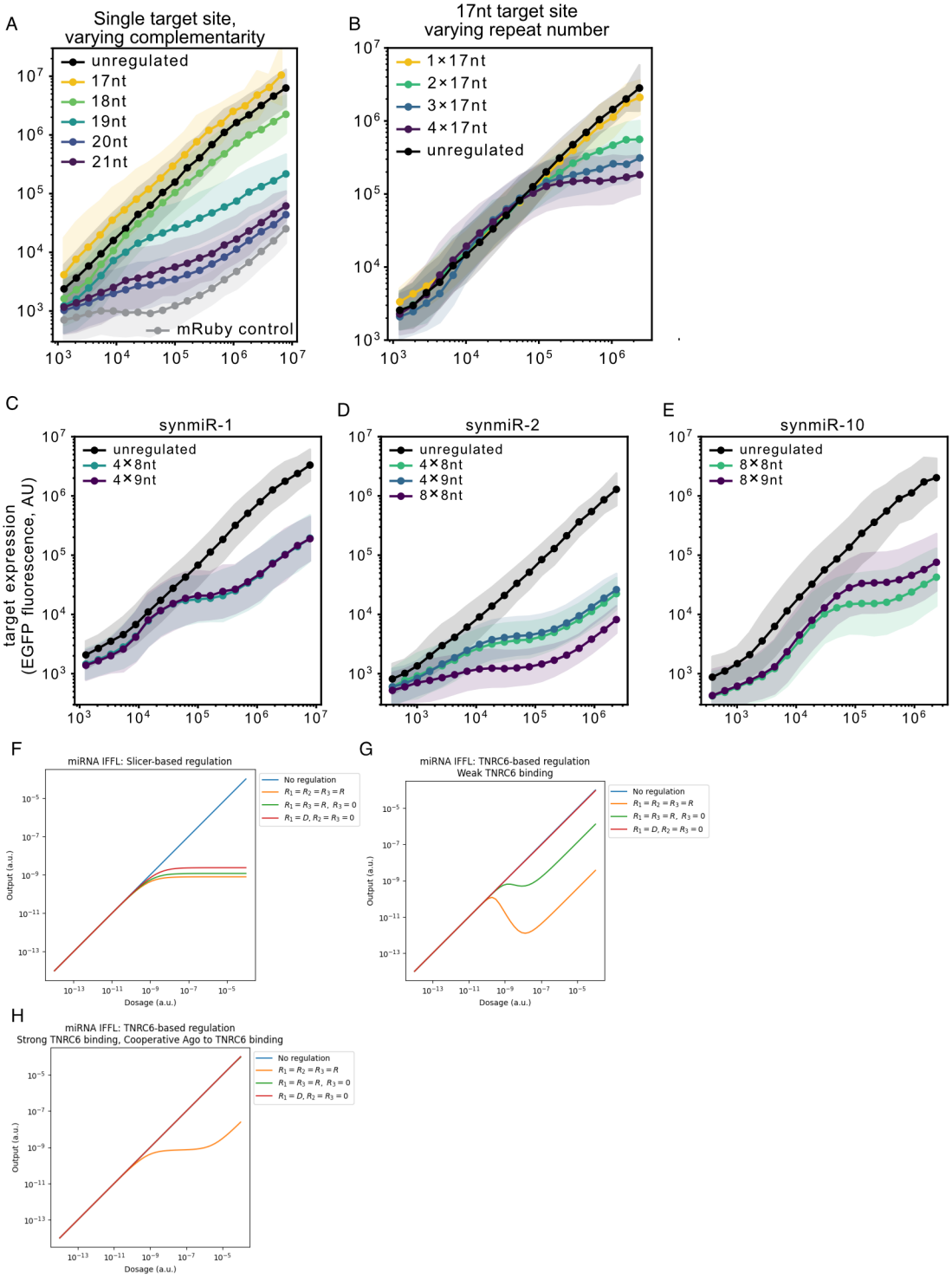

Figure 4.1: **Observations on miRNA. (A)** miRNAs are not able to regulate via a single target below a certain limit in complementarity. **(B)** cooperativity enables regulation of targets for low complementarity targets, and yields a region of constant expression in an IFFL. **(C-E)** miRNAs can mediate strong repression (>10-fold) via 8 nucleotide, canonical target sites. **(F-H)** The model most consistent with this data is strong TNRC6 binding model.

projects demonstrate that a “build to understand” approach can both advance understanding and help humanity.