Principles for Composing Genetic Circuits in Mammalian Cells with a Focus on miRNA Sensing

by

Jeremy Jonathan Gam

B.S.E, University of Michigan (2011)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Author .....
Department of Biological Engineering
July 12, 2018

Certified by .....
Ron Weiss
Professor of Biological Engineering and EECS
Thesis Supervisor

Accepted by .....
Forest M. White
Professor of Biological Engineering
Chair, Department Committee on Graduate Theses

Principles for Composing Genetic Circuits in Mammalian Cells with a Focus on miRNA Sensing

by

Jeremy Jonathan Gam

Submitted to the Department of Biological Engineering
on July 12, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

In this thesis, we developed two synthetic biology frameworks to facilitate the construction of useful genetic circuits, with a focus on circuits that sense miRNAs. miRNAs are an attractive biomarker for sensing since they regulate virtually all biological pathways in plants and animals, and because miRNA sensors can be easily designed by incorporating sequences complementary to the miRNA into a genetic circuit. Therefore, circuits that sense endogenous miRNAs can dynamically respond to cellular signaling or classify between cell types. However, the development of genetic circuits, and especially multi-input miRNA sensors, has traditionally been iterative, costly, and time-consuming.

To this end, we have developed a framework to measure miRNA activity and generate accurate predictions for sensors with multiple miRNA inputs. We started by building the largest library of miRNA sensors to date (620 sensors) and used the library to measure miRNA activity in several cell lines. We then constructed multi-input sensors and determined design rules for predicting their function, namely that miRNAs repress targets synergistically in opposite UTRs and antagonistically within the same UTR.

In our second framework, we developed a 'one-pot' method for high-information transfection and analysis that allows researchers to quickly determine performance of many tuned circuit variants in a single well. We used our one-pot method to quickly characterize a variety of genetic elements and to optimize the design of a miRNA sensor with inverted logic, a circuit topology we found difficult to design using traditional methods.

Finally, we incorporated synthetic miRNAs into the genome and used miRNA sensors to measure changes in endogenous gene expression in live cells, enabling modular sensing of transcription. These advances contribute to our knowledge of miRNA regulation, accelerate the design, construction, and testing of genetic circuits, and improve our ability to sense dynamic biological processes.

Thesis Supervisor: Ron Weiss
Title: Professor of Biological Engineering and EECS

Preface

Parts of this thesis are the result of collaborations and therefore involved the hard work of many people. As a result, the words and work presented here are not solely mine. A portion of the work in Chapter 1 concerning microfluidic DNA assembly was in collaboration with MIT Lincoln Lab and has been published in the article "Open-source, community-driven microfluidics with Metafluidics". Authors include David Kong, Todd Thorsen, Jonathan Babb, Scott Wick, myself, Ron Weiss, and Peter Carr. Chapter 2 describing the Ant/Syn miRNA repression model has been peer reviewed and published under the title "A mixed antagonistic/synergistic miRNA repression model enables accurate predictions of multi-input miRNA sensor activity" and was the work of myself, Jonathan Babb, and Ron Weiss. Chapter 3 was in a collaboration with Life Technologies, including Sanjay Kumar, Jon Chesnut, myself, Jon Babb, Yinqing Li, and Ron Weiss. Parts of Chapter 4 have been prepared for imminent submission for peer review, and were the work of myself, Breanna DiAndreth, Ross Jones, Jin Huh, and Ron Weiss. The work showcased in this thesis was supported in part by the National Institutes of Health (R01CA173712 and P50GM098792), Life Technologies, National Science Foundation Synthetic Biology Engineering Research Center (SynBERC), and the National Cancer Institute (core grant P30CCA14051).

Acknowledgements

Here are some of the giants on whose shoulders I'm grateful to have stood. First of all, a great many thanks to my advisor Ron Weiss for his tireless support, advice, and mentorship over the years. Having come to the SBC with virtually no knowledge of synthetic biology, I think it's safe to say that everything I've learned and accomplished has been due in large part to the Ron's guidance and the environment he has cultivated within the lab. I'd also like to thank my committee members Doug Lauffenburger and Phil Sharp for their valuable input and encouragement.

Special thanks to Bre DiAndreth and Ross Jones for bringing balance, perspective, and teamwork - without you this thesis would not have been remotely possible. Thanks also to Jon Babb for being a great mentor and showing me the ropes at the start of my PhD. I'd like to acknowledge the many collaborators I've had the pleasure to work with over the years: Jin Huh and the rest of the members of team cancer in the Weiss lab; the organoids team in the Weiss lab including Allen Tseng, Casper Enghuus, Patrick Fortuna, Sebastian Palacios, Katherine Kiwimagi, and Aishwarya Jagtap; collaborators at Lincoln lab including Peter Carr, David Walsh, David Kong, Todd Thorsen, Scott Wick; Jon Chesnut and Sanjay Kumar at Life Technologies; along with the rest of the members of the Weiss lab - thanks to all. I'm also grateful to everyone who kept the Weiss lab running every day, namely Brian Teague, Steve Firsing, John Scarpa, Cameron Haase-Pettingell, and Kalpana Jagtap. And I'd also like to thank everyone who has supported my career as a scientist, including Thommey Thomas and Alina Kotlyar at MNIMBS and Luis Hernandez-Garcia and Steven Allen at the fMRI lab at Michigan.

To my family, thank you for all your encouragement and support - it's cliché but I wouldn't be where I am without you. Thank you for giving me the resources and mindset for pursuing my dreams. Hope to make you guys proud every day.

And finally, shout out to all the friends I've made at MIT, especially the Biological Engineering class of 2012 graduate students. It's been a great run.

Contents

Preface **5**

Acknowledgements **6**

List of Figures **10**

List of Tables **14**

1 Technologies enabling miRNA sensing **15**

1.1 Introduction and background 15

1.1.1 Classification of live cells in synthetic biology 16

1.1.2 miRNAs 17

1.1.3 miRNA sensors and classifiers 20

1.1.4 Limitations and ways forward 24

1.2 miRNA sensor architecture and assembly 26

1.2.1 DNA assembly methods 26

1.2.2 DNA assembly using microfluidics 30

1.2.3 Golden Gate - PCR - Gibson Assembly 35

1.3 Delivery of miRNA sensors *in vitro* 39

1.4 Sensor designs for *in vivo* measurements 40

2 Accurate predictions of multi-input miRNA sensor activity **48**

2.1 Overview of miRNA repression and models 49

2.2 miRNA activity characterization and modeling 52

2.3 miRNA target site sets exhibit antagonistic interactions within a UTR and synergistic interactions across UTRs . . . 55
2.4 A novel Ant/Syn model uses miRNA activity to predict multi-input miRNA classifier performance . . . 57
2.5 miRNA sensors and classifiers based on Ant/Syn guidelines . . . 58
2.6 Discussion . . . 60
2.7 Modeling . . . 115
2.8 Design considerations . . . 118

3 miRNA gene tags for measuring transcription 124
3.1 Background and challenges . . . 125
3.2 Initial design of miRNA gene tag . . . 127
3.3 Gene tag to measure RHOXF2 transcription . . . 129
3.4 Discussion . . . 130

4 One-pot optimization of genetic circuits 137
4.1 Abstract . . . 137
4.2 Introduction . . . 138
4.3 Building a miRNA classifier using traditional design processes . . . 139
4.4 Results . . . 146
4.5 Poly-transfection characteristics . . . 153
4.6 Activation by Gal4-VP16 . . . 155
4.7 Characterization of Tet3G system . . . 157
4.8 Characterization of artificial miRNA for use as synthetic biology parts 159
4.9 Characterization of translational repressor L7Ae . . . 160
4.10 Building up to a cell-type classifier responding to miR-21-5p . . . 161
4.11 Additional analysis . . . 163
4.11.1 Complete workflow for one-pot optimization experiments . . . 215

5 Conclusions and future directions 228

6 Appendix 237
6.1 Methods for DNA assembly with microfluidics . . . 237

6.1.1 Genetic circuit assembly 237
6.1.2 Ligation assembly protocol 238
6.1.3 Ligation-assembled GFP expression circuit biological validation 239
6.1.4 Gateway assembly protocol 239
6.1.5 Gateway-assembled GFP expression circuit biological validation 240
6.1.6 Gibson assembly protocol 241
6.1.7 Gibson-assembled cell density dependent gene expression circuit biological validation 242
6.1.8 Golden Gate assembly protocol 243
6.1.9 Golden Gate-assembled aTc-inducible GFP circuit biological validation 244
6.1.10 Microfluidic device fabrication 244
6.1.11 Microfluidic controller 245
6.1.12 Microfluidic genetic circuit assembly 246
6.2 Methods for miRNA activity measurements and multi-input predictions 248
6.2.1 Construction of miRNA sensor library 248
6.2.2 Construction of miRNA classifiers 248
6.2.3 Cell culture 249
6.2.4 Transfection of miRNA sensors and classifiers into cell lines . . 250
6.2.5 Flow cytometry 251
6.2.6 Model-based fitting and prediction of miRNA activities 251
6.2.7 Statistical analysis 252
6.2.8 Data availability 252
6.3 Methods for one-pot optimization 253
6.3.1 DNA assembly framework 253
6.3.2 Cell culture 254
6.3.3 Transfections 254
6.3.4 Flow cytometry 255
6.3.5 Data analysis 255
6.3.6 Statistical analysis 257

List of Figures

1-1 **miRNA sensors and classifiers** 22
1-2 **Multi-input miRNA classifier design and function** 23
1-3 **Golden Gate-Gibson assembly of multi-TU circuits.** 28
1-4 **AATG overhang in mMoClo reduces expression by approximately 2-fold.** 29
1-5 **Ring mixer microfluidic device for DNA assembly.** 32
1-6 **Results from assembly of genetic constructs in a microfluidic ring mixer device.** 35
1-7 **Assembly of multi-TU plasmids without intermediate vectors.** 38
1-8 **Design of a miRNA sensors for *in vivo* and *ex vivo*.** 41

2-1 **A biochemical model explains miRNA repression measured using miRNA sensors.** 67
2-2 **Methods for combining miRNA activities in predictions of multi-input low sensors.** 68
2-3 **Interactions between miRNA target site sets in the 3' UTR sensors appear antagonistic.** 70
2-4 **Positioning miRNA target site sets in separate UTRs yields synergistic interactions** 72
2-5 **Ant/Syn model provides accurate predictions of 4-input miRNA classifiers.** 74
2-6 **miRNA target site position affects cell classifier performance substantially.** 76

2-7 **Ant/Syn model predicts the effects of target site position and number on the trade-off between sensitivity and specificity.** . 78
2-8 **ROC curves for classifiers illustrate specificity and sensitivity tuning by miRNA target set location and number.** 80
2-9 **Low sensor library data in HEK293FT and HeLa cells.** 82
2-10 **Enhanced repression by cooperation with up to four target sites.** 84
2-11 **Effect of spacer length on miRNA low sensors.** 86
2-12 **Goodness of fit for the miRNA repression model.** 89
2-13 **The model explains different miRNA repression curve shapes** 90
2-14 **Workflow for predictions based on a synergistic-only repression model vs the Ant/Syn model.** 92
2-15 **Gating of EYFP+ cells (HEK293FT) from EYFP- cells (HeLa and HepG2).** 95
2-16 **Antagonistic effects are maintained in miRNA sets containing distinct target sites.** 96
2-17 **Speculative mechanistic model for miRNA repression observed in this study.** 98
2-18 **Antagonistic interactions not due to repression limits** 100
2-19 **Speculative ODE model for antagonistic and synergistic repression.** 102
2-20 **Minimal resource sharing between miRNA sensors of related and unrelated miRNAs.** 104
2-21 **Repression of sensors bearing targets with and without uORFs.** 106
2-22 **Predictions for Ant/Syn model** 107
2-23 **Predictions for antagonist-only model** 108
2-24 **Predictions for the synergistic-only model** 109
2-25 **Comparison of predictions for all three models** 110

3-1 **miRNA gene tag overview** 126
3-2 **Gene tag measurements from endogenous Sox2.** 133
3-3 **miRNA gene tag in RHOXF2.** 135

4-1 **Circuit diagram for the initial miRNA high classifier.** 142
4-2 **Tests of initial miRNA high classifier with TALER14 repressor.** 143
4-3 **Design of miRNA classifier variants.** 144
4-4 **Output from miRNA classifier variants.** 145
4-5 **Ratios in output from classifier variants.** 146
4-6 **Comparison of co-transfection and poly-transfection methods.** 167
4-7 **Rapid optimization of a miRNA classifier using poly-transfection.** 170
4-8 **Co-transfection experiment for dCas9-VPR transcriptional activation.** 174
4-9 **One-pot characterization of L7Ae translational repression.** . . 176
4-10 **One-pot characterization of Doxycycline-induced Tet3G activation.** 177
4-11 **One-pot characterization of artificial miRNA activity and orthogonality.** 179
4-12 **Comparison of co- and poly-transfections for artificial miRNA experiment.** 181
4-13 **miRNA activity for each artificial miRNA tested.** 183
4-14 **Additional information for the miR-21-5p classifier.** 184
4-15 **Demonstration of subsampling for poly-transfections.** 186
4-16 **Tuning gene expression using upstream open reading frames (uORFs).** 189
4-17 **Tuning gene expression using promoter truncations.** 191
4-18 **Qualitative comparison of methods to characterize and optimize genetic circuits and systems.** 192
4-19 **Simplified assembly strategy for one-pot-ready plasmids.** . . . 193
4-20 **Poly-transfection distributions can be changed by titrating DNA.** 194
4-21 **Expression distribution from sequential electroporation and lipid transfection.** 195
4-22 **Poly-TX with different cationic transfection reagents.** 196
4-23 **Screening of fluorescent proteins to determine compatible sets.** 198

4-24 **Normalization of poly-transfection data to reduce effects of transfection distribution.** 200
4-25 **Rapid characterization of a miRNA classifier and its subcomponents.** 203
4-26 **Data for 4-component miRNA classifier.** 204
4-27 **One-pot characterization of Gal4-VP16 activation.** 207
4-28 **ROC-like curves to compare 3- and 4-component miRNA classifiers.** 208
4-29 **Analysis of 4-component classifier with machine learning.** . . 211
4-30 **Heatmap for binned VPR activation data across the full input parameter space.** 212

List of Tables

1.1 **Assembly efficiency with microfluidic assembly compared to conventional format.** 31
1.2 **Assembly efficiency with Golden Gate-PCR-Gibson assembly.** 39

2.1 **Best candidate miRNAs with specific activities among tested cell lines.** 111
2.2 **miRNA target combinations for sensors with targets in the 5' UTR.** 112
2.3 **Rate constants used for miRNA repression model** 113
2.4 **Golden Gate overhangs used for assembly** 114

4.1 **Quick start guide for poly-transfections.** 213
4.2 **DNA amounts for co-transfection testing of miR-21-5p classifier** 214

Chapter 1

Technologies enabling miRNA sensing

1.1 Introduction and background

Synthetic biology aims to build new and useful systems from biological components. While applications of synthetic biology are incredibly diverse - ranging from production of biofuels to engineering of the immune system to treat cancer - the central principles for designing and testing biological systems are unifying. In general, researchers will mine and characterize genetic 'parts' that perform a biological process (for instance activating or repressing transcription), generate models for composing multiple parts together into higher order systems termed genetic circuits, physically construct genetic circuits, test the performance of the genetic circuits, and modify the models according to the new results. These steps form the basis of a design-build-test-learn (DBTL) cycle that allows iterative improvements towards a given specification. To date, many major advances have been made at each of the DBTL

steps, accelerating the pace at which functional genetic circuits can be implemented. Even so, continued efforts to more accurately model and predict genetic circuits are required to make it feasible to understand and engineer systems that meet or surpass the complexity already possible with natural systems. In this thesis, we have worked to expand the characterization, modeling, and testing knowledge base to include miRNAs, small RNA molecules important for regulation. We believe several of the developments shown here are also translatable to other areas in synthetic biology, while also advancing the potential for smarter therapeutics that can actively respond to changes in miRNA signaling.

1.1.1 Classification of live cells in synthetic biology

In many fields, classification is achieved with a 'classifier' which takes in several attributes comprising an observation and determines which type or category that observation belongs to. Examples of classifiers in medicine and biology include measurements of expression of multiple genes for diagnostics, computer vision for pathology, distinguishing between species in taxonomy, and classification of different cell types. Classification of cell types is of particular interest in many application areas, for instance using biomarker expression to distinguish cancerous cells from normal ones, or to identify intermediate cell types in processes like stem cell differentiation or reprogramming. However, classification is often performed with fixed cells or cell lysate since techniques like high throughput RNA-seq or microarrays require access to cellular RNA or DNA. This limitation uncouples classification from the ability to initiate downstream effects in live cells and hampers followup experiments with a particular subset of cells. Also, live cell classification would enable therapeutics to respond to signaling changes to potentially reduce off-target effects or actuate

dynamic outputs in cells. Therefore, we are interested in developing live cell classifiers that can respond to biomarkers and restrict expression of a signal to only a particular cell state (e.g. cells in a cell cycle stage, in a certain differentiated state, or exhibiting abnormal signaling) or cell type (e.g. to mitigate off-target effects by expressing toxic proteins in cancer cells rather than normal cells, or to reduce off-target expression in liver to reduce side effects). Circuits that can exploit cell-state and cell-type classification will be immensely more useful than those that cannot, since large gains in safety, efficacy, and circuit complexity would be possible. We are especially interested in using miRNAs as biomarkers for cell classification due to the ability to easily measure miRNA activity using reporters with Watson-Crick complementarity.[1]

1.1.2 miRNAs

microRNA (miRNA) molecules are short sequences of non-coding RNA that are important for post-transcriptional regulation of mRNA. miRNAs are typically transcribed from DNA as primary miRNA transcripts by RNA polymerase II. Primary transcripts may include multiple miRNAs as a cluster and may be expressed from intergenic regions of the genome or from introns or untranslated regions (UTRs) of genes[2]. The sequence encoding the miRNA forms a stem loop structure; a portion of the stem encodes the particular miRNA which is ~22 bp in length and may be located either toward the 5' and/or 3' end of the loop depending on the particular miRNA[3]. The primary transcript is then processed into a pre-miRNA transcript which retains only the stem loop structure by the microprocessor complex which contains Drosha and DiGeorge Critical Region 8 (DGCR8). The pre-miRNA transcript is then exported from the nucleus into the cytoplasm by exportin, a process which

requires Ran-GTP. A Dicer complex cleaves the loop portion of the transcript leaving short double stranded RNA with ~2bp 3' overhangs[4,5]. Finally, the mature miRNA guide strand is loaded into the RNA induced silencing complex (RISC) which then mediates mRNA destabilization or translational repression of transcripts bearing a miRNA target site[6]. Target sites often correspond to regions complementary to the seed sequence (nucleotides 2-8) of the miRNA, though less canonical binding can occur if other factors like signficant complementarity outside of the seed are met.[7,8]

Many advances have been made in characterizing miRNA concentration and activity in cell lines. Landgraf et al. sequenced over 350 small RNA library preparations cloned from a variety of tissues in humans, mice and rats[9]. Cloning frequency was taken as a measure of miRNA expression and correlated well with concentration measured by Northern blot. However, the amount of information gained from the experiments was relatively small compared to modern RNA-seq approaches, and the authors noted potential biases in the cloning protocol which may limit comparisons to other datasets. In another study, expression of 847 miRNAs was assayed in the NCI-60 cell lines using microarrays[10]. Though only 495 unique miRNAs were detected and only moderate correlation was observed in the four cell lines with existing expression data, cell lines were generally clustered according to miRNA expression and miRNAs remained differentially expressed according to tissue type. However, information about miRNA sensor performance that can be extracted from these studies of miRNA abundance is limited since correlations between miRNA abundance and apparent activity is weak[11]. Nevertheless, measures of abundance remain important for estimating whether a miRNA is active until more detailed miRNA activity measurements can be produced. One area where high throughput sequencing strategies have been crucial is the identification of thousands of potential miRNAs in humans, with several hundred of these meeting established criteria for designation

as high confidence miRNAs[12]. Moreover, methods like Argonaute CLIP-seq[13] and CLASH[14] have enabled the high-throughput discovery of miRNA target sites across the transcriptome.

Initial experiments reported that miRNAs can act to translationally repress their targets,[15] while later studies found that miRNAs can also cause miRNA degradation.[16,17] As a result, there has been debate as to which miRNA repression mechanisms may dominate and in what contexts.[18] Researchers have used microarrays and RNA-seq from polysomes to conclude that effects from translational inhibition are more modest compared to effects from changes in mRNA quantity.[19,20] Timing effects have also been shown, with translational inhibition showing greater effects compared to mRNA degradation only at shorter timepoints.[21] Following these studies, sequencing to determine poly(A) length (PAL-seq) was able to show that the difference in repression mechanism could be due mainly to shortening of the poly(A), where at early developmental stages shortening of the poly(A) acts mainly to decrease translational initiation and at later developmental stages acts to decrease mRNA concentration.[6] As a result, the current understanding of miRNA repression mechanism is that translational repression accounts for rapid but low magnitude responses to miRNA, while mRNA deadenylation and subsequent degradation account for the majority of effects at later time points.[22] Others have also noted that different mRNA isoforms may also affect the degree of translational inhibition.[23] The current model for miRNA repression involves Argonaute (Ago) binding to TNRC6 which also interacts with poly(A)-binding protein (PABPC) and recruits the CCR4-NOT complex which deadenylates the mRNA, resulting in mRNA decapping and degradation.[24] The CCR4-NOT complex also recruits DDX6 which may inhibit translation.[25]

miRNAs are understood to mediate most of their effects on endogenous targets at target sites within the 3' UTR. However, analysis of CLIP-Seq datasets or target

site conservation suggest the presence of many CDS and 5' UTR target sites, albeit at lower frequency than targets in the 3' UTR.[26,27] Measurement of miRNA effects on targets in the CDS across the transcriptome show small but significant effects on their targets, and enhance regulation from targets in the 3' UTR.[28] Similarly, mRNAs with miRNA targets in the 5' and 3' UTRs exhibit enhanced regulation[27] and mRNAs with targets introduced into the 5' UTR can repress as well as targets in the 3' UTR.[29] As a result, characterizing the effects of target sites outside of the 3' UTR should be important, especially for synthetic biology where target sites can be placed wherever they may be most effective. Of related interest is the notion that translational inhibition appears to occur through blocking of initiation or ribosome drop-off near the start site,[30–32] and the demonstration that LIN41 RNA binding protein silences mRNA using tranlsational repression when binding the 5' UTR and mRNA degradation for the 3' UTR.[33] These results set the stage for our study of synthetic miRNA sensors that have miRNA targets within the 5' and 3' UTRs.

1.1.3 miRNA sensors and classifiers

As a result of the number of miRNAs and prevalance of target sites, miRNAs can serve as useful biomarkers for cell type and cell state.[34,35] Therefore, researchers have worked towards developing sensors to detect diseases ranging from cancer to Alzheimer's disease and heart disease, often for circulating miRNAs.[36–38] A natural extension for miRNA sensing is to apply genetically-encoded sensors in live cells to provide information about cell state in real time or for long-term monitoring. These genetically-encoded miRNA sensors offer several advantages important for research purposes, most significantly, the ability to directly actuate transcriptional responses.[1,35,39,40] Our group has demonstrated a miRNA-based synthetic DNA cir-

cuit for distinguishing a given cell line from a set of six other background cell lines.[35] The basic sensor unit was designed as a construct containing four tandem repeats of the sequence perfectly complementary to the miRNA of interest (called a miRNA target site 'set') located in the 3' UTR of a gene of interest. miRNA low sensors (which provide output when the sensed miRNA activity is low) were developed as a fluorescent reporter (DsRed) followed by miRNA target sites while high sensors (which provide output when miRNA activity is high) were constructed with a LacI repressor followed by miRNA target sites controlling the output (Fig. 1-1). Previous studies have shown four target site repeats to be sufficient for repression of targets with further increases in the number of targets having diminishing effects on repression and greater potential for unwanted miRNA titration effects.[1,41,42] Multiple low and high sensors may be combined to form multi-input miRNA classifiers with greater cell-type specificity as shown in Figure 1-2.

miRNA sensors should ideally exhibit the following properties: significant increase or decrease in expression within a desired range of miRNA expression activity, minimal titration effects on endogenous miRNA activity, and output specific to only the miRNA of interest. These characteristics may be obtained with optimal design of the underlying logic function and proper tuning of expression (eg. by promoter strength, repressor strength, or small molecule induction), and are especially important in multi-input sensors. Though expression changes in single sensors are typically modest (~2 to 10-fold), with the proper design of multi-input miRNA profilers greater specificity can be achieved.[35]

The use of miRNA sensors to distinguish cells in a variety of conditions began with Brown et al.[1] Knockdown was shown 1) in dendritic cells induced with lipopolysaccharide (LPS) compared to a no LPS control using a sensor consisting of GFP followed by four repeats of miR-155 target sites, 2) in hepatocytes and Kupffer

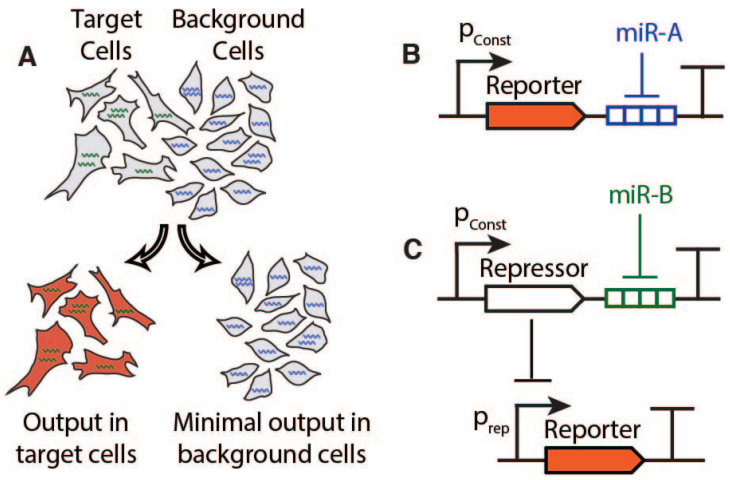


Figure 1-1: miRNA sensors and classifiers
A. An ideal classifier. Illustration showing behavior of an ideal classifier with output exclusively in target cells and not background cells; **B.** Circuit diagram of an example low sensor. With high activity of miR-A, reporter expression is knocked down, producing little output. With low activity of miR-A, reporter expression is unsuppressed; **C.** Circuit diagram of an example high sensor. With high activity of miR-B, repressor expression is knocked down, allowing the reporter to be expressed. With low activity of miR-B, repressor expression is unsuppressed and the reporter is repressed effectively.

cells compared to endothelial cells in mouse liver using GFP followed by miR-142-3p and miR-122a target sites, and 3) in embryoid bodies using separate sensors for miR-372, miR-302a, and miR-124a. This study demonstrated the utility of using miRNA sensors for actuating cell-state and cell-type specific responses, despite the limited number of miRNAs that were tested.

An advance in the number of miRNAs tested for activity was made by Mullokandov et al.,[11] with the use of pooled single sensor transduction into cells followed by fluorescence-activated cell sorting and next generation sequencing to determine en-

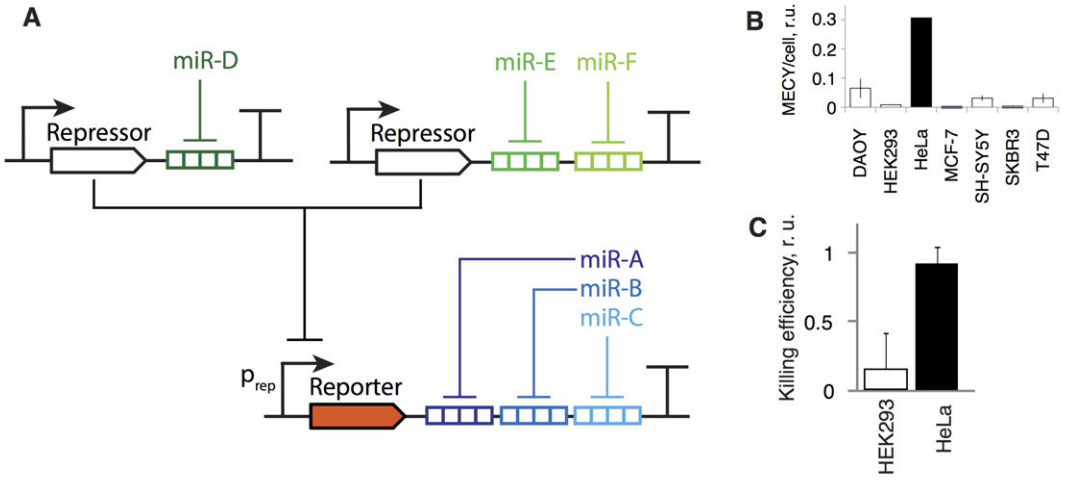


Figure 1-2: Multi-input miRNA classifier design and function
A. Design of a multi-input miRNA sensor for distinguishing HeLa cells by Xie et al.[35] A combination of LacI and FF4 miRNA was used as the repressor. Repressor expression was activated by separate activators not shown here. Reporter was a DsRed reporter or hBax apoptosis regulator. miRNA identities were as follows, A: miR-141, B: miR-142-3p, C: miR-146a, D: miR-21, E: miR-17, F: miR-30a; **B.** DsRed fluorescence for HeLa cells compared to 6 background cell lines. Expression was at least 5-fold higher in HeLa cells compared to the cell line with next greatest expression; **C.** Selective killing of HeLa cells. Apoptosis was induced preferentially in HeLa cells compared to HEK293 cells when the reporter was replaced by the hBax protein.

richment of miRNAs in high sensor expression or low sensor expression bins. Sensors with significant enrichment in low sensor expression showed downregulation when tested by conventional flow cytometry. The pooling approach allowed 291 miRNA target sites (both perfectly complementary and bulged versions) to be tested simultaneously per cell line, a much greater number than had been tested previously. Interestingly, only 40% of measured miRNAs had significant activity in the three cell lines tested. miRNAs that were more abundantly expressed were more likely to show suppressive activity (80% of suppressive miRNAs were expressed above 100

reads per million), though abundance and activity remained poorly correlated even in this regime. Perfectly complementary target sites were generally more suppressed than bulged target sites and are likely to demonstrate lower probability for miRNA titration effects.[42] One limitation to this approach includes limited resolution of the degree of repression for an individual miRNA in a given condition, since enrichment is only measured across 4 expression bins. While the statistical significance for miRNA activity could be determined, fold-changes in activity - an important metric for design of multi-input circuits - were difficult to quantify.

Another development included the use of adeno-associated virus (AAV) miRNA sensors to measure activity of 115 miRNAs for 12 cell lines[43]. The authors again demonstrated limited correlation between miRNA abundance and activity on the AAV platform. Ability to distinguish cell states using this method was also shown as expected changes in miRNA activity profiles were obtained when inducing K562 cells with 12-O-tetradecanoylphorbol-13-acetate. Benefits for the AAV platform include low cost and simple storage, but drawbacks include small circuit size and complexity in preparation.

1.1.4 Limitations and ways forward

Despite these studies, our understanding of how to efficiently design genetically encoded miRNA sensors, especially those sensing multiple miRNAs simultaneously, has been limited. Existing models of miRNA sensors have explained the threshold-like nature of miRNA repression, but have been largely explanatory rather than predictive - thus limiting their use for sensor design.[41,44] Additionally, the existing data measuring miRNA activity has used fold enrichment metrics which are not ideal for generating predictions.[11] As a result, our work has focused on building a synthetic

biology framework for measuring miRNA activities, generating rules for combining miRNA activities for sensors responding to multiple miRNAs, and then comparing predictions to measurements from multi-input miRNA sensors.

Using synthetic biology techniques to probe design rules has been an effective way to study biological systems.[45] For example, in *E. coli*, the study of combinatorial promoters with shuffled transcription factor (TF) binding sites and copy numbers elucidated several design rules, including the limits of regulation and effects of repressor and activator locations.[46,47] Another study where transcriptional and translational efficiencies were varied using promoters and RBSs showed how those parameters contributed to noise in expression.[48] In mammalian systems, researchers have studied how chromatin regulators dynamically control gene expression using synthetic fusion of chromatin regulators to a DNA binding protein. They found that each regulator silences or reactivates in an all-or-nothing fashion, and could be modeled by stochastic traversal through three states: active, reversibly silent, and irreversibly silent.[49] Study of the bone morphogenetic protein (BMP) pathway has shown that the complex system involving 20 ligand and 7 receptor types could be explained by competitive receptor-ligand interactions, resulting in complex behavior in some ways reminscent of a neural network with hidden layers.[50] As a result, the pathway can give many different outputs depending on the combination of input ligands. In another example, reconstitution of Hedgehog signaling gradients using engineered sender and receiver cells and resulting models were able to explain what signaling activities were essential to gradient generation.[51] Specifically the PTCH receptor required bifunctional intracellular and extracellular inhibition to recapitulate robust and fast responses to morphogens. And finally, a paper demonstrating a synthetic system for titration of miRNA using additional target sites showed that ceRNA was possible only with an extremely large (non-physiological) amount of tar-

gets introduced.[52] Therefore, our work described here using miRNA sensors to probe the limits of miRNA repression serves as another example where synthetic biology can contribute to knowledge of how a biological system functions.

1.2 miRNA sensor architecture and assembly

1.2.1 DNA assembly methods

A fundamental part of synthetic biology today involves the cloning and assembly of genetic circuits using DNA. This allows genetic 'parts' including promoters, insulators, protein coding sequences, and transcriptional terminators (among others) to be placed within designated positions so that the correct functionality can be achieved. Once DNA has been assembled, downstream steps like transformation, transfection, or in vitro transcription or translation are then possible.

To assemble circuits for mammalian cells, we initially planned to use existing hierarchical DNA assembly methods. A Gateway-Gibson method[53] would allow for promoters and genes to be assembled into a common backbone, but with a few caveats. First, the sequences surrounding the promoter-gene pair cannot be easily changed, constraining the circuits to only those that do not require regulation of sequences like the 5' and 3' UTRs or insulators. One possible workaround would be to append the UTR sequences to the promoter or gene sequences but this would require the cloning of many different gene/UTR combinations, drastically reducing the modularity of such a cloning system. Since miRNA sensors require insertion of miRNA target sites into the UTRs, such a system would not be amenable to miRNA sensor cloning.

The mammalian Modular Cloning (mMoClo) system[54] remedies several of the

limitations of the Gateway-Gibson system. In the mMoClo system, each insulator, promoter, 5' UTR, coding sequence, 3' UTR, and poly-A sequence is inserted into its own plasmid, designated as plasmid level 0 (or pL0). Then transcription units (TUs) are assembled into a level 1 (pL1) plasmid which contains one of each of the pL0 parts. This assembly is mediated by a Golden-Gate step[55] where a type IIS restriction enzyme, usually BsaI, which generates unique matching overhangs on only parts that should be assembled adjacently to each other. After this step, a level 2 plasmid (pL2) can be assembled using a different type IIS restriction enzyme, to give a final plasmid encoding multiple TUs. The mMoClo system should be sufficient to allow for assembly of the most complex genetic circuits that have been demonstrated in mammalian cells to date. However, we encountered some restrictions, with the largest being that the efficiency of pL2 plasmid assembly fell significantly as more transcription units were added.

As a result, we modified the mMoClo system to use Gibson assembly for pL2's instead of Golden Gate, since Gibson assembly is better able to assemble several large fragments together (Fig. 1-3).*[56] Moreover, this approach allows pL2 plasmids to incorporate the expression vectors from Gateway-Gibson assembly in case those have already been assembled previously. As an aside, other assembly techniques are arising as possible future alternatives to Gibson, notably ligase cycling reaction and twin primer assembly which should enable 12 or more fragments to be assembled at a time.[56,57] Also, we built several pL1 plasmid backbones with Gibson sequences reverse complemented such that orientation of transcription units could be switched. A final modification involved changing the overhang between promoter and 5' UTR from AATG to CAGA, in order to eliminate the possibility for an upstream open reading

*Work initiated by Jin Huh in the Weiss lab

frame. Changing this overhang increased output by 2-fold, a moderate increase but potentially important for outputs that need to be highly expressed in cells for a phenotype (Figure 1-4*).

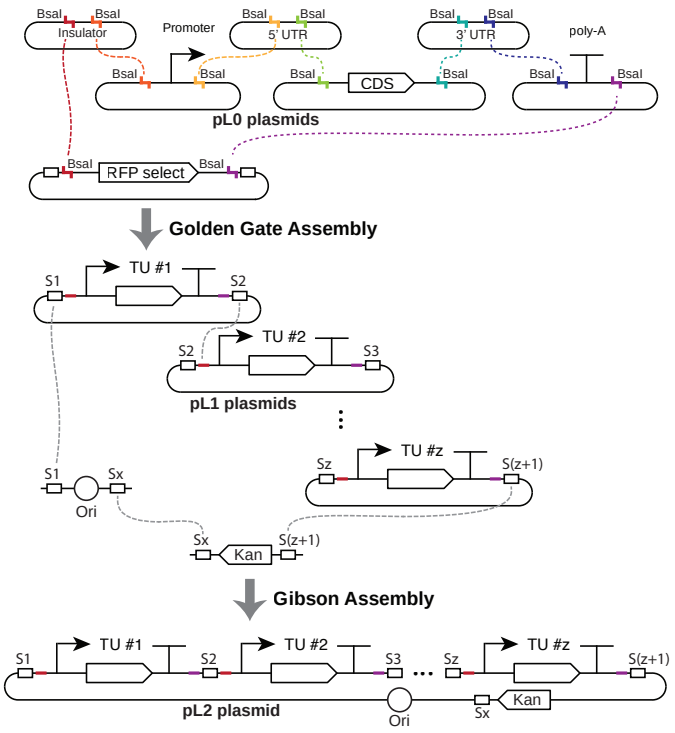


Figure 1-3: Golden Gate-Gibson assembly of multi-TU circuits.
This diagram illustrates the two assembly steps that comprise much of the assembly framework used in this thesis. In the first Golden Gate step, genetic elements encoded on plasmid level 0 (pL0) vectors - including insulators, promoters, 5' UTRs, coding sequences, 3' UTRs, and poly-A sequences - are assembled into transcription units (TUs). This assembly is performed using a type IIS restriction enzyme to generate unique matching overhangs for each of these genetic elements, indicated here by the different colors. T4 ligase can then ligate the matching overhangs to generate the complete circular DNA. The resulting plasmids are designated plasmid level 1 (pL1). If multiple transcription units are required to be on a single construct, the

*Data can be found within folder `/jgam/Data/flow/2015-04-08_Extra_ATG/`

pL1 plasmids can be digested with I-SceI restriction enzyme to expose 45 bp long regions (S1, S2, ... S(z+1), Sx) that are used for homology in Gibson assembly. In Gibson assembly, a 5' exonuclease chews back DNA ends to generate long overhangs which anneal to matching overhangs on other pL1's. A polymerase fills in any gaps generated by the exonuclease and a DNA ligase eliminates nicks in the DNA. The final product of the Gibson assembly therefore encodes expression of multiple transcription units necessary for function of the genetic circuit.

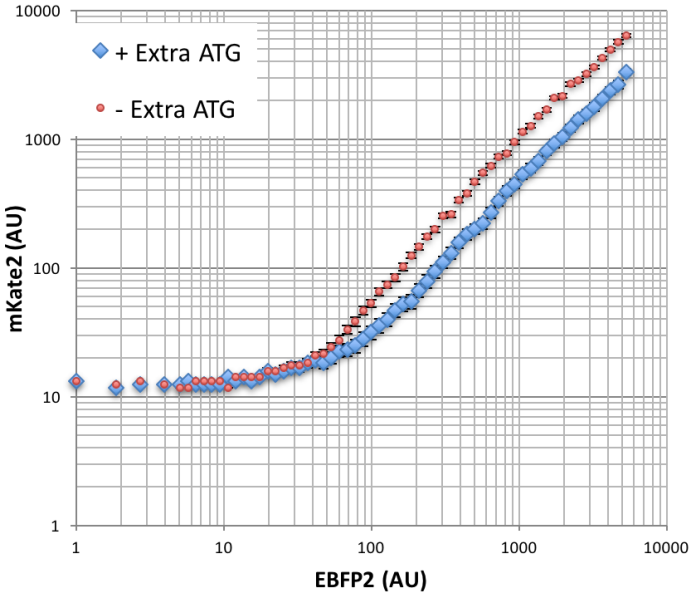


Figure 1-4: AATG overhang in mMoClo reduces expression by approximately 2-fold.
In an experiment to determine the effect of the AATG overhang present in the mammalian MoClo workflow,[54] we constructed a single plasmid encoding hEF1a-mKate2 and hEF1a-EBFP2 with either an AATG overhang between promoter and 5' UTR for the mKate2 TU or an AAAG overhang. The construct with the AATG overhang, which contains and extra ATG upstream of the start codon for mKate2, showed ∼2-fold reduced expression across all positive transfection levels, demonstrating that the AATG overhang should be changed to obtain higher possible expression levels and dynamic range. Values represent geometric means for mKate2 fluorescence binned across EBFP2 fluorescence.

1.2.2 DNA assembly using microfluidics

In addition to advancing the molecular biology platform, we also prototyped the use of microfluidics to scale down the assembly reaction volumes, reducing cost and increasing throughput. In our system, a 'ring-mixer' was built using conventional fabrication with polydimethylsiloxane (PDMS) and controlled using an open-source controller with 32 programmable channels. In this design, a circular ring can be loaded with up to three different liquids (e.g. DNA, enzyme mix, buffer) using dead-end filling. Then valves can seal off the ring and mixing can be performed using three channels that act as a peristaltic pump. After incubating the reaction, the mixture can be flowed out, collected, and transformed into e. coli for propagation. The ring-mixer uses approximately 10-fold less reaction volume compared to traditional reactions (300-650 nL compared to 2-10 uL) and we were able to demonstrate assembly using all four different assembly methods we tested: ligation, Gateway, Gibson, and Golden Gate. Note that while efficiency was decreased in each format (fewer colonies for ligation, Gateway, Gibson; lower proportion of correct colonies for Golden Gate) further optimization of microfluidic assembly protocols could allow for efficiencies to more closely match the larger format in the future (Table 1.1).

We tested the biological function of DNA assembled either microfluidically or in the conventional format. For each assembly, function was similar in both formats, showing that useful biological data could be obtained from microfluidically-assembled DNA. Thus we demonstrated that microfluidics could be used to scale down diverse types of of assembly reactions, and should be instrumental in accelerating large-scale assemblies in the future.*

*For further information, see Appendix 6.1 and our article "Open-source, community-driven microfluidics with Metafluidics" in Nature Biotechnology (2017)

Assembly Method	**Number of Parts**	**Final Size (bp)**	**cfu/uL (μfl. vs tube)**	**% Correct (μfl. vs tube)**
Ligation	2	5,362	7.0e3 vs 5.6e4	100% vs 100%
Gateway	2	3,479	2.5e3 vs 1.1e4	73% vs 75%
Gibson	4	4,708	2.6e2 vs 1.7e4	90% vs 100%
Golden Gate	5	5,571	5.1e4 vs 4.8e4	64% vs 85%

Table 1.1: Assembly efficiency with microfluidic assembly compared to conventional format.
Comparison of efficiency between microfluidic and conventional DNA assembly shows moderate drops (<10-fold) in the microfluidic format for all assemblies besides Gibson. Golden Gate assembly shows mostly comparable efficiency between both formats, indicating that future microfluidic assembly would require less optimization compared to the other assemblies.

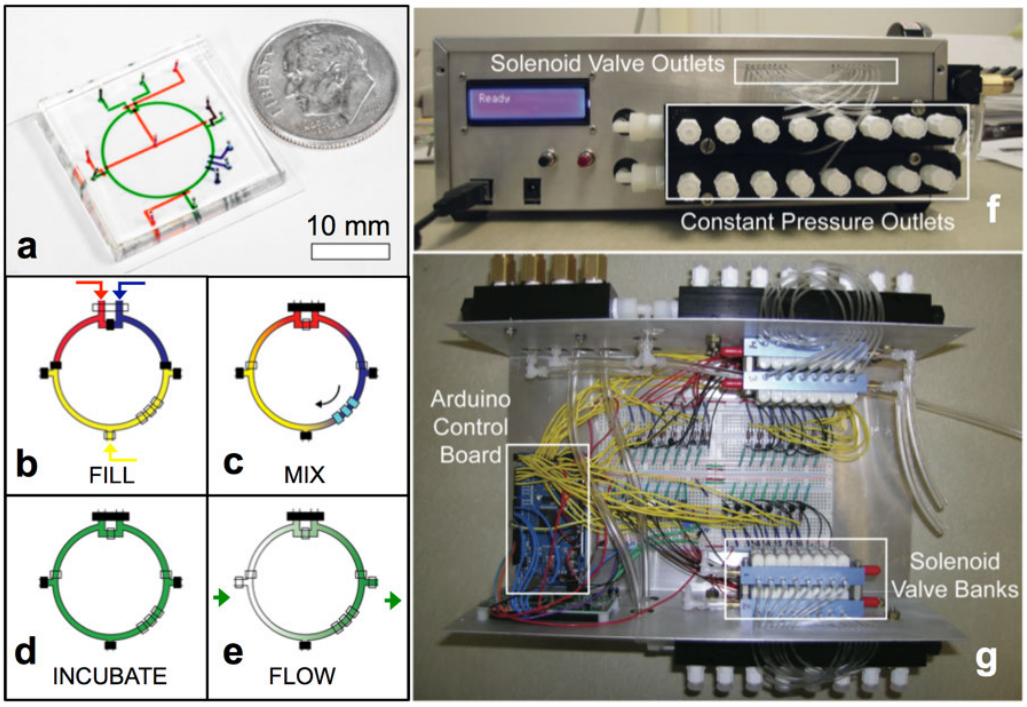
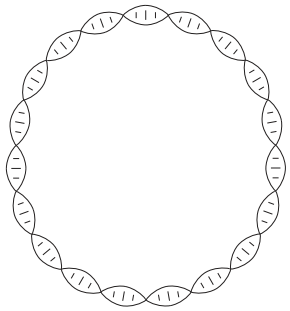


Figure 1-5: Ring mixer microfluidic device for DNA assembly.
(a) Photograph of the ring mixer device with chambers filled with food coloring. Green indicates chambers where DNA, enzymes, and buffer can be loaded. Red indicates control valves control valves. Blue indicates valves for mixing. **(b-e)** Different stages of DNA assembly with the ring mixer. DNA, enzymes, and buffer are loaded in the device with dead end filling (b), the reaction is mixed using valves with are sequentially actuated to induce peristalsis (c), the reaction is incubated at the temperature required (d), and the final reaction product is flowed out of the device (e). **(f)** Image of the 32-channel open source microfluidic controller used to operate the ring mixer. Solenoid and constant pressure outlets are highlighted. **(g)** Internal layout of the microfluidic controller. An Arduino and breadboard controls four sets of eight solenoid valve banks.



This page intentionally left blank to keep captions near figures.

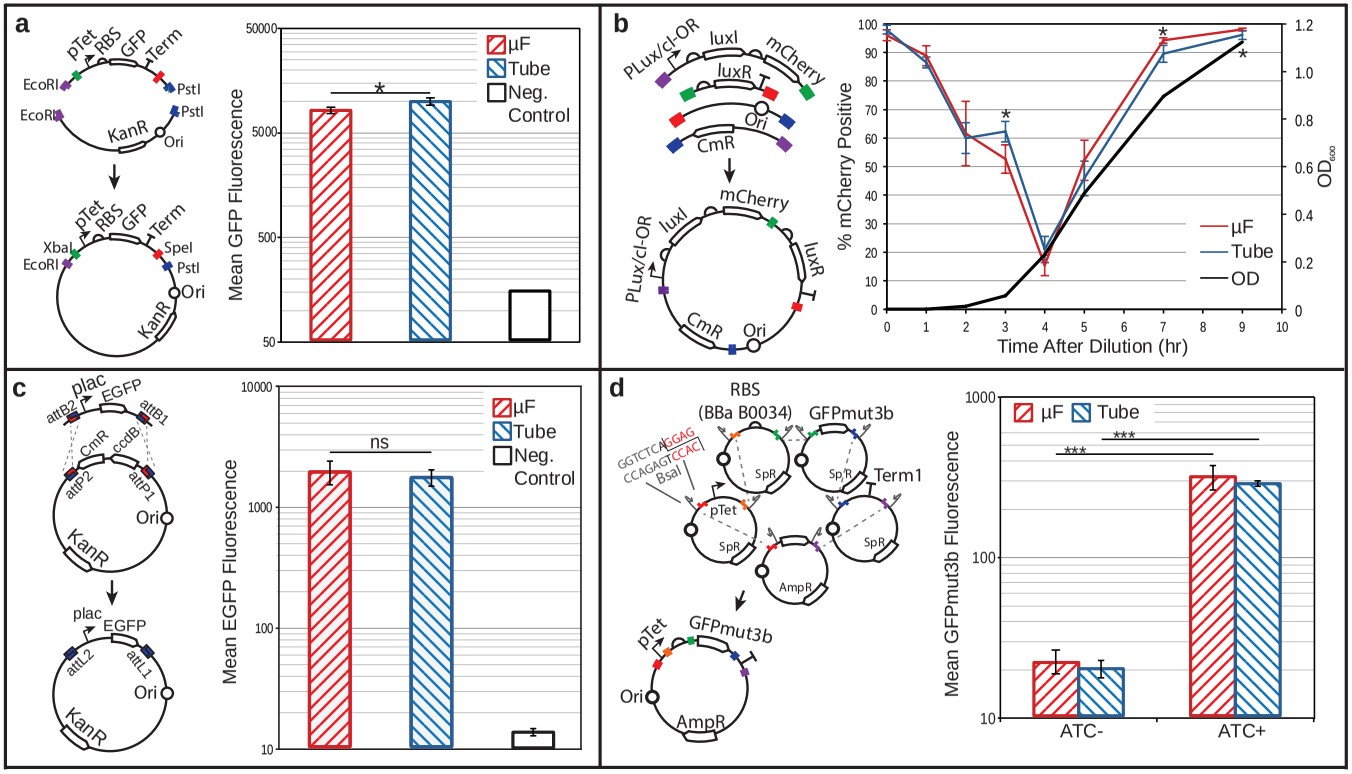

a
pTet
RBS
GFP
Term
EcoRI
PstI
EcoRI
PstI
KanR
Ori
pTet
RBS
GFP
Term
XbaI
EcoRI
SpeI
PstI
Ori
KanR
Mean GFP Fluorescence
50000
5000
500
50
*
μF
Tube
Neg. Control
b
PLux/cI-OR
luxI
mCherry
luxR
Ori
CmR
mCherry
PLux/cI-OR
luxI
luxR
CmR
Ori
% mCherry Positive
OD600
Time After Dilution (hr)
μF
Tube
OD
c
plac
EGFP
attB2
attB1
CmR
ccdB
attP2
attP1
Ori
KanR
plac
EGFP
attL2
attL1
Ori
KanR
Mean EGFP Fluorescence
10000
1000
100
10
ns
μF
Tube
Neg. Control
d
RBS
(BBa B0034)
GFPmut3b
GGTCTCA
CCAGAG
BsaI
EGAG
TCAC
SpR
SpR
Term1
pTet
SpR
SpR
AmpR
pTet
GFPmut3b
Ori
AmpR
Mean GFPmut3b Fluorescence
1000
100
10
μF
Tube
***
***
ATC-
ATC+

Figure 1-6: Results from assembly of genetic constructs in a microfluidic ring mixer device.
(a) Schematic of BioBrick-based ligation assembly. **(b)** Constitutive GFP fluorescence with ligation-assembled circuit. GFP fluorescence was compared for *Escherichia coli* transformed with microfluidic (four clones) and tube-assembled (three clones) circuits to *E. coli* lacking the circuit (negative control). When compared to cells lacking the circuit, both microfluidic and tube clones showed similar (>50-fold) increased fluorescence, with microfluidic clones exhibiting 58-fold increased fluorescence and tube clones showing 70-fold increased fluorescence. **(c)** Schematic of BP Gateway Assembly. (d) Constitutive EGFP fluorescence with Gateway-assembled circuit. EGFP expression was measured by flow cytometry for microfluidic (four clones) and tube (three clones) assembled circuits. Microfluidic and tube clones demonstrated >100-fold increased fluorescence compared to cells lacking the EGFP expression plasmid and did not show significantly different fluorescence relative to each other. **(e)** Schematic of Gibson assembly. **(f)** Cell-density-dependent gene expression with Gibson-assembled circuit. Following dilution from overnight culture at time zero, the fraction of cells expressing mCherry decreased to ∼20% after 4 hours. This percentage then rebounded as cell density increased starting at OD_{600} = 0.2. Microfluidic (five clones) and tube-assembled (three clones) reactions performed similarly, with no significant difference in the percentage of mCherry-positive cells in 5/8 time points. Of the time points with significantly different expression (indicated with asterisks) the greatest discrepancy was at 3 h with a 9.6% difference in percentage of cells expressing mCherry. **(g)** Schematic of Golden Gate Assembly. **(h)** aTc induction of GFPmut3b with Golden Gate–assembled circuit. With both microfluidic and tube-assembled circuits, addition of aTc induced more than a tenfold increase in GFPmut3b fluorescent signal in TetR-expressing cells. Microfluidic and tube-assembled circuits showed similar behavior with no significant difference in the induced and uninduced cases. Error bars indicate s.d.

1.2.3 Golden Gate - PCR - Gibson Assembly

In addition to microfluidics, we also prototyped accelerated DNA assembly methods to make the build portion of the DBTL cycle less of a potential bottleneck. We identified intermediate propagation steps in *e. coli* as one of the largest areas for

improvement. Namely, if the propagation steps could be eliminated, complex circuits containing multiple transcription units could be assembled in a single day, rather than the several day long process currently required. For a typical hierarchical assembly the following steps and times are needed for each hierarchical step: transcription units (i.e. promoters driving genes, along with surrounding regulatory elements) are assembled (several hours), DNA transformed into *e. coli* cells (1 hr), colonies grown on agar plates (16 hrs), colonies picked and grown in liquid media (12 hrs), DNA extracted from cells (1hr). With the propagation steps removed, assembly of multi-TU circuits (i.e. two hierarchical steps) could be accelerated from a $\sim$60 hour process to a $\sim$10 hour one, a significant improvement.

We initially tested whether we could circumvent intermediate propagation steps by simply taking the assembly products from the first step and instead of transforming them, using them directly for the second assembly step. We reasoned that if this initial step, a Golden Gate assembly in our framework, were efficient enough, there may be sufficient amounts of correct assemblies to allow for downstream Gibson assembly. However in our initial tests we were unable to recover significant numbers of correctly assembled constructs with multiple transcription units. We then sought to use cell-free amplification steps to replace amplification in *e. coli*. In this protocol, Golden Gate assembly products were used as template for PCR reactions using primers based on the Gibson sequences identified earlier[53]. Then PCR products could be used directly for Gibson assembly to produce multi-TU plasmids (Fig. 1-7a). After this assembly protocol, we identified DNA bands on agarose gel matching the expected product size of $\sim$4 kb. After transforming Gibson products into e. coli, selecting with Kanamycin, and counting colonies expressing GFP and/or mCherry from the two TUs assembled together, we found that constructs were correctly and efficiently assembled at a rate of 95.8%. The efficiency and number of

colonies were similar to the positive control reaction where the transcription units were obtained from PCR of sequence validated plasmid, rather than a Golden Gate assembly product (Table 1.2).

Therefore we demonstrated a proof of concept for faster assembly where propagation of assembly products is achieved using cell-free techniques. Testing of assembly with many more TUs should be conducted to determine the limits for the Golden Gate - PCR - Gibson method. Since we conducted the proof of concept reactions, others have shown the ability to assembly up to six transcription units using a similar protocol.[58] We anticipate that similar procedure could be used to speed up MoClo protocols, if pL1 plasmids were modified to have distinct regions for PCR amplification outside of the Golden Gate cut sites. Other interesting future directions may include the use of rolling circle amplification (RCA) or other methods, in order to circumvent the final propagation step in *e. coli* to obtain transfection-ready DNA. Products from RCA would likely need to be digested with restriction enzyme prior to transfection to allow for delivery into cells and the nucleus, but RCA would likely be faster than growing and harvesting from bacteria.

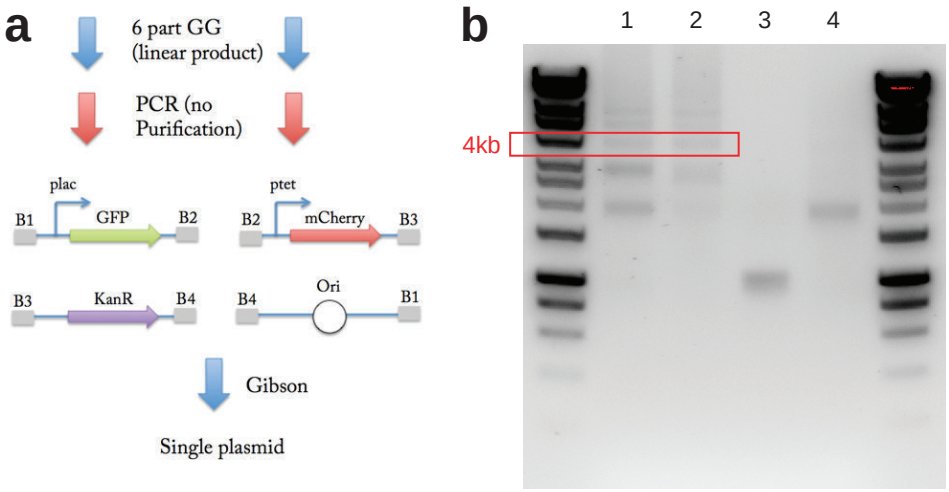


Figure 1-7: Assembly of multi-TU plasmids without intermediate vectors. **(a)** Assembly diagram for Golden Gate - PCR - Gibson protocol. Two different Golden Gate reactions were separately performed to make templates encoding constitutive bacterial expression of GFP and mCherry. The Golden Gate products were used directly for PCR without a purification step using primers based on the Gibson sequences (B1, B2, B3, and B4). The two resulting PCR products were combined with PCR products encoding antibiotic resistance for Kanamycin and the bacterial origin, then used in Gibson assembly to give a single plasmid encoding both GFP and mCherry. **(b)** Agarose gel for assembly products. Lane 1 shows the final assembly product for Golden Gate - PCR - Gibson, with a band at the expected 4 kb product along with other side products. Lane 2 indicates the positive control with PCR from sequence validated plasmid template (instead of Golden Gate product), followed by Gibson assembly. Lane 3 is a negative control for the PCR inputs without conducting Gibson assembly; all fragments were around 1 kb in length. Lane 4 shows a negative control where Gibson assembly was conducted without addition of GFP or mCherry TUs. The single band is at 2kb as expected of a linear product containing only KanR and the origin.

	GFP only	mCherry only	Neither	Both	% correct
GG-PCR-Gib	5	4	3	271	95.8
PCR-Gib	3	5	1	186	95.4

Table 1.2: Assembly efficiency with Golden Gate-PCR-Gibson assembly. Counts of colonies expressing GFP and mCherry show highly efficient assembly with either the Golden Gate - PCR - Gibson (GG-PCR-Gib) assembly or the positive control with PCR then Gibson (PCR-Gib).

1.3 Delivery of miRNA sensors *in vitro*

In mammalian cells, one of the simplest methods to deliver exogenous DNA (or other molecules like RNA and protein) into cells to express genetic circuits is through transient transfection. Commonly a cationic lipid or polymer is complexed to the negatively charged DNA which are endocytosed by cells, before endosomal escape of the complexes and import of DNA into the nucleus.[59,60] Chemical transfection methods are generally safe, easy to perform, and inexpensive,[61] and for these reasons we chose to use chemical transfection for our large-scale measurements of miRNA activity.

In order to scale down and simplify these measurements, we used a reverse transfection method[62] to make many replicates of 'transfection plates' onto which a DNA/transfection reagent mix is spotted. These transfection plates allow for transfection of cells simply by plating cells into them. Since many such plates can be made at a time and stored for later use, new cell lines can be rapidly tested without having to prepare the transfection mix each time. Here we adapted the reverse transfection protocol to 96-well plates, enabling measurements via flow cytometry with a suitable plate sampler. While microscopy could be used to measure fluorescence as in the

original reverse transfection procedure, many more cells can be collected using flow cytometry without having to store large image files and conduct more complicated image processing steps. Briefly, our reverse transfection protocol involves mixing a specified amount of DNA with a modified effectene reagent and then adding gelatin. A large batch of each mix can be made at a time (e.g. >100 uL) and 4 uL spotted into wells in a 96-well plate. Plates are then covered with sterile covers and can be stored at -80C until cells are added. In this way, we reverse transfected HEK293FT, HeLa, and HepG2 cells with the full miRNA low sensor library.

It should be noted that many cell lines and cell types - especially primary or non-dividing cells - demonstrate low transfection efficiency with chemical transfection. Alternative platforms exist for transducing these cells including 96-well electroporation[63] and AAV or other viruses.[43] Though since our work in developing design rules for multi-miRNA repression was more foundational and could be performed in easy-to-transfect cell lines, we chose to focus on easier transfection methods.

1.4 Sensor designs for *in vivo* measurements

In order to expand the types of cells and tissues we can measure, we constructed miRNA sensors with alternate outputs for miRNA activity measurements *in vivo* and *ex vivo*.* These sensors use a bi-directional CMV promoter to drive Nluc-mNeonGreen with regulation from miRNA target sites in the 3' UTR and also constitutive Fluc-TagBFP as a reference signal (Fig. 1-8) This delivery platform would allow the usual flow cytometry measurements using mNeonGreen and TagBFP signals, and those data would be amenable to analysis and modeling discussed

*In collaboration with Sophie Strobel and Jin Huh

later in Chapter 2. The addition of NanoLuc and firefly luciferase allow for measurements from luciferase assays and bioluminscence imaging, which can be done *in vivo* or *ex vivo*. The bi-directional promoter allows for a smaller design, with the total length of 5.3 kb excluding bacterial origin and antibiotic resistance on the plasmid backbone. Finally, the plasmid contains sequences for integration and propagation in an HSV delivery platform, allowing for delivery of the miRNA sensor into hard to transfect cells or *in vivo* models. These sensors should enable researchers to measure miRNA activity in more directly related to application areas like disease models and therapeutics.

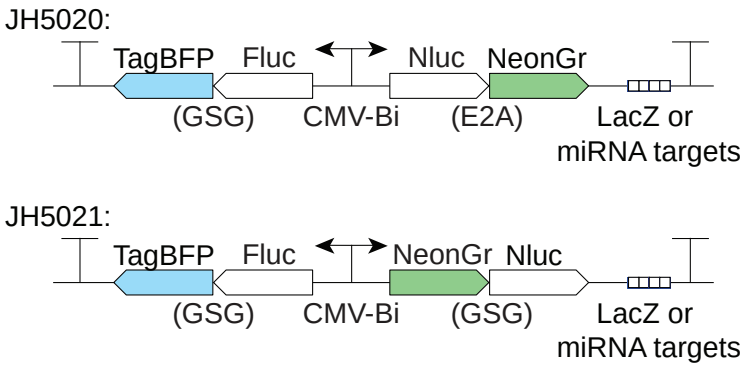


Figure 1-8: Design of a miRNA sensors for *in vivo* and *ex vivo*.
We designed miRNA low sensors for use beyond *in vitro* experiments. These constructs express both fluorescent proteins and luciferase such that measurements can be made using flow cytometry, fluorescent microscopy, or bioluminescence. Two versions were constructed where Nluc was linked to mNeongreen either with a 2A linker or by making a fusion. In both cases, TagBFP was fused to Fluc since we only had one existing construct with those two proteins linked.

References

1. Brown, B. D. *et al.* Endogenous microRNA Can Be Broadly Exploited To Regulate Transgene Expression According To Tissue, Lineage and Differentiation State. *Nature Biotechnology* **25,** 1457–1467 (2007).

2. Lin, S. L., Miller, J. D. & Ying, S. Y. Intronic microRNA (miRNA). *Journal of Biomedicine and Biotechnology* **2006,** 1–13 (2006).

3. Kozomara, A., Hunt, S., Ninova, M., Griffiths-Jones, S. & Ronshaugen, M. Target repression induced by endogenous microRNAs: large differences, small effects. *PloS one* **9,** e104286 (2014).

4. Slack, F. J. *MicroRNAs in Development and Cancer* (2011).

5. Talebizadeh, Z. Regulation of Gene Expression by Small RNAs. *The American Journal of Human Genetics* **86,** 328–330 (2010).

6. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508,** 66–71 (2014).

7. Grimson, A. *et al.* MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell* **27,** 91–105 (2007).

8. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Molecular Cell* **64,** 320–333 (2016).

9. Landgraf, P. *et al.* A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129,** 1401–1414 (2007).

10. Sokilde, R. *et al.* Global microRNA analysis of the NCI-60 cancer cell panel. *Molecular Cancer Therapeutics* **10,** 375–384 (2011).

11. Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods* **9,** 840–846 (2012).

12. Kozomara, A. & Griffiths-Jones, S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42,** 68–73 (2014).

13. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460,** 479–486 (2009).

14. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153,** 654–665 (2013).

15. Olsen, P. H. & Ambros, V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology* **216,** 671–680 (1999).

16. Bagga, S. *et al.* Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122,** 553–563 (2005).

17. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769–773 (2005).

18. Cloonan, N. Re-thinking miRNA-mRNA interactions: Intertwining issues confound target discovery. *BioEssays* **37,** 379–388 (2015).

19. Hendrickson, D. G. *et al.* Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biology* **7,** 25–29 (2009).

20. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).

21. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science* **336,** 233–237 (2012).

22. Eichhorn, S. W. *et al.* MRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular Cell* **56,** 104–115 (2014).

23. Clancy, J. L. *et al.* mRNA isoform diversity can obscure detection of miRNA-mediated control of translation. *RNA* **17,** 1025–1031 (2011).

24. Chen, C. Y. A., Zheng, D., Xia, Z. & Shyu, A. B. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nature Structural and Molecular Biology* **16,** 1160–1166 (2009).

25. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics* **16,** 421–433 (2015).

26. Guo, Z. W. *et al.* MtiBase: A database for decoding microRNA target sites located within CDS and 5'UTR regions from CLIP-Seq and expression profile datasets. *Database* **2015,** 1–9 (2015).

27. Lee, I. *et al.* New class of microRNA targets containing simultaneous 5' -UTR and 3'-UTR interaction sites. *Genome Research* **19,** 1175–1183 (2009).

28. Fang, Z. & Rajewsky, N. The impact of miRNA target sites in coding sequences and in 3' UTRs. *PLoS ONE* **6,** 1–6 (2011).

29. Lytle, J. R., Yario, T. a. & Steitz, J. a. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 9667–9672 (2007).

30. Pillai, R. S. *et al.* Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells. *Science* **309,** 1573–1576 (2005).

31. Humphreys, D. T., Westman, B. J., Martin, D. I. K. & Preiss, T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 16961–6 (2005).

32. Petersen, C. P., Bordeleau, M. E., Pelletier, J. & Sharp, P. A. Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell* **21,** 533–542 (2006).

33. Aeschimann, F. *et al.* LIN41 Post-transcriptionally Silences mRNAs by Two Distinct and Position-Dependent Mechanisms. *Molecular Cell* **65,** 476–489.e4 (2017).

34. Gurtan, A. M. & Sharp, P. A. The role of miRNAs in regulating gene expression networks. *Journal of Molecular Biology* **425,** 3582–3600 (2013).

35. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333,** 1307–1311 (2011).

36. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).

37. Leidinger, P. *et al.* A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biology* **14,** R78 (2013).

38. Ikeda, S. *et al.* Altered microRNA expression in human heart disease. *Physiological Genomics* **31,** 367–373 (2007).

39. Miki, K. *et al.* Efficient Detection and Purification of Cell Populations Using Synthetic MicroRNA Switches. *Cell Stem Cell* **16,** 699–711 (2015).

40. Sayeg, M. K. *et al.* Rationally Designed MicroRNA-Based Genetic Classifiers Target Specific Neurons in the Brain. *ACS Synthetic Biology* **4,** 788–795 (2015).

41. Bloom, R. J., Winkler, S. M. & Smolke, C. D. A quantitative framework for the forward design of synthetic miRNA circuits. *Nature Methods* **11,** 1147–53 (2014).

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

43. Tian, W. *et al.* High-Throughput Functional MicroRNAs Profiling by Recombinant AAV-Based MicroRNA Sensor Arrays. *PLoS ONE* **7,** 1–10 (2012).

44. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature genetics* **43,** 854–9 (2011).

45. Mukherji, S. & van Oudenaarden, A. Synthetic biology: Understanding biological design from synthetic circuits. *Nature Reviews Genetics* **10,** 859–871 (2009).

46. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Molecular Systems Biology* **3** (2007).

47. Kinkhabwala, A. & Guet, C. C. Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS ONE* **3** (2008).

48. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & Van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **31,** 69–73 (2002).

49. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351,** 720–724 (2016).

50. Antebi, Y. E. *et al.* Combinatorial Signal Perception in the BMP Pathway. *Cell* **170,** 1184–1196.e24 (2017).

51. Li, P. *et al.* Morphogen Gradient Reconstitution Reveals Hedgehog Pathway Design Principles. *Science* **360,** 543–548 (2018).

52. Yuan, Y. *et al.* Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 3158–63 (2015).

53. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Research* **41** (2013).

54. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Research* **42,** 13440–13451 (2014).

55. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS ONE* **4** (2009).

56. Kok, S. D. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synthetic Biology* **3,** 97–106 (2014).

57. Liang, J., Liu, Z., Low, X. Z., Ang, E. L. & Zhao, H. Twin-primer non-enzymatic DNA assembly: An efficient and accurate multi-part DNA assembly method. *Nucleic Acids Research* **45,** e94 (2017).

58. Halleran, A. D., Swaminathan, A. & Murray, R. M. Single day construction of multi-gene circuits with 3G assembly, 1–9 (2018).

59. Felgner, J. H. *et al.* Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *The Journal of Biological Chemistry* **269,** 2550–2561 (1994).

60. James, J. *et al.* A Novel Mechanism Is Involved in Cationic Lipid-Mediated Functional siRNA Delivery. *Molecular Pharmaceutics* **6,** 763–771 (2009).

61. Awais, M., Voronina, S. G. & Sutton, R. An efficient method is required to transfect non-dividing cells with genetically encoded optical probes for molecular imaging. *Analytical Sciences* **31,** 293–8 (2015).

62. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110 (2001).

63. McCoy, A. M., Collins, M. L. & Ugozzoli, L. A. Using the Gene Pulser MX-cell Electroporation System to Transfect Primary Cells with High Efficiency. *Journal of Visualized Experiments* **35,** 1–7 (2010).

Chapter 2

Accurate predictions of multi-input miRNA sensor activity

miRNAs regulate a majority of protein-coding genes, affecting nearly all biological pathways. However, many aspects of miRNA biology have not been quantitatively characterized, including frequency and potency of target sites outside of the 3' UTR, the impact of target site location on repression mechanism, composition rules for multiple target sites, and limits of cooperativity for genes regulated by many miRNAs. To this end, we explore miRNA biology at a quantitative single-cell level using a library of 620 miRNA sensors and reporters that are regulated by many miRNA target sites at different positions. Interestingly, we found that miRNA target site sets within the same untranslated region exhibit combined miRNA activity described by an antagonistic relationship while those in separate untranslated regions show synergy. Our resulting Antagonistic/Synergistic (Ant/Syn) computational model generated significantly enhanced prediction of activity for miRNA sensors containing many miRNA targets, accelerating development of sophisticated sensors for clinical and

research applications. These results shed light on the aforementioned understudied but important aspects of miRNA biology.

2.1 Overview of miRNA repression and models

microRNA (miRNA) molecules are short sequences of non-coding RNA that are important for post-transcriptional regulation of mRNA. Despite active study of miRNAs since their discovery, several aspects of miRNA repression remain unknown or controversial.[18] For instance, many of the proteins and mechanisms involved in miRNA repression and interactions between them have yet to be elucidated.[25,64] Also, most studies have focused on miRNA target sites in the 3' UTR,[65] but recent research has shown that targets in the coding sequence and 5' UTR can be important for modulating activity, especially in combination with other target sites.[27,28,66] Due to incomplete study of these interactions between target sites, there has been a lack of consensus for the importance of target sites outside of the 3' UTR and also insufficient knowledge to generate design rules and models for miRNA target site composition. To address these needs, we created a large library of reporter constructs with composable miRNA target sites and used them in various combinations to explore the effects of multi-miRNA regulation from 5' and 3' targets. We found that miRNA target site interactions follow an Antagonistic/Synergistic (Ant/Syn) model where sets of miRNA target sites exhibit antagonistic interactions within the same UTR (i.e. the amount of knockdown depends strictly on the miRNA target sites with highest activity), and synergistic interactions across UTRs (i.e. knockdown is a multiplicative combination of miRNA target sites). In contrast to previous computational models,[41,44] our Ant/Syn model accurately predicts simultaneous repression effects from many different miRNAs.

miRNA repression mechanisms are varied and complex, with molecular players including the Argonaut proteins, GW182, PABPC, the CCR-NOT complex, the PAN2-PAN3 complex, and decapping proteins.[25] Recent evidence shows even more possible interactions with DDX6, phosphorylation by CSNK1A1 and dephosphorylation by the ANKRD52-PPP6C complex.[64,67] Further complicating matters is the fact that miRNAs can mediate multiple modes of regulation including deadenylation, decapping, cleavage, and translational repression. Here we use highly expressed synthetic miRNA sensors to probe the limits of miRNA regulation, since quantitative measurements made at biological extremes can provide mechanistic insight otherwise difficult to obtain via conventional knockout or sequencing based techniques.[18,45] Our results suggest that the repression mechanisms for targets in the 5' UTR and 3' UTR may be distinct (e.g. translational repression in the 5' UTR and mRNA destabilization in the 3' UTR).

Importantly, our Ant/Syn model accurately captures repression behavior for transcripts simultaneously regulated by many different miRNAs. Nature is replete with examples of such highly miRNA-regulated genes; on average 7.3 different miRNAs repress each miRNA-regulated gene and 47 distinct genes are regulated by >40 miRNAs,[68] with p21Cip1/Waf1 experimentally verified to be targeted by 28 miRNAs.[69] Additionally, emerging evidence indicates a class of transcripts regulated by simultaneous 5' and 3' UTR targets of the same miRNA.[27] Based on our observation in synthetic circuits of synergistic interactions across UTRs even with targets for different miRNAs, we propose that this class of 5' and 3'-regulated transcripts should be broadened to include those with targets for different miRNA sequences.

Our results contribute to an enhanced quantitative understanding of miRNA biology which can then be applied to create better nucleic acid-based therapeutics (e.g. ones that are regulated dynamically by complex biomarker profiles). We are espe-

cially interested in using miRNAs as indicators of cell type and cell state, since there are thousands of distinct miRNAs which regulate >5,300 genes across almost all cellular pathways.[34,70,71] We and others have shown that genetically encoded miRNA sensors can be constructed by placing miRNA target sites in the UTRs of a reporter.[1,35,39,40] While most efforts have focused on sensing a single miRNA, our new approach more closely mimics endogenous biological regulation in that many miRNAs can regulate a single transcript, improving specificity and redundancy. Several studies have used miRNA profiles to identify diseases including cancer[36], Alzheimer's disease,[37] and heart disease.[38] Genetically encoded miRNA sensors (which sense a single miRNA input) and cell classifiers (which sense multiple miRNA inputs simultaneously) can provide information about disease state, actuate responses in cells specifically expressing either a diseased or healthy miRNA profile,[1,35,39] distinguish between subtypes of cells *in vivo*,[40] and help biologists study complex processes like stem cell differentiation.[72]

The desire for sophisticated miRNA classifier designs that perform increasingly more complex operations necessitates a deeper understanding of the composition rules that govern regulation of transcripts by many miRNAs. In this study we introduce miCAD - a workflow for measuring output of single-input miRNA sensors in cell lines, characterizing miRNA activity from miRNA sensor data using a biochemical model, utilizing the measured miRNA activity to make accurate predictions of multi-input miRNA classifiers using the Ant/Syn to inform better designs, and testing the best classifier candidates in cells. The accurate predictions generated by miCAD reduce or eliminate the need for multiple iterations of physical classifier designs by instead simulating a large range of classifier designs *in silico* and testing only the best candidate(s) experimentally. Interestingly, the use of our Ant/Syn model in miCAD often leads to improved classifier designs that are counterintuitive.

For instance, classifiers may perform dramatically better with rearranged target sites or even with fewer target sites, depending on the application. These and other design rules, both intuitive and counterintuitive, can help explain regulation of natural endogenous transcripts and improve the design of new sensor circuits, illustrating the power of convergence between quantitative biological modeling and analytical synthetic circuit design.

2.2 miRNA activity characterization and modeling

We designed, constructed, and tested a miRNA sensor library containing all 620 sequences of mature human miRNA designated as high confidence in miRBase 21.[12] Our library enables high-information-content screening of miRNA activity in cells and also serves as a source for sequence-validated templates of miRNA targets when building multi-input sensors. We synthesized miRNA target site sets bearing four repeats of perfectly complementary sequence to the miRNA and inserted them into the 3' UTR of a reporter construct (Fig. 2-1a). Throughout this work, a miRNA target site "set" refers to four repeats of a given miRNA target site. We chose to include four target sites per set since minimal increase in dynamic range was obtained with >4 targets (Fig. 2-10, 2-11). We chose to include perfectly complementary target sites to reduce the possibility of miRNA sponging effects[42] and increase the dynamic range of repression. The resulting single-input sensors demonstrate up to several-hundred-fold dynamic range, indicating the utility of miRNAs as potent biomarkers (Fig. 2-1a). To our knowledge, ours is the largest miRNA sensor library reported to date (the previous largest containing target sites for 291 miRNAs[11]) and the only one allowing for Golden Gate assembly[55] from single-input sensors into multi-input classifiers in a single assembly step. The library will be available on

Addgene for researchers to use in advancing the understanding of miRNA biology and the development of better DNA and RNA-based therapies.

We implemented a computational miRNA repression model to better understand and quantify the behavior of our sensors and classifiers. The model for single-input sensors is based on previous deterministic models[41,44] for miRNA activity, which we extended to multi-input sensors using the Ant/Syn model. At the single-input level, the model comprises the following reactions: 1) mRNA and protein molecules for two fluorescent reporters are transcribed and translated at first order rates and are degraded, 2) mKate2 reporter mRNA may be reversibly bound by a miRNA-containing complex which may then catalytically degrade the bound mRNA, 3) parameters corresponding to miRNA concentration (M, the total effective number miRNA molecules either free or bound to reporter mRNA), and Michaelis constant ($K_m = \langle k_{off} + k_{cat} \rangle / k_{on}$, a measure related to the repression strength per miRNA molecule) are fit using fluorescence data obtained experimentally from our sensors and together accurately describe the behavior of a given miRNA sensor (Fig. 2-1b). In concordance with previous models[41,44], our model predicts thresholding behavior with three output regimes when sensing a single miRNA. At low EBFP2 expression (i.e. low transfection efficiency) a repressed regime exists where mKate2 is significantly repressed relative to EBFP2. At high EBFP2 expression, an excess of reporter mRNA is present and saturates the miRNA machinery, resulting in a derepressed regime. And at intermediate EBFP2 expression, a threshold regime switches between repressed and derepressed behavior. Note that in this model we omit the effects of cellular division and associated dilution, since the other rate constants within the model operate at much faster time scales (ie. 48 hours for division compared to minutes to hours for other rates). Relatedly, it may be interesting generally explore the dynamics of gene expression and maturation of fluorescent proteins and their effects

on measurements from miRNA sensors. Such analysis may be especially important when designing miRNA sensors that act in a cascade or toggle.

We tested the ability of our single-input repression model to explain changes to experimental perturbations of M and K_m. For decreasing K_m, our model predicts no shift in threshold region - only a decrease of output within the repressed region (Fig. 2-1c, left). We modulated K_m experimentally by varying the number of target sites for miR-21-5p in the sensor. Ase expected from the model, we observed that increasing the number of target sites increased k_{on}, resulting in a decrease in K_m and increase in repression (Fig. 2-1c, right). This result was also confirmed for endogenous high activity miRNA in HEK293FT and HeLa cells (Fig. 2-10). For increasing total miRNA concentration (M), the model predicts a shift in threshold region to the right and a concomitant decrease of output within the repressed region (Fig. 2-1d, left). We modulated M by transfecting various amounts of miR-21-5p mimic in HEK293FT cells which have low endogenous miR-21-5p activity. The expected decrease in output was observed across increasing miR-21-5p amounts (Fig. 2-1d, right).

To experimentally determine the values of M and K_m for each of our miRNA sensors, we used a reverse transfection protocol[62] to introduce our library of sensor plasmids into common cell lines HEK293FT and HeLa. We then used computational methods to generate fits of M and K_m for each miRNA sensor in both cell lines, generating a database of miRNA activities. Fits based on the model captured the effects of miRNA-mediated repression well as indicated by low and normally distributed errors (Fig. 2-12). Most miRNA sensors exhibited low or no miRNA activity, consistent with previous reports of miRNA sensor measurements in several cell lines (Fig. 2-9).[11] To compare miRNA activity to miRNA expression in our cell lines, we submitted HEK293FT and HeLa for small RNA sequencing and estimated

miRNA expression levels using existing computational tools.[73] miRNA activity was weakly correlated with miRNA expression (Fig. 2-9), again consistent with observations in other cell lines.[11] We also show that miRNA activities measured using the sensor library are reproducible, with two biological replicates showing good correlation of a M/K_m metric which describes the maximal fold repression of a sensor due to miRNA activity (Fig. 2-9). Analysis of miRNA activities in HEK293FT, HeLa, and HepG2 helped reveal miRNAs specific to each cell line relative to the others, which we later used to construct a HEK293FT cell classifier (Table 2.1).

2.3 miRNA target site sets exhibit antagonistic interactions within a UTR and synergistic interactions across UTRs

Following the characterization of single-input miRNA sensors, we sought to use the obtained parameters to make predictions for more complicated multi-input cell classifiers. Because our initial purely synergistic model did not provide accurate predictions for multi-input sensor function, we evaluated models that include additive, antagonistic, or synergistic interactions (Fig. 2-2). These three types of interactions were implemented as in the Chou-Talalay method[74], which comprises a standard set of equations used to determine whether inhibitors interact with each other. Additive predictions were made assuming mutually exclusive inhibitors, antagonistic predictions were made using similar assumptions but with a Hill coefficient value near zero, and synergistic predictions were made by multiplying contributions from each target site set (Fig. 2-2b,c).

We then tested 3-input sensors (Fig. 2-3a) encoding several combinations of

representative low and high activity miRNA target site sets in the 3' UTR of the output and compared obtained data to the predictions (Fig. 2-3b). Antagonistic interactions best predicted the multi-input data as indicated by lower maximal fold error and mean squared error, while additive and synergistic predictions consistently overestimated miRNA activity (Fig. 2-3c). Interestingly, inclusion of up to three high activity target sets (12 target sites) had minimal effect on increasing repression further than the single highest activity target set, i.e. the 3-input sensor exhibits activity reflective of only the highest activity target set.

To fully explore the general composition rules for miRNA target sets, we tried spacing target sets further apart or placed them in separate UTRs and investigated whether interactions were antagonistic, additive, or synergistic. We first tested separating target sets withing the 3' UTR with spacers up to 600bp but still observed antagonistic interactions even with the longest spacers (Fig. 2-11). In contrast, we found that miRNA target sets in the 5' UTR exhibited synergistic miRNA repression when combined with target sets in the 3' UTR (Fig. 2-4). The synergistic interactions we observed suggest that cooperativity may be important for RISC binding or repression when target site sets are placed in different UTRs. Based on these results we hypothesized that models for predicting miRNA activity would need to take into account miRNA target set position in addition to M and K_m.

2.4 A novel Ant/Syn model uses miRNA activity to predict multi-input miRNA classifier performance

Next we tested whether we could generate more accurate predictions than previous repression models by taking into account antagonistic interactions of miRNA target sites within UTRs and synergistic interactions across UTRs. This novel model, which we term an Ant/Syn model, first calculates (on a per transfection marker basis) the antagonistic interactions within each of the UTRs (approximated as the minimum reporter expression observed for any single-input sensor within one UTR) and then calculates the synergistic interactions between the UTRs by multiplying the contributions of each UTR to obtain the final output (Fig. 2-14). This is in contrast to the antagonistic-only model where only maximal repression for any single miRNA is taken, or the synergistic-only model where all contributions for each miRNA target site are multiplied, regardless of UTR position.

To test the Ant/Syn model experimentally, we built several variants of 4-input miRNA sensors (i.e. classifiers) bearing two miRNA target site sets in the 5' UTR and two miRNA target sets in the 3' UTR (Fig. 2-5a). We selected twelve miRNA target sites for study based on their miRNA activity measured from single-input sensors - four each of high, medium, and low activity miRNAs. We assembled and assayed reporter expression of 36 different 4-input sensors in HEK293FT cells representing a diverse panel of activities (Table 2.2), one example of which is shown in Fig. 2-5b. In all 36 tested cases, the Ant/Syn model predicted final output similarly or better than the antagonistic-only or synergistic-only models (Fig. 2-5c). For all classifiers with behavior accurately predicted by the synergistic-only model, the antagonistic-only

model underestimated activity, while classifiers with behavior accurately predicted by the antagonistic-only model had activity overestimated by the synergistic model.

2.5 miRNA sensors and classifiers based on Ant/Syn guidelines

We next investigated whether the Ant/Syn model could be used to design better single-input sensors and multi-input cell classifiers containing miRNA targets in both 5' and 3' UTRs which would allow sensors and cell classifiers to leverage antagonistic and synergistic interactions to obtain better sensitivity or specificity in classifying different cell types. Such multi-input constructs also inform how endogenous miRNAs could utilize 5' UTR target sites to enhance conventional activity from the 3' UTR.[27] We built two versions of a sensor containing multiple target sets for miR-21-5p, which shows high activity in HeLa and HepG2 cells but low activity in HEK293FT cells. In one variant, target site sets were placed in only the 3' UTR, while in another variant, sets were placed in both the 5' and 3' UTRs. We also constructed a 3-input cell classifier containing target sites for miR-21-5p, miR-23a-3p, miR-106b-3p where targets were either all in the 3' UTR or separated into different UTRs (Fig. 2-6a).

When separately transfected into HEK293FT, HeLa, and HepG2 cells, all sensors and classifiers showed reporter knockdown in HeLa and HepG2 cells and retained high reporter expression in HEK293FT cells as expected (Fig. 2-6b). The classifiers with targets in separate UTRs showed a marked reduction in reporter expression for off-target cells at high transfection levels resulting in greater on/off ratio (280-fold on/off for miR-21-5p, 210-fold for 3-input), in contrast to classifiers with targets only in the 3' UTR (14-fold on/off for miR-21-5p, 29-fold for 3-input). Thus our new

constructs with target sets in both UTRs achieved approximately 20-fold (miR-21-5p sensor) and 7-fold (3-input classifier) improved on/off ratio, compared to constructs with target sets in only the 3' UTR. We also observed a reduction in false positive output consistent with the synergistic interactions anticipated for miRNA target sets placed in separate UTRs. We then tested whether we could enhance classification in a mixture of cell types. We chose to test in a co-culture of HEK293FT, HeLa, and HepG2 cells where HEK293FT cells expressed genomically integrated EYFP in order to provide an independent measure of whether cells classified as HEK293FT by our classifiers were indeed HEK293FT in origin (Fig. 2-15). Reporter expression in co-culture was similar to that in separate transfections, with classifiers designed to utilize synergistic interactions exhibiting reduced false positive rate at high transfection levels (Fig. 2-6c). Moreover, quantification of sensitivity and specificity shows that while sensitivity was high (>95%) for all designs, specificity was improved (>85%) only in the new designs based on the Ant/Syn model (Fig. 2-6c). We chose to place only miR-21-5p target sets into the 5' UTR of these classifiers to minimize the length of the 5' UTR - reducing the potential for 5' UTR structure effects on reporter expression - while still utilizing synergism across UTRs.

While it is possible to duplicate all miRNA target sites across both UTRs, this may not be desirable in cases where there is a size constraint, where target sites bear an 'ATG' (which can initiate translation of an upstream ORF), or when it is important to have synergistic interactions for different miRNA species (with duplicated target sets only the single miRNA species with highest activity would show synergism). Moreover target site duplication can be adjusted to optimize the trade-off between sensitivity and specificity according to whichever is more important for the given application. For example, high sensitivity may be more important for cell classifiers designed to specifically kill cancer cells to reduce the chance of cancer cells

escaping, while high specificity may be more important for classifiers designed to induce differentiation of a particular cell type at a particular differentiation stage (e.g. driving liver differentiation in a subset of cells in the endoderm but absolutely not elsewhere). We tested several variants of cell classifiers with and without target site duplication and examined the effects of duplication on classifier sensitivity and specificity. The Ant/Syn model predicts that high sensitivity and lower specificity will be obtained by removing target set duplication from both UTRs, which we observed in the data (Figure 2-7). The effects of miRNA target location and duplication were further analyzed using ROC curves (Figure 2-8). Again, classifiers with targets in both UTRs demonstrated improved accuracy and area under the curve (AUC) compared to those with targets in only the 3' UTR (Figure 2-8a-b), while those with target set duplication showed tradeoffs between specificity and sensitivity (Figure 2-8c) and in certain cases improved performance without target duplication (Figure 2-8d). These results also suggest that some endogenous transcripts may use target sites in the 5' UTR to tune sensitivity and specificity when responding to miRNA activity.

2.6 Discussion

We report miCAD, a framework for making accurate predictions of multi-input miRNA sensors using parts level characterization. Several aspects of miRNA target regulation comprise our current design framework. First, concatenation of miRNA target sites in a UTR increases repression of a target reporter up to approximately 4 repeats (Fig. 2-10, 2-16). 1-2 repeats are often insufficient to obtain significant repression and 3-4 repeats may be required even with high activity miRNAs, suggesting possible roles of cooperativity on miRNA repression. Second, concatenating

target 'sets' of different miRNAs within the 3' UTR does not result in increased repression, indicating an antagonistic interaction where the combined activity reflects only the miRNA target set with the greatest activity. Third, target sets in both 5' and 3' UTRs exhibit synergistic interactions which can be approximated as the multiplication of fold repression contribution from the 3' UTR with that from the 5' UTR. Finally, antagonistic and synergistic interactions are combined into our proposed Ant/Syn model. Based on our observations, the Ant/Syn model is better able to explain data from a panel of 4-input classifiers than either the antagonistic-only or synergistic-only models. It may be interesting in the future to further explore if there are specific situations (individual miRNAs, combinations of miRNAs, certain miRNA activity regimes, etc.) where the Ant/Syn model makes better or worse predictions. Though such work may require testing of more than the 36 combinations we analyzed here. Our model represents a considerable advance compared to previous computational models of miRNA activity[41,44] in that it accurately models circuits that respond to multiple different miRNAs.

We also examined the potential for cooperative effects when concatenating multiple target sites into sets.[75] Interestingly, we observed apparent cooperativity across target sites for miRNAs even when different target sites are interleaved to make new mixed target sets. In other words, a set of interleaved target sites for four different high activity miRNAs represses similarly to a set of four targets of the same high activity miRNA (Fig. 2-16). This effect suggests that cooperativity could be mediated by Ago-Ago or Ago-protein-Ago interactions. One proposed protein mediator is GW182 which can interact with multiple Ago molecules simultaneously, forming a complex containing many Ago molecules.[76–79] When one miRNA target site-containing transcript is bound by Ago, other Ago molecules within the complex are more likely to bind the transcript, enabling cooperative repression (Fig. 2-17).

One potential complication for this model is that cleavage-independent repression mechanisms (e.g. mRNA deadenylation, target sequestration) should dominate over cleavage-dependent mechanisms. Otherwise each target site could be independently cleaved when bound by Ago2, leading to synergistic repression contributions for each target site. Since addition of exogenous siRNA along with complementary target sites was able to enhance repression of transcripts already containing three high activity target sets (i.e. siRNA targets act independently of miRNA targets), our data suggests that there could be two different repression mechanisms for siRNA and miRNA even when target sites are perfectly complementary, with siRNA utilizing cleavage and miRNA using cleavage-independent mechanisms (Figure 2-18). While siRNA and miRNA are known to repress via different mechanisms for endogenous transcripts[80], mechanisms for synthetic sensors with perfectly complementary target sites have been less studied. Other relevant evidence in plant biology suggests that cleavage-independent mechanisms like translational inhibition are important even though most miRNA target sites in plants are perfectly complementary[81] and it has been suggested that non-cleaving repression may be the default mechanism for miRNA repression in both mammals and plants, despite near-perfect complementarity for the latter.[82] Additionally, examples of miRNA-RISC complexes lacking cleavage capability have been described previously[83]. For instance, usage of Argonaut proteins 1, 3, or 4, which lack the cleavage capability of Ago2, could explain the antagonistic behavior we observed. In that case, processes downstream of Argonaut binding like sequestration, deadenylation, or translational repression could be the bottleneck for downregulation, which would explain why target site sets do not act independently. Future studies using overexpression or knockdown/knockout of Ago2 could potentially shed light on whether usage of other Argonauts besides Ago2 can explain why we observed Ant/Syn behavior.

In general we and others[29] have observed that miRNA target site sets placed in the 5' UTR showed similar repression levels to those in the 3' UTR. However, the role of 5' UTR miRNA target sites in endogenous transcripts has not been studied extensively. Since there appears to be ample potential for effective miRNA regulation at the 5' UTR, it appears likely that endogenous miRNA-mRNA targets within the 5' UTR should be more abundant than the six validated examples currently described in the literature.[84] Several studies have shown miRNA repression in 5' UTRs could be important in some transcripts, even though such repression is less-studied than that in the 3' UTR despite hundreds of possible 5' target sites revealed by computational predictions.[27,29] For comparison, the number of predicted target sites in the 3' UTR is on the same order of magnitude (though several fold greater) when the same computational approach is used.[27] Analysis of CLIP-seq data has also identified hundreds of thousands of target sites in CDS and 5' UTR regions in mammals.[26] Thus miRNA regulation of 5' UTRs may be an understudied aspect of miRNA biology and we provide a model that predicts the effects of such regulation.

Our observation of synergistic interactions across UTRs supports the notion that endogenous transcripts bearing miRNA targets in both UTRs, termed miBridges, would have enhanced repression.[27] Moreover, it may be possible that miBridges are not limited to transcripts bearing target sites for the same miRNA molecule in both UTRs, since we observed that different miRNA targets still exhibit synergistic interactions when placed in opposite UTRs (Figs. 2-4 and 2-5) and that combinations of different miRNAs act similarly to combinations with the same miRNA (Fig. 2-16). In this case miBridges could be a more general and common motif in biology, where miRNA targets in the 5' UTR enhance regulation in the 3' UTR, similarly to measured effects from targets in the CDS to those in the 3' UTR.[28]

An important remaining question is whether there is a predominant repression

mechanism for targets in the 5' UTR and whether the mechanism is different than that in the 3' UTR. The observation of distinct synergistic and antagonistic behavior based on whether target sites are on separate or the same UTRs suggests that distinct mechanisms for miRNA repression may exist for targets in the 5' UTR compared to the 3' UTR, otherwise we would expect to observe only antagonistic effects even with target sets in separate UTRs. The repression effects we observed could be explained if target sites in separate UTRs separately saturate distinct repression machinery (e.g. different subunits of the CCR4-NOT complex mediating deadenylation or translational repression). Indeed, numerical simulation of models incorporating shared or separate repression machinery were able to recapitulate the antagonistic and synergistic interactions respectively (Fig. 2-19). There are several alternative explanations for the mechanistic basis for the Ant/Syn behavior we observed. We tested whether steric effects (Fig. 2-11), miRNA sponging (Fig. 2-20), and inherent biological repression limits (Fig. 2-18) could explain the antagonistic interactions, though none were able to. Results from our and other studies[29] suggest that synergistic effects across UTRs could be due to the use of separate repression machinery depending on the UTR location. Since targets in the 3' UTR are generally considered to mediate mostly mRNA destabilization[22], it is possible that targets in the 5' UTR act mostly to translationally repress the targeted transcript (Fig. 2-17). In the 3' UTR of endogenous transcripts, approximately 6% to 26% of miRNA repression has been attributed to translational repression.[22] However, due to closer proximity of 5' UTR targets to translational machinery, it may be possible that a majority of repression there is due to translational repression. Interestingly, for the RNA binding protein LIN41 in *C. elegans*, different modes of repression have been observed depending on UTR location (i.e. translational repression in the 5' UTR and mRNA destabilization in the 3' UTR).[33] Results for LIN41 combined with our data, pro-

vide the attractive hypothesis that other RNA binding proteins like Ago may use a similar mechanism to mediate different modes of repression depending on location within the UTRs. Our results highlight the need for further studies on elucidating differences between repression from the 5' UTR, 3' UTR, and coding region for both synthetic miRNA sensors and endogenous transcripts.[18]

Also, we hypothesize that antagonistic interactions could act to reduce noise and increase redundancy in highly regulated genes, since if the miRNAs have similar activity, a decrease in activity of any single miRNA would not have significant impact on gene expression. This is in contrast to the stereotypical additive or independent (i.e. synergistic) models where fluctuation of any single miRNA would have effects on gene expression, propagation of noisy miRNA activity toward target gene expression. This mechanism of noise reduction could work in tandem with other modes where miRNAs can suppress noise at the network level.[85] As such, it may be possible that highly miRNA-regulated genes use antagonistic interactions within the same UTR to help reject noise and use synergistic interactions across UTRs to increase repression.

In summary, we describe the development of miCAD, which uses a new Ant/Syn model taking into account miRNA target set number, location, and interactions in order to accurately predict the behavior of multi-input miRNA sensors from well-characterized single-input sensors. In contrast to previous methods yielding fold enrichment metrics[11], characterization via miCAD provides highly detailed information of miRNA activity across a wide spectrum of transduction efficiencies and output levels, allowing accurate predictions of multi-input miRNA classifier function using our novel Ant/Syn miRNA repression model. miCAD contributes to synthetic biology by enabling predictable responses to miRNA inputs, and to biology by showing that miRNA repression must take into account location and number of target sites.

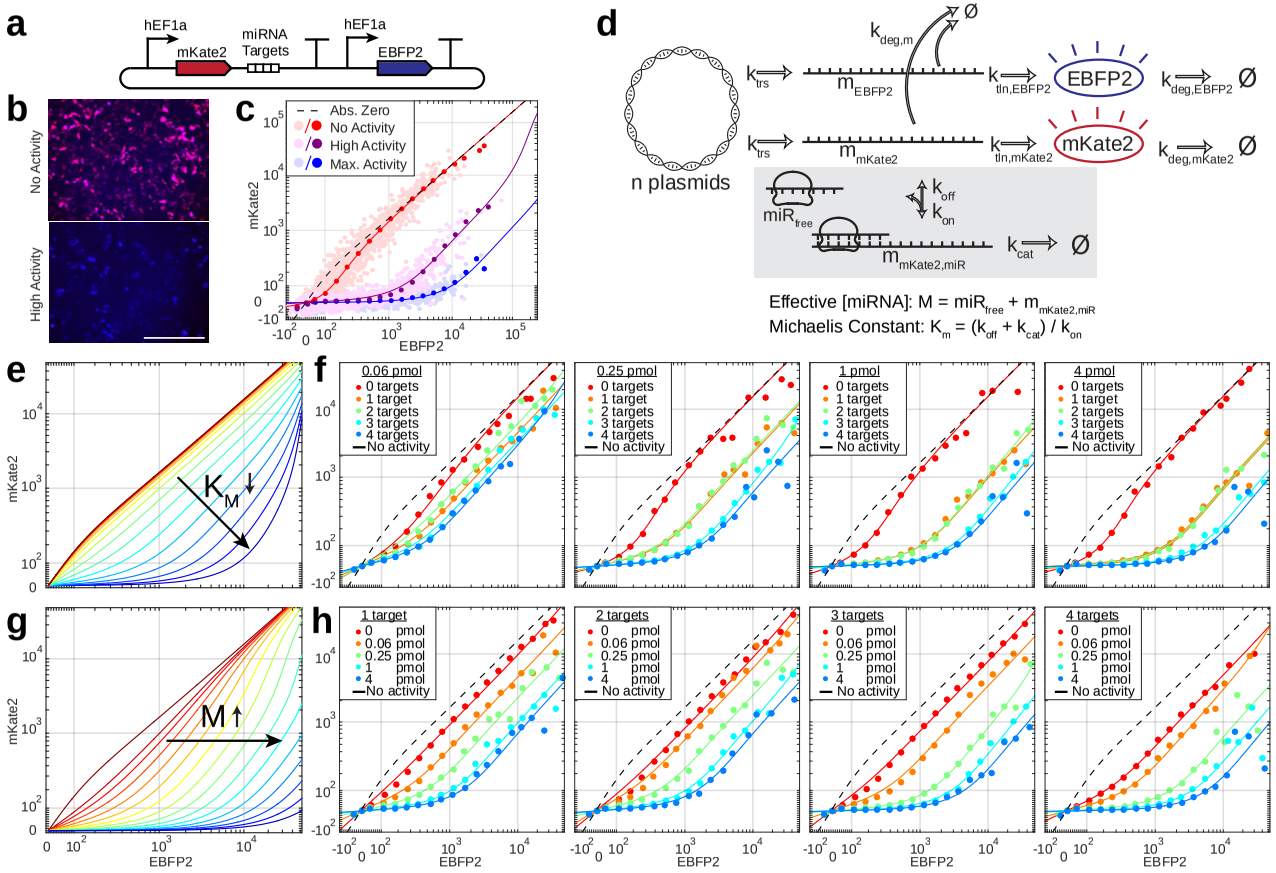

a
hEF1a
mKate2
miRNA Targets
hEF1a
EBFP2
b
No Activity
High Activity
c
Abs. Zero
No Activity
High Activity
Max. Activity
mKate2
EBFP2
d
n plasmids
EBFP2
mKate2
Effective [miRNA]: M = miR_free + m_mKate2,miR
Michaelis Constant: K_m = (k_off + k_cat) / k_on
e
mKate2
K_M
f
0.06 pmol
0.25 pmol
1 pmol
4 pmol
0 targets
1 target
2 targets
3 targets
4 targets
No activity
g
mKate2
M
EBFP2
h
1 target
2 targets
3 targets
4 targets
0 pmol
0.06 pmol
0.25 pmol
1 pmol
4 pmol
No activity
EBFP2

Figure 2-1: A biochemical model explains miRNA repression measured using miRNA sensors.
(a top) A schematic of the single-input miRNA low sensors used in this study. Four repeats of perfectly complementary target sites are placed in the 3' UTR of the mKate2 fluorescent reporter. EBFP2 serves as a transfection marker. **(a left)** Fluorescence microscopy in HEK293FT cells for a control sensor with miRNA target sites with no activity (FF4) compared to a sensor with target sites for a high activity miRNA (miR-106a-3p). Cells expressed both mKate2 and EBFP2 with the no miRNA activity sensor, while mKate2 fluorescence was greatly reduced in the presence of a high activity miRNA. Both images are merged blue and red channels. **(a right)** Corresponding flow cytometry data in HEK293FT cells for no activity (red) and high activity (purple) sensors. Data for a positive control using exogenous siRNA to repress mKate2 is shown for reference (blue). Light scatter points correspond to data; dark points represent median values for data binned by EBFP2 fluorescence; lines indicate fits to the data using a biochemical model; dashed line shows modeling results with absolutely zero miRNA activity. Fluorescence is shown in terms of arbitrary units and with logicle scaling for the axes[86]. **(b)** The single-input repression model annotated with relevant species and rate constants. Each cell is transfected with some number of plasmids (N) which are transcribed (at rate k_{trs}) and translated (at rates $k_{tln,EBFP2}$ and $k_{tln,mKate2}$) to yield fluorescent proteins. mKate2 transcripts may be reversibly bound by a miRNA-containing complex (miR_{free}) forming a bound species ($m_{mKate2,miR}$). When bound, mKate2 transcripts can be catalytically degraded at some rate (k_{cat}). Two parameters, effective miRNA concentration (M) and Michaelis constant (K_m) fully characterize miRNA activity. RNA and protein species are non-specifically degraded at a fixed rate. All rate constants except for M and K_m are approximated from literature values (Table 2.3). **(c left)** Model predictions for decreasing K_m while holding M constant. As K_m decreases, repression increases while switching threshold is maintained. **(c right)** Experimental perturbation of K_m. HEK293FT cells, which exhibit low endogenous miR-21-5p activity, were transfected with sensors containing varying numbers of target sites for miR-21-5p and varying amounts of exogenous miR-21-5p mimic. As the number of target sites was increased (decreasing K_m), repression increased as predicted. **(d left)** Model predictions for increasing M while holding K_m constant. As M increases, the switching threshold increases and repression increases as a result. **(d right)** Experimental perturbation of M. As miR-21-5p mimic concentration is increased, repression increased as predicted.

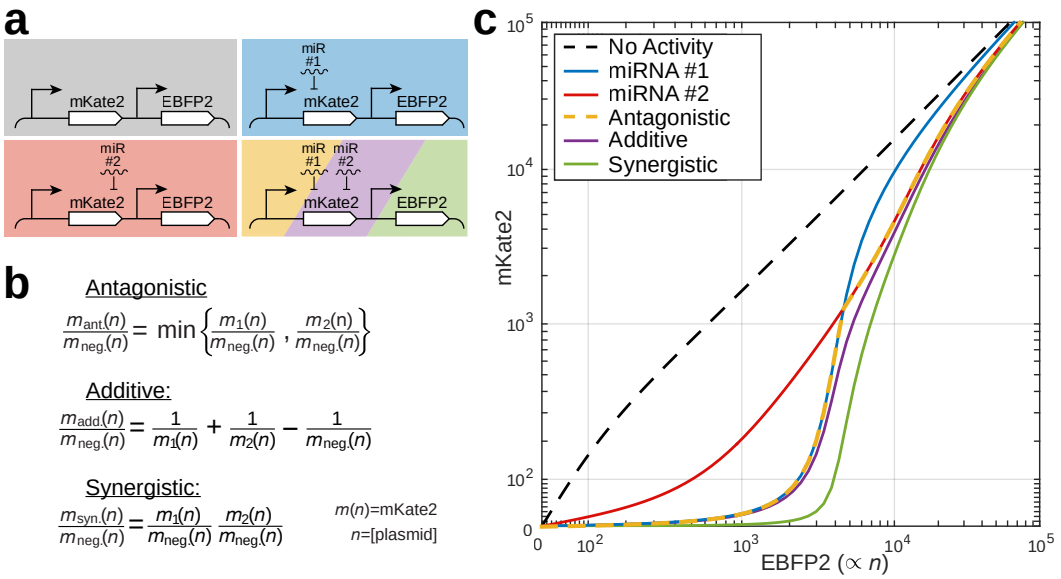


Figure 2-2: Methods for combining miRNA activities in predictions of multi-input low sensors.
(a) Schematics for 0-input control (gray), 1-input (blue and red), and 2-input miRNA sensors (yellow/purple/green stripe) single and two-input sensors The 1-input miRNA sensors are first characterized, and the characterization used to predict the behavior of multi-input sensors. **(b)** Equations based on the Chou-Talalay method adapted here for predicting miRNA activity of multi-input sensors according to antagonistic, additive, and synergistic interactions. Predictions of free mKate2 mRNA concentration for the three interaction types ($m_{ant.}(n)$, $m_{add.}(n)$, and $m_{syn.}(n)$) are computed based on the characterized output of k different miRNAs ($m_i(n)$). All mKate2 mRNA concentration predictions are functions of the number of plasmids (n) in the cell, which is proportional to EBFP2 fluorescence. $m_{neg.}(n)$ refers to the output of the negative control where no miRNA target sites are present. **(c)** Calculation of 2-input predictions from single-input activity. Colors for the curves correspond with those from (a). The repression curve for the negative control construct without target sites (black) is shown for reference. miRNA repression curves are illustrated for different single miRNAs (blue and red), which are generated by fitting M and K_m to sensor data using the single-input model illustrated in Figure 2-1. The blue and red single-input repression curves intersect because in this example miRNA #1 exhibits lower M and K_m but higher M/K_m compared to miRNA #2, leading to greater repression at lower EBFP2 expression levels and vice versa. Antagonistic predictions (yellow) closely follow the miRNA

with highest activity (minimal mKate2 expression) for a given level of EBFP2 expression, additive predictions (purple) are similar but with increased overall activity when the two constituent miRNAs have similar individual activity, and synergistic predictions (green) always exhibit higher miRNA activity than either antagonistic or additive predictions. Generally, discrepancy between the two types of predictions varies depending on the number and miRNA activities of the single-inputs, with greater discrepancies when many similar activity single-inputs are combined.

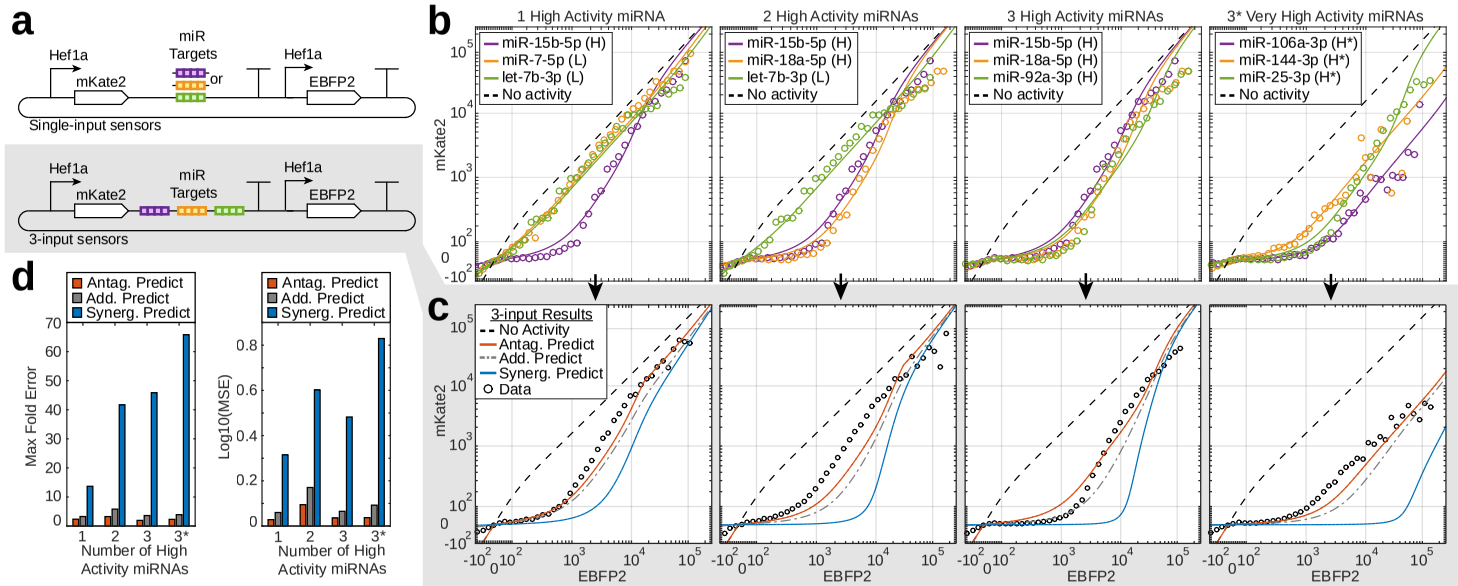

a
Hef1a
mKate2
miR
Targets
or
Hef1a
EBFP2
Single-input sensors
Hef1a
mKate2
miR
Targets
Hef1a
EBFP2
3-input sensors
b
1 High Activity miRNA
2 High Activity miRNAs
3 High Activity miRNAs
3* Very High Activity miRNAs
miR-15b-5p (H)
miR-7-5p (L)
let-7b-3p (L)
miR-18a-5p (H)
miR-92a-3p (H)
miR-106a-3p (H*)
miR-144-3p (H*)
miR-25-3p (H*)
No activity
mKate2
c
3-input Results
No Activity
Antag. Predict
Add. Predict
Synerg. Predict
Data
EBFP2
d
Antag. Predict
Add. Predict
Synerg. Predict
Max Fold Error
Log10(MSE)
Number of High Activity miRNAs

Figure 2-3: Interactions between miRNA target site sets in the 3' UTR sensors appear antagonistic.

(a) Schematic for single-input sensors bearing one set of target sites (top) and 3-input sensors bearing three different sets of target sites (bottom) where all miRNA target sites are located in the 3' UTR. **(b top)** miRNA repression data and model fits for different single-input sensors. Plots are ordered by increasing number of high activity miRNA target site sets. miRNA activities are denoted as low (L), high (H), or very high (H*). **(b bottom)** Comparison of predicted miRNA repression to data obtained for 3-input sensors. Predictions for 3-input sensors are computed based on M and K_m parameters measured from single-input sensors. Using the assumption that miRNA target sites act antagonistically yields predictions (red) which are equivalent to taking the maximum activity (minimum mKate2 expression) of the three single-input sensors for each EBFP2 expression level. Additive predictions (gray dashed) are made using the Chou-Talalay method. Synergistic predictions are made by multiplying fold repression from each of the single-input sensors within each EBFP2 transfection bin. Predictions using the antagonistic or additive models are markedly better than those from an synergistic model, indicating that miRNA repression for multi-input sensors is not the simple multiplicative effect. While antagonistic and additive models were close in these example, predictions can diverge drastically when several miRNA inputs are combined and when input activities are very similar. **(c)** Analysis of prediction errors for 3-input sensors. Errors are measured by computing the maximum fold difference between predictions and data across all bins of EBFP2 expression (max fold error) or by computing the mean squared error (MSE). For both metrics and for all tested combinations of low/high/very high (*) activity miRNAs, antagonistic interaction explains the data better than an additive or synergistic interaction.

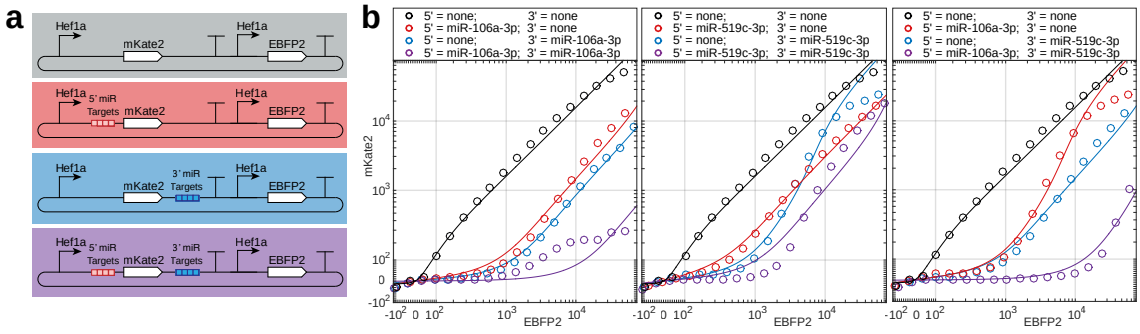
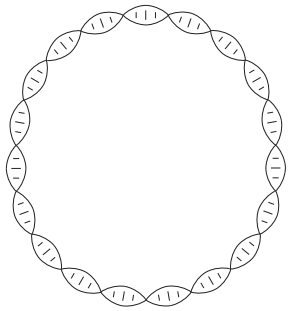


Figure 2-4: Positioning miRNA target site sets in separate UTRs yields synergistic interactions

(a) Diagram of tested constructs bearing target sites in 5' UTR only (red), 3' UTR only (blue), both 5' and 3' UTRs (purple), and no target sites (gray; shown as black in panel b). Constructs were designed to sense endogenous miRNAs with detectable activity in the tested cell line (HEK293FT). **(b)** miRNA repression data (circles) and model fits (lines) obtained for each of the four constructs with miR-106a-3p target sites, miR-519c-3p target sites, or a combination of the two. Constructs bearing target sites in both 5' and 3' UTRs showed greater repression than constructs with target sites in only a single UTR, demonstrating a synergistic effect not observed in constructs where target site sets were confined to the 3' UTR only.



This page intentionally left blank to keep captions near figures.

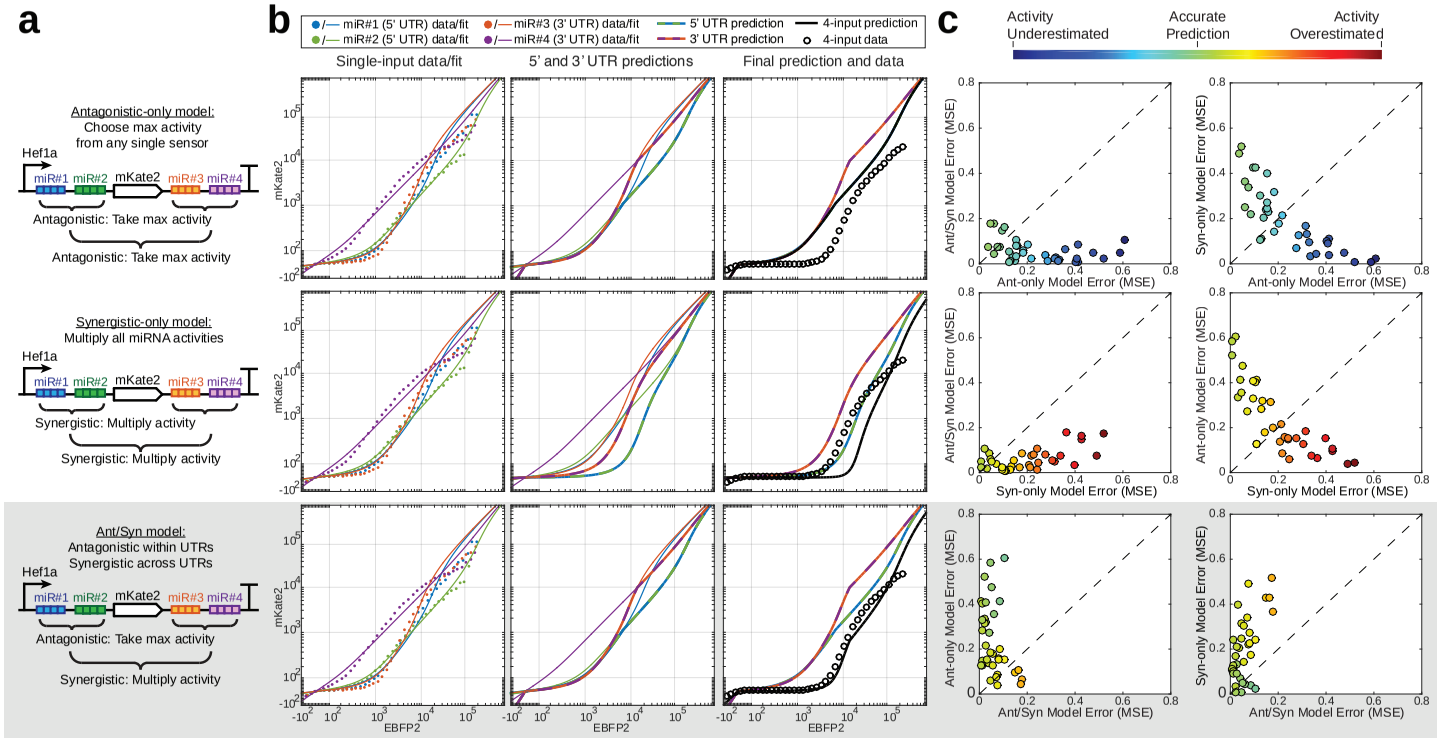

a
b
c
Antagonistic-only model:
Choose max activity
from any single sensor
Hef1a
miR#1 miR#2 mKate2 miR#3 miR#4
Antagonistic: Take max activity
Antagonistic: Take max activity
Synergistic-only model:
Multiply all miRNA activities
Synergistic: Multiply activity
Synergistic: Multiply activity
Ant/Syn model:
Antagonistic within UTRs
Synergistic across UTRs
miR#1 (5' UTR) data/fit
miR#2 (5' UTR) data/fit
miR#3 (3' UTR) data/fit
miR#4 (3' UTR) data/fit
5' UTR prediction
3' UTR prediction
4-input prediction
4-input data
Single-input data/fit
5' and 3' UTR predictions
Final prediction and data
mKate2
EBFP2
Activity
Underestimated
Accurate
Prediction
Activity
Overestimated
Ant/Syn Model Error (MSE)
Ant-only Model Error (MSE)
Syn-only Model Error (MSE)

Figure 2-5: Ant/Syn model provides accurate predictions of 4-input miRNA classifiers.

(a) Diagrams for 4-input sensors with associated logic for combining miRNA activity annotated for the antagonistic-only model (top row), synergistic-only model (middle row), and Ant/Syn model (bottom row). Two sets of distinct miRNA target sites were placed in each 5' and 3' UTR. **(b)** Example 4-input predictions based on single-input data. Rows across panels depict analysis of the sensor configuration in panel a. Single-input miRNA sensor data and model fits are shown for each of the four miRNAs (left column). Predictions are made separately for the 5' and 3' UTRs (center column) by taking the maximum activity for any single sensor (antagonistic-only model and Ant/Syn model) or alternatively by multiplying fold changes (synergistic-only model) in each transfection marker bin. Final predictions are made (right column) by combining activities from 5' and 3' UTRs by multiplying fold changes for each UTR (synergistic-only model and Ant/Syn model) or by taking maximum activity (antagonistic-only model). For this example, miR#1 = miR-17-5p, miR#2 = miR-31-3p, miR#3 = miR-519a-3p, and miR#4 = miR-16-5p. All 36 miRNA target combinations tested with associated predictions and data can be found in Figures 2-22 through 2-25. **(c)** Comparison of error in antagonistic-only, synergistic-only, and Ant/Syn model predictions for 36 different classifiers. Errors calculated as mean squared error between prediction and obtained data are plotted for each of the three possible comparisons. Additionally, points are colored by the degree of miRNA activity underestimation or overestimation according to the model plotted along the x axis (i.e. top row colored by antagonistic-only model errors, middle row by synergistic-only model error, bottom row by Ant/Syn model errors), resulting in six total graphs. For all 36 tested combinations of miRNAs, errors from the Ant/Syn model were similar or better than those from the other two models, as shown by most points falling within upper-left triangle. In comparison, the antagonistic-only model tended to underestimate repression and the synergistic-only model tended to overestimate repression.

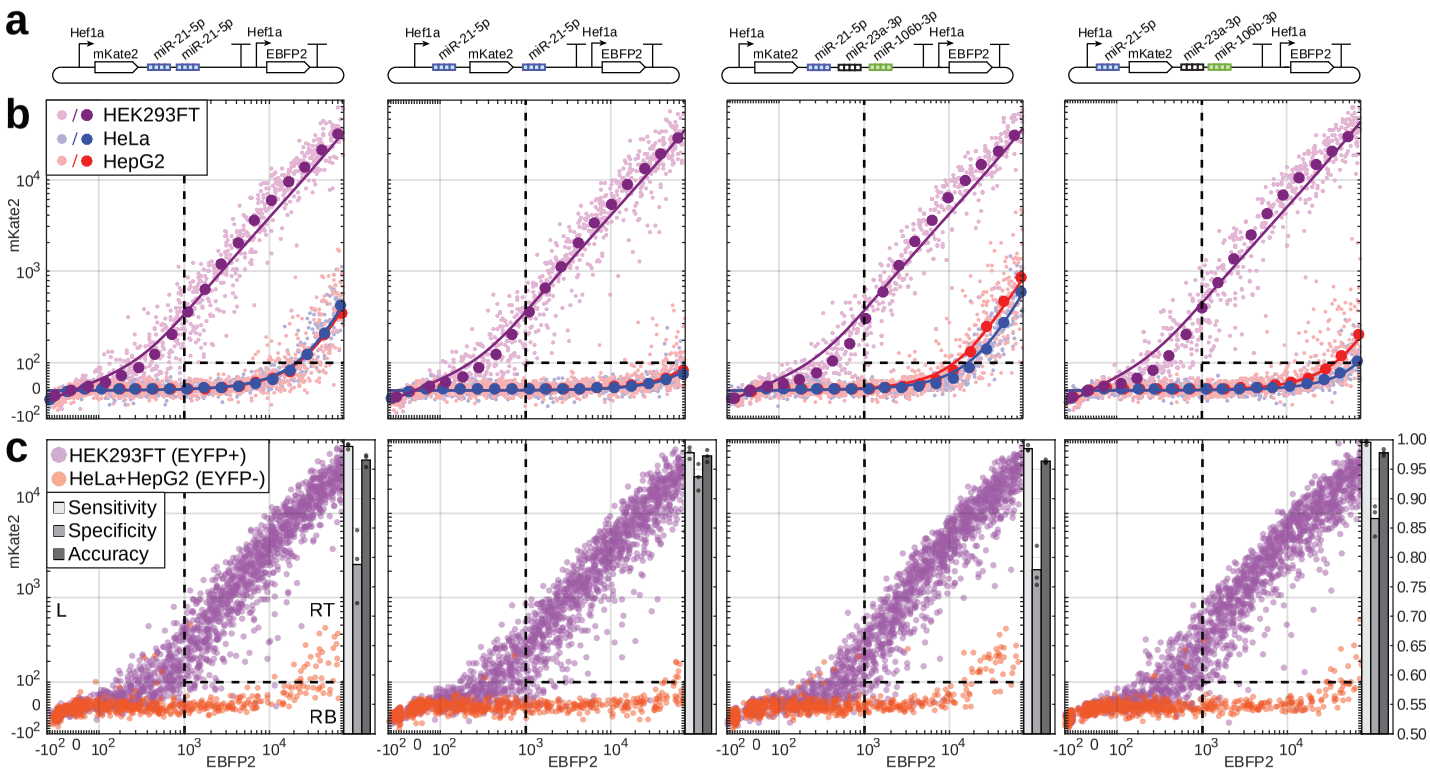

a
Hef1a
mKate2
miR-21-5p
miR-21-5p
Hef1a
EBFP2
Hef1a
miR-21-5p
mKate2
miR-21-5p
Hef1a
EBFP2
Hef1a
mKate2
miR-21-5p
miR-23a-3p
miR-106b-3p
Hef1a
EBFP2
Hef1a
miR-21-5p
mKate2
miR-23a-3p
miR-106b-3p
Hef1a
EBFP2
b
HEK293FT
HeLa
HepG2
mKate2
c
HEK293FT (EYFP+)
HeLa+HepG2 (EYFP-)
Sensitivity
Specificity
Accuracy
L
RT
RB
EBFP2

Figure 2-6: miRNA target site position affects cell classifier performance substantially.
(a) Circuit diagrams for miRNA classifiers to distinguish HEK293FT cells from HeLa and HepG2. Two pairs of classifiers were tested, some encoding miRNA targets only in the 3' UTR and others with one set of miR-21-5p targets moved to the 5' UTR. miRNAs exhibiting high activity in HeLa and HepG2 cells but not HEK293FT were selected for inclusion. **(b)** Fluorescence observed after separate transfections of the corresponding classifiers from (a) into HEK293FT, HeLa, and HepG2 cells. Expression of mKate2 reporter remained high in HEK293FT cells for all classifiers, but knockdown of mKate2 in HeLa and HepG2 was enhanced only when miR-21-5p targets were placed in the 5' UTR (columns 2 and 4) compared to the 3' UTR (columns 1 and 3), likely due to the synergistic interactions obtained from miRNA targets in different UTRs. Subsampled raw data are indicated as light points and binned data/fits are indicated as dark points/lines. Subsampling was performed to normalize the number of cells within each EBFP2 expression level bin in order to minimize effects of transfection efficiency on sensitivity and specificity measurements. **(c)** Classification of three cells lines in co-culture format. HEK293FT cells with genomically integrated constitutive expression of EYFP were co-cultured with HeLa and HepG2. Transfections were performed in cell mixtures for each classifier. After flow cytometry, gating for EYFP+ and EYFP- cells was used to determine whether each cell was HEK293FT or HeLa/HepG2 in origin. Subsampled data is shown along with dotted lines that demarcate three regions indicating fluorescent value ranges used for determining whether cells were classified as HEK293FT (RT: right top), HeLa/HepG2 (RB: right bottom), or undetermined due to low transfection levels (L: left). Sensitivity or true positive rate is calculated using $TP/(TP+FN) = [RT^+]/([RT^+]+[RB^+])$, specificity or true negative rate is calculated using $TN/(TN+FP) = [RB^-]/([RB^-]+[RT^-])$, and accuracy is calculated using $(TP+TN)/(TP+FP+FN+TN) = ([RT^+]+[RB^-])/([RT^+]+[RT^-]+[RB^+]+[RB^-])$ where TP, FN, TN, and FP denote the number of cells that are true positive, false negative, true negative, and false positive respectively while $^+$ and $^-$ denote EYFP+ and EYFP- respectively. Classifiers with separate target sites in 5' and 3' UTR were better able to distinguish different cell types by reducing off-target expression. Quantification of sensitivity and specificity shows that while sensitivity and accuracy remained high for all classifiers, greater specificity (>85%) was obtained only with classifiers bearing miRNA target sites in both UTRs.

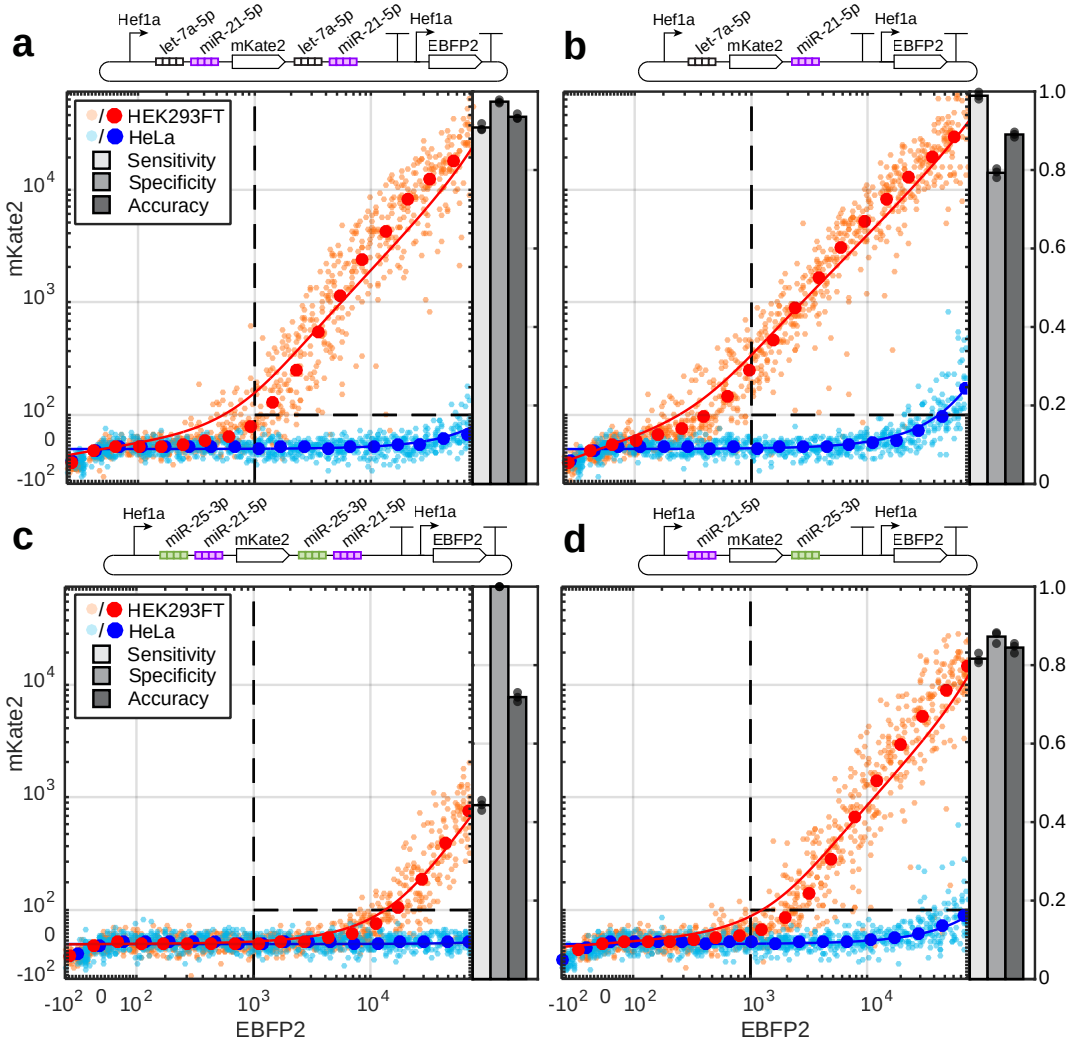


Figure 2-7: Ant/Syn model predicts the effects of target site position and number on the trade-off between sensitivity and specificity.
Several classifiers were tested in order to further explore the effect of target set position and duplication on classifier sensitivity, specificity, and accuracy. We chose to sense three different miRNAs: miR-21-5p and let-7a-5p have high activity in HeLa and slight activity in HEK293FT, miR-25-3p has medium activity in HeLa and HEK293FT. In this example, miR-25-3p would be important for classifying against a third cell type not shown here (e.g. glioma[87]). **(a-b)** To test the effects of miRNA target set duplication on classification, versions of 2-input classifiers were built containing sets of target sites for let-7a-5p and miR-21-5p together in both UTRs for

a total of four target sets (a), or in separate UTRs for a total of two target sets (b). Output in the classifier with four target sets was dominated by synergy between miR-21-5p activity from both UTRs, resulting in very low output in HeLa (which exhibits high miR-21-5p activity) and high specificity (97%). However, this high specificity required a slight trade-off in sensitivity (91%) since output in HEK293FT (which exhibits slight miR-21-5p activity) was reduced with target site duplication. In contrast, high sensitivity (99%) but lower specificity (79%) was obtained with the classifier with only a single set of each miRNA target (b) since output in HEK293FT cells was attenuated by only one miR-21-5p target set. **(c-d)** Classifiers similar to (a-b) were built containing sets of target sites for miR-21-5p and miR-25-3p with duplication (c) and without (d). Again duplication resulted in high specificity (100%) but with a significant trade-off in sensitivity (44%) due to synergistic effects from both sets of miR-25-3p targets, which allows two medium activity target sets to exhibit high activity when combined in HEK293FT cells. Also, miR-25-3p targets contain an ATG sequence resulting in further reduction in output. In contrast, the classifier with only two target sets avoided multiplicative effects from miR-25-3p and introduction of uORFs. These effects result in greatly improved sensitivity (82%) with the classifier containing only two target site sets, at a cost of slightly reduced specificity (87%). The latter classifier also showed higher overall accuracy (84% compared to 72%).

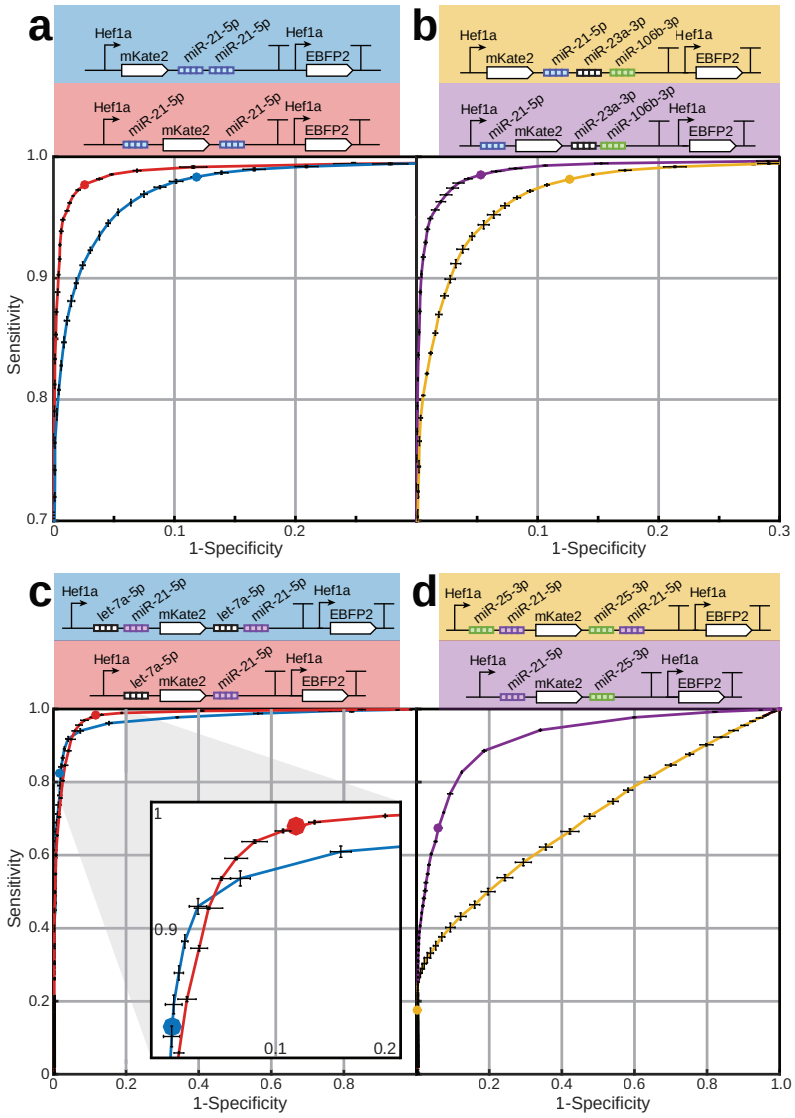


Figure 2-8: ROC curves for classifiers illustrate specificity and sensitivity tuning by miRNA target set location and number.

Receiver operator characteristic (ROC) curves derived from data in Figs. 2-6 and 2-7 were generated by varying the mKate2 threshold used to classify HEK293FT cells and then calculating sensitivity and specificity at each threshold. For reference, indicated circles along the curves show sensitivity and specificity at the fixed threshold of 10^2 used for Figs. 2-6 and 2-7. **(a)** Circuit diagrams and ROC curves for miRNA classifiers with varying target site position as tested in Fig. 2-6. Circuit shading

color corresponds to ROC curve color. The classifier with miR-21-5p target sets in different UTRs (red) shows greater maximal accuracy compared to that with target sites in only the 3' UTR (blue) as indicated by nearer approach towards the upper left and greater area under the curve (AUC = 0.9947 vs 0.9895). At the example 10^2 mKate2 threshold, both circuits prioritized sensitivity somewhat at the expense of some specificity. **(b)** Similarly, the 3-input classifier with miR-21-5p in the 5' UTR and miR-23a-3p and miR-106b-3p in the 3' UTR (purple) exhibits better overall classification than the variant with target sets in only the 3' UTR (yellow), with an AUC of 0.9952 and 0.9881 respectively. Again, at the 10^2 mKate2 threshold circuits prioritized sensitivity over specificity. **(c)** Circuit diagrams and ROC curves for miRNA classifiers with or without target site duplication across UTRs for let-7a-5p and miR-21-5p. Data for (c) and (d) corresponds to that from Fig. 2-7. Both classifier variants show similar maximum accuracy when classifying between HEK293FT and HeLa cells, though at the example threshold of 10^2 the variant without duplication (red, AUC = 0.9833) prioritizes sensitivity while the variant with target set duplication (blue, AUC = 0.9750) prioritizes specificity as shown previously. See inset for magnified view of region where sensitivity and specificity are tuned. **(d)** For classifiers with target sites for miR-25-3p and miR-21-5p, the circuit without duplication (purple, AUC = 0.9235) exhibits better classification compared to that with duplication (yellow, AUC = 0.7018), since repression was determined by both miR-25-3p and miR-21-5p rather than being dominated by miR-21-5p and also since potential uORFs are avoided by removing miR-25-3p targets from the 5' UTR (both effects enhance sensitivity).

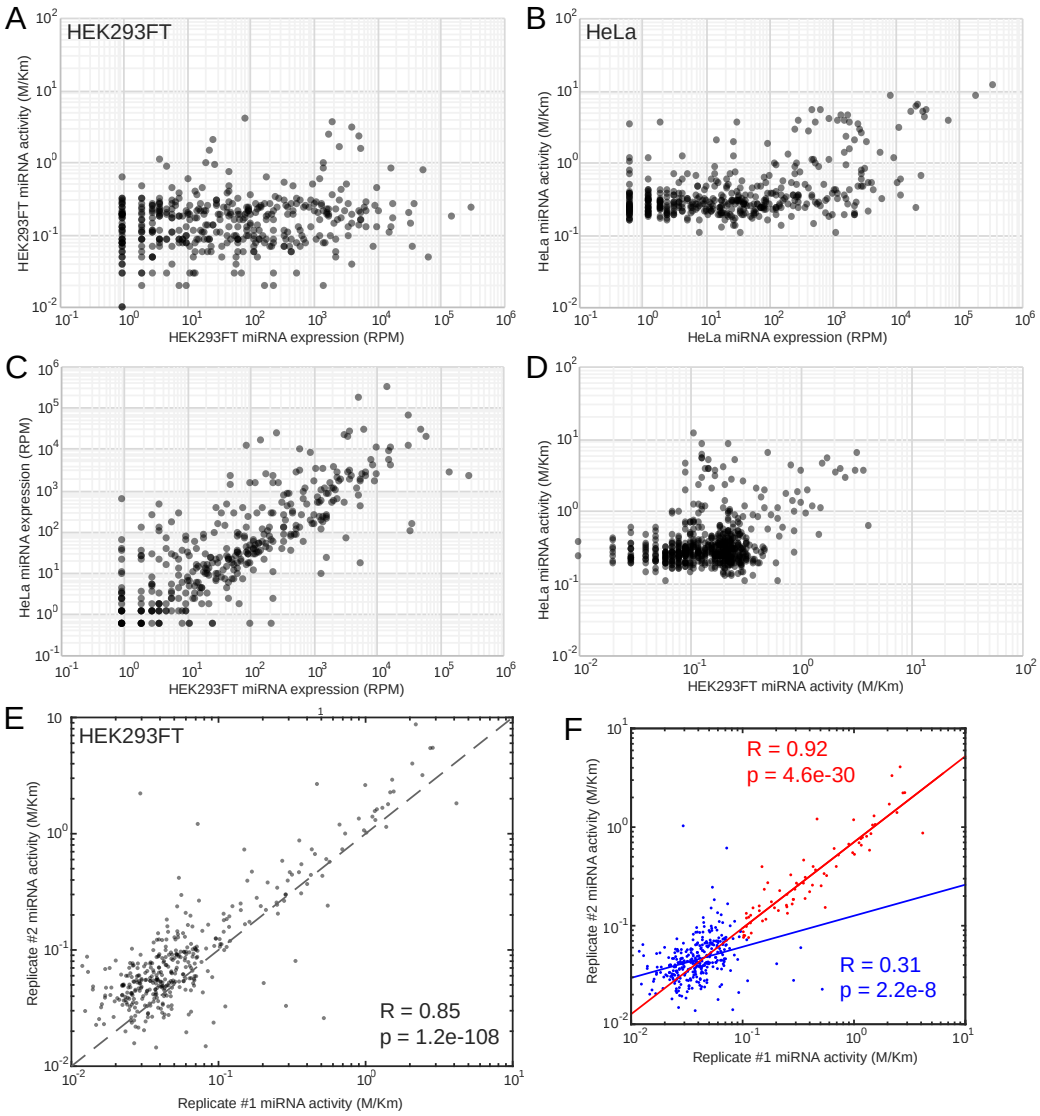


Figure 2-9: Low sensor library data in HEK293FT and HeLa cells. **(A,B)** A miRNA activity metric of M/K_M derived from parameter fits is plotted against miRNA expression data in reads per million (RPM) obtained by high throughput sequencing. Similarly to other results comparing miRNA activity and abundance, a poor correlation between the two measures is observed for both cell lines tested here. While several hypotheses exist for the discrepancy (ceRNA hypothesis, cellular localization, miRNA modifications) the extent of contributions from each of these and other possible effects is yet to be determined. **(C)** miRNA expression for

the two cell lines are plotted against each other. Many miRNAs are differentially expressed between the two lines. **(D)** miRNA activity observed using our sensors in HEK293FT and HeLa are plotted. Several miRNAs show greater activity in HeLa vs HEK293FT (HeLa-specific) and a few show greater activity in HEK293FT vs HeLa (HEK293FT-specific). Note the difference between expression data in (C) vs activity data in (D) **(E)** The reproducibility of the M/K_M metric was tested by repeating two biological replicates of reverse transfection in HEK293FT cells and parameter fitting of the low sensor library. The M/K_M metric appears reproducible over the two replicates suggesting M/K_M may be used as a proxy for miRNA activity if a single measure for activity is required. High throughput sequencing was performed using the NEBNext small RNA kit according to the manufacturer's protocol and sequencing was performed using an Illumina HiSeq 2000. Analysis was performed using cutadapt to trim the 5' SR adapters, fastq-multx to demultiplex, cutadapt to trim the 3' SR adapter and barcodes, followed by miRExpress[88] to quantify miRNA expression based on sequences in miRBase 21[3]. **(F)** Correlation coefficient and p-values were calculated for groups of either high activity (M/Km >0.1, red) and low activity (M/Km <0.1, blue) sensors in HEK293FT. Both groups show statistically significant correlation though there was a large reduction in correlation for low activity sensors, limiting reproducibility in low activity ranges. However, it should be noted that miRNA activities within this low regime (M/Km between 0.01 and 0.1) behave similarly with little detectable repression. Variance for low activity measurements may be due to noise in model fitting or amplified by the log transform.

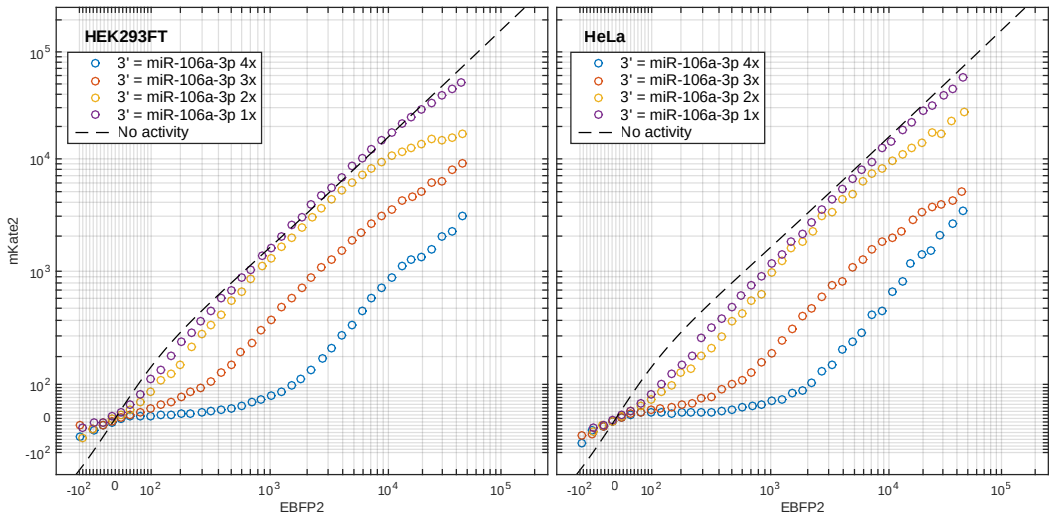
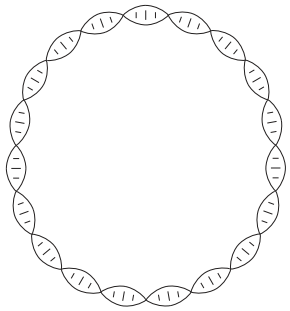


Figure 2-10: Enhanced repression by cooperation with up to four target sites.

In both HEK293FT and HeLa cells, combining up to four miR-106a-3p target sites together to form a set results in greater repression. However, further increase from 4x to 8x repeats does not result in further cooperative effects (see figure 2-11). Interestingly, 1x and 2x repeats show very little activity. These results suggest that the 4x repeats we used for our sensor library is a close to optimal balance that can achieve sufficient repression without incurring excessive DNA synthesis cost.



This page intentionally left blank to keep captions near figures.

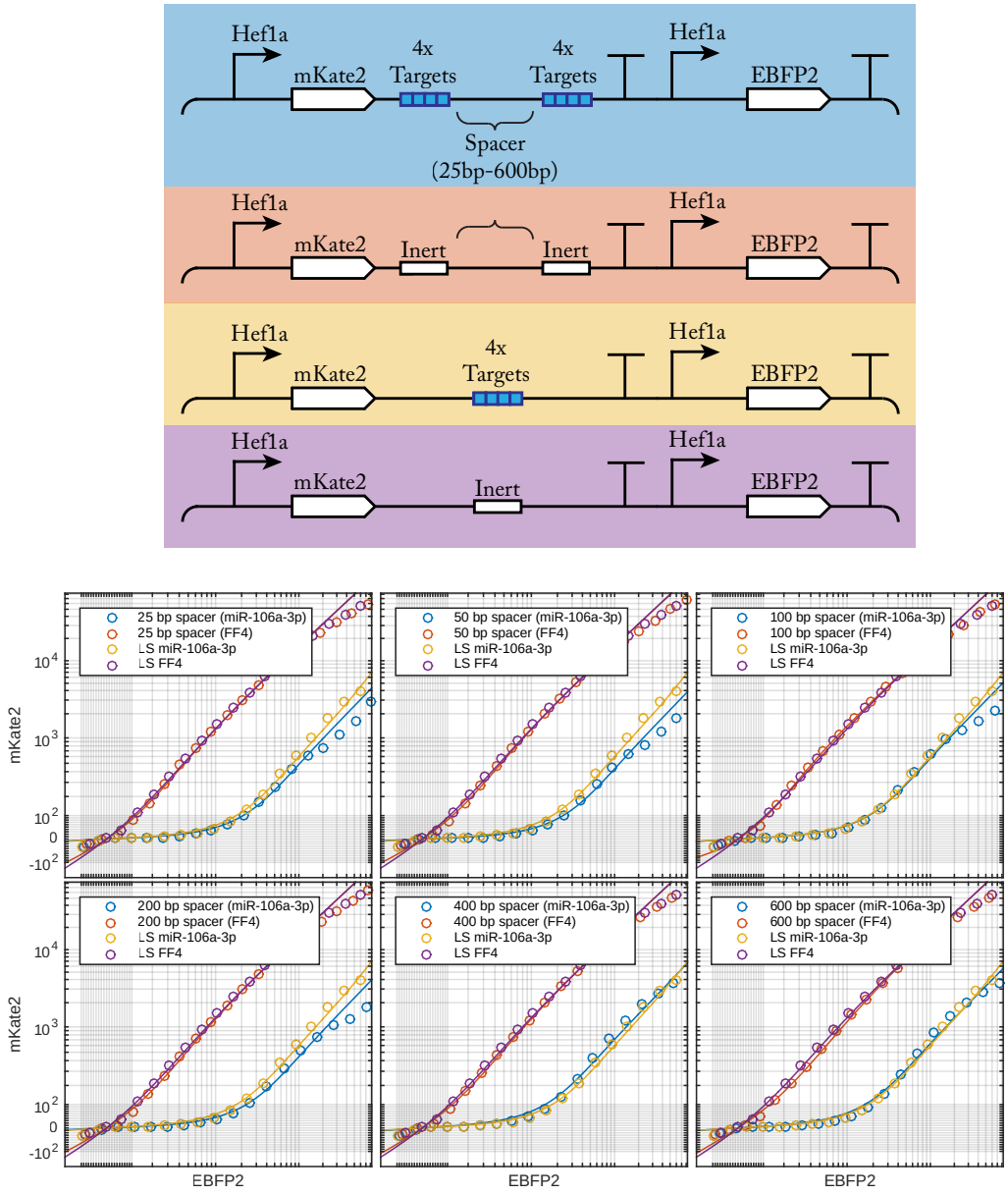


Figure 2-11: Effect of spacer length on miRNA low sensors.
To determine whether antagonistic miRNA activities within the 3' UTR are due to steric effects, several variants of a low sensor with miRNA target site sets spaced by different distances were built and tested. For all spacer lengths, sensors with two sets of target sites performed the same as a sensor with only a single set of target sites, indicating almost complete antagonistic miRNA activity for spacers as long as

600 bp - approximately the same length that separates target sites in sensors with target sets across UTRs which show synergistic activity (figure 2-3). This result rules out steric or length-dependent effects for observed antagonism. Assembly was performed using LSBr as backbone and LSB/JG107 providing miRNA target sites, with JG106 providing different length spacers derived from mKate2 coding sequence. Transfections were performed in HEK293FT cells which possess high endogenous miR-106a-3p activity

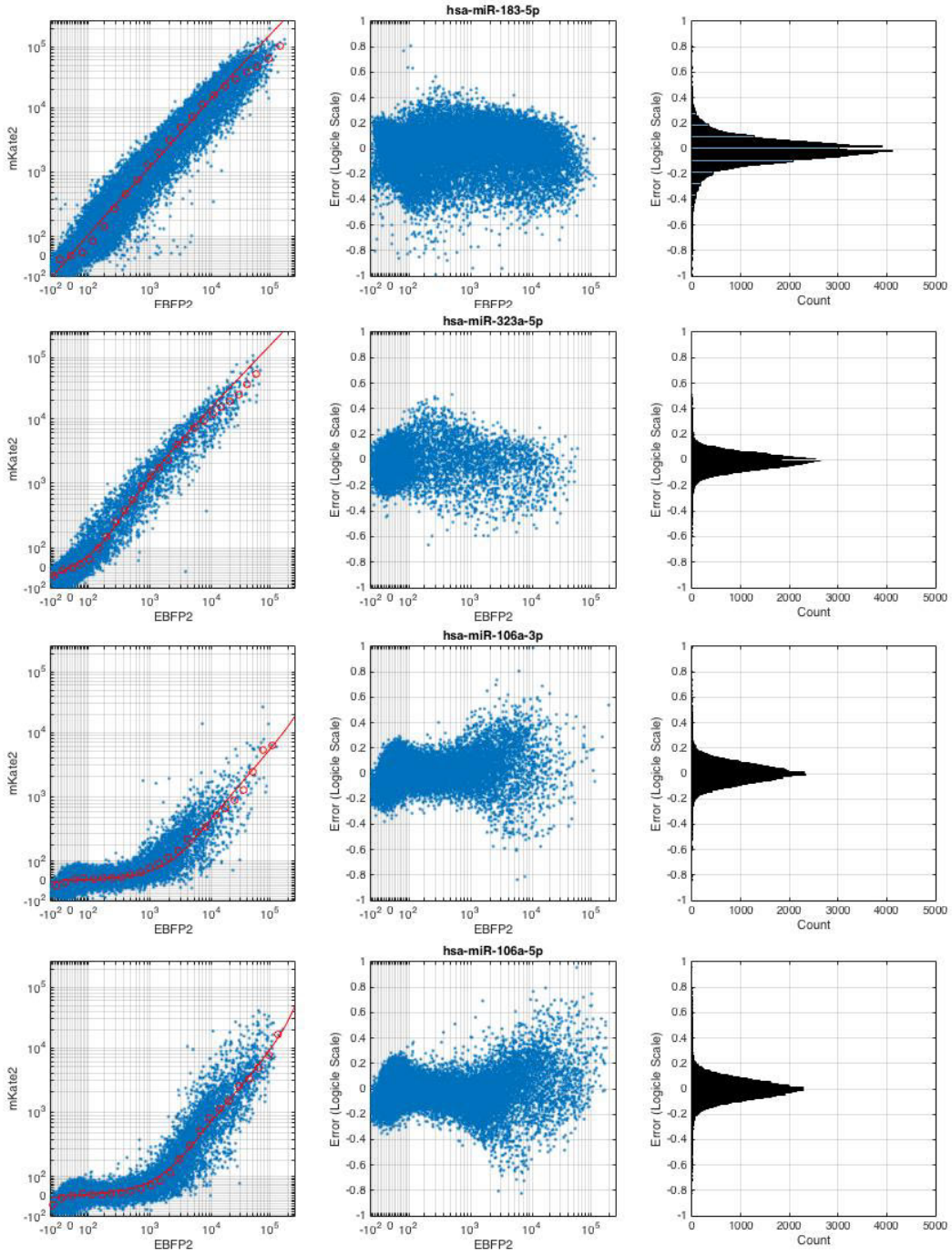

hsa-miR-183-5p
hsa-miR-323a-5p
hsa-miR-106a-3p
hsa-miR-106a-5p
mKate2
EBFP2
Error (Logicle Scale)
Count

Figure 2-12: Goodness of fit for the miRNA repression model.
After transfection of miRNA low sensor libraries into HEK293FT cells, we sought to determine whether observed miRNA activities could be adequately explained by a simple repression model. After fitting M and K_M to the data, residuals and a histogram of the residuals was plotted for several miRNAs with different activities. Residuals are relatively constant across EBFP2 expression levels and centered about zero, with slightly increased variance at high transfection levels. While minor deviations were found in some cases, most of the data is captured by the model.

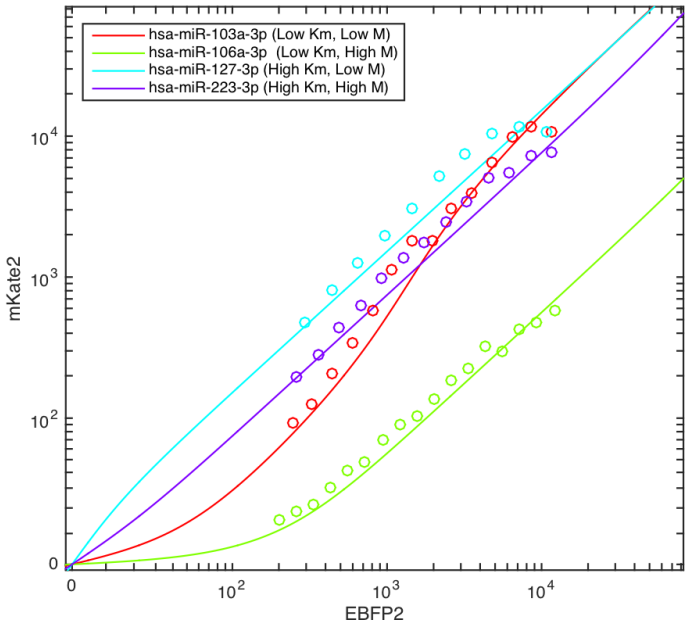
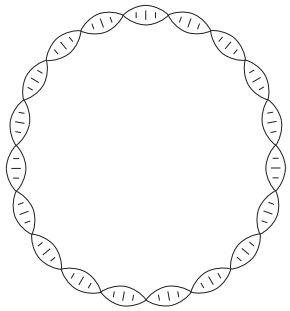
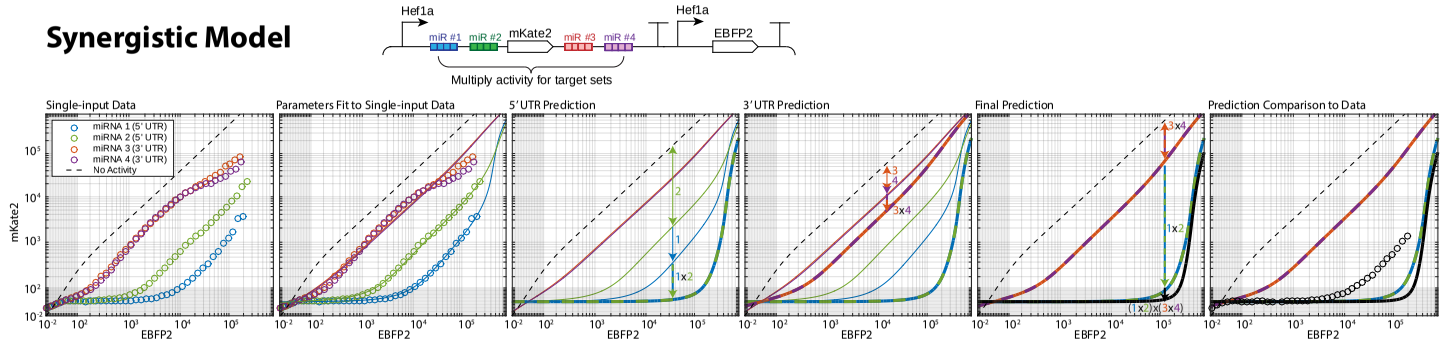


Figure 2-13: The model explains different miRNA repression curve shapes Four miRNA sensors demonstrating distinct miRNA repression curves are plotted post transfection into HEK293FT cells and after inning analysis. These shapes can be explained by different combinations of M and K_M. For miR-103a-3p, the threshold between repressed/unrepressed regions is low, indicative of a low M, while magnitude of repression is relatively high, indicative of a low K_M. For miR-106a-3p, K_M is similarly low since greater repression is observed, but the threshold region is shifted far to the right resulting in high M. For miR-127-3p, no apparent activity is observed which is reflect in both a high K_M and low M. For miR-223-3p, a slight repression across all transfection levels indicates that the threshold is far to the right (high M) but repression magnitude is low (low K_M).



This page intentionally left blank to keep captions near figures.


Synergistic Model
Hef1a
miR #1
miR #2
mKate2
miR #3
miR #4
Hef1a
EBFP2
Multiply activity for target sets
Single-input Data
Parameters Fit to Single-input Data
5'UTR Prediction
3'UTR Prediction
Final Prediction
Prediction Comparison to Data
miRNA 1 (5' UTR)
miRNA 2 (5' UTR)
miRNA 3 (3' UTR)
miRNA 4 (3' UTR)
No Activity
mKate2
EBFP2
3x4
1x2
(1x2)x(3x4)

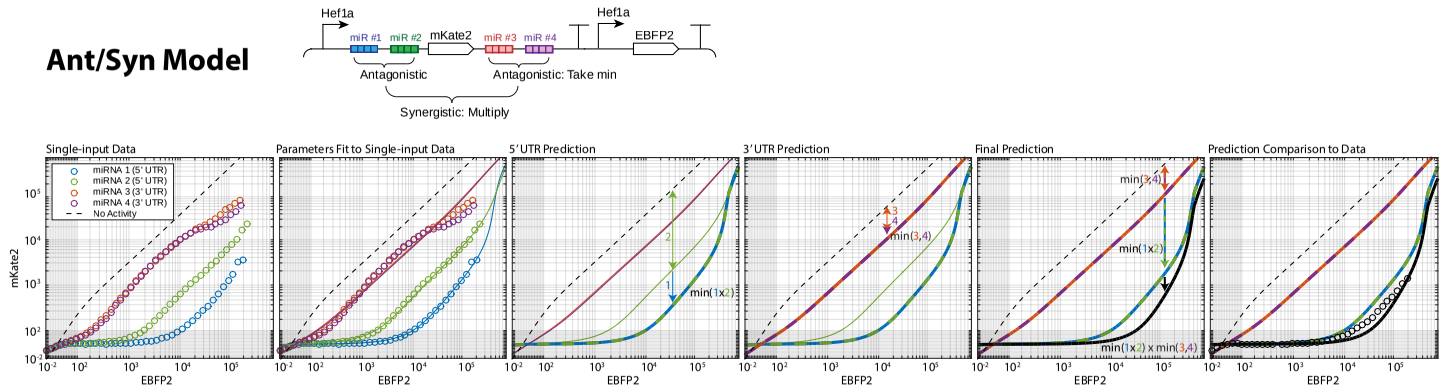

Ant/Syn Model
Hef1a
miR #1
miR #2
mKate2
miR #3
miR #4
Hef1a
EBFP2
Antagonistic
Antagonistic: Take min
Synergistic: Multiply
Single-input Data
Parameters Fit to Single-input Data
5'UTR Prediction
3'UTR Prediction
Final Prediction
Prediction Comparison to Data
miRNA 1 (5' UTR)
miRNA 2 (5' UTR)
miRNA 3 (3' UTR)
miRNA 4 (3' UTR)
No Activity
mKate2
EBFP2
min(1x2)
min(3,4)
min(1x2) x min(3,4)

Figure 2-14: Workflow for predictions based on a synergistic-only repression model vs the Ant/Syn model.
An illustration of the mathematical operations used to make predictions for synergistic-only vs Ant/Syn models is shown. In this example for HEK293FT cells, miR #1 = hsa-miR-363-3p, miR #2 = hsa-miR-196a-5p, miR #3 = hsa-miR-33b-5p, miR #4 = hsa-miR-340-5p. 1) Single input data are binned according to EBFP2 fluorescence and medians taken. 2) Parameters are fit to the binned data, generating the basis for all predictions using those miRNAs. 3) The 5' UTR prediction is calculated by multiplying the fold changes from miR #1 and miR #2 from the no activity reference point for each EBFP2 fluorescence level (synergistic model) or by taking the minimum mKate2 fluorescence within miR #1 or miR #2 for each EBFP2 level (Ant/Syn model). 4) 3' UTR prediction is made similarly to the 5' UTR prediction except using miR #3 and miR #4. 5) The final prediction is determined by multiplying fold changes from the two separate UTR predictions again relative to the no activity reference. 6) Data observed by transfecting the 4-input sensor construct into cells is compared to the prediction made only from single-input sensor data.

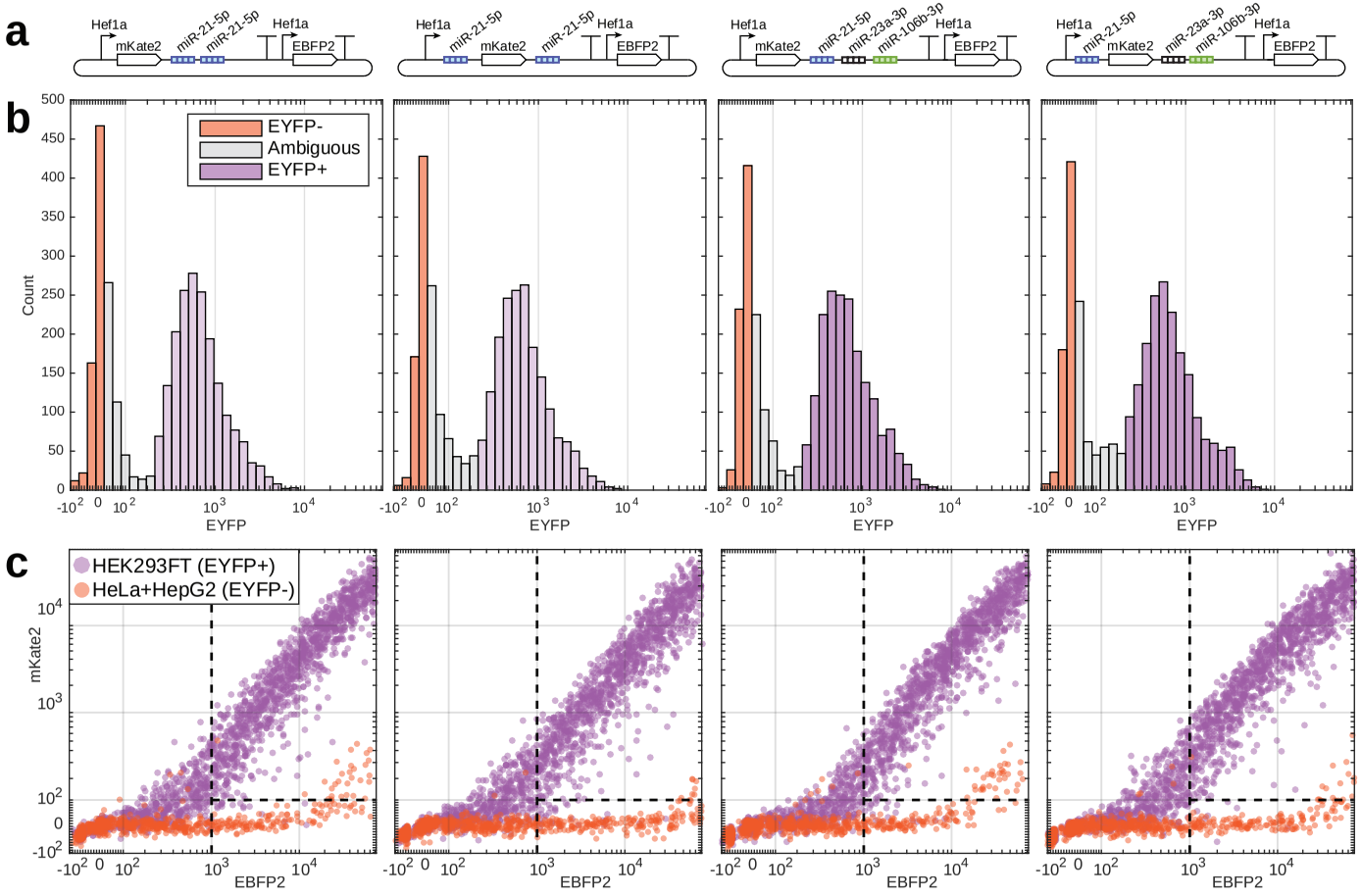

a
Hef1a
mKate2
miR-21-5p
miR-21-5p
Hef1a
EBFP2
Hef1a
miR-21-5p
mKate2
miR-21-5p
Hef1a
EBFP2
Hef1a
miR-21-5p
miR-23a-3p
miR-106b-3p
Hef1a
mKate2
EBFP2
Hef1a
miR-21-5p
mKate2
miR-23a-3p
miR-106b-3p
Hef1a
EBFP2
b
EYFP-
Ambiguous
EYFP+
Count
EYFP
c
HEK293FT (EYFP+)
HeLa+HepG2 (EYFP-)
mKate2
EBFP2

Figure 2-15: Gating of EYFP+ cells (HEK293FT) from EYFP- cells (HeLa and HepG2).
We used HEK293FT cells expressing EYFP from the genome to determine whether cells in coculture were indeed HEK293FT or HeLa/HepG2. **(a)** The same sensors and classifiers from Figure 2-5 are shown. Data for each construct is shown in the column below the respective circuit diagram. **(b)** Histogram of EYFP for cocultured cells is illustrated. Cells that could be unambiguously assigned to EYFP+ and EYFP- populations are colored purple and red respectively. **(c)** Data from Figure 2-5 with the EYFP+ and EYFP- cells colored in the same fashion as (b).

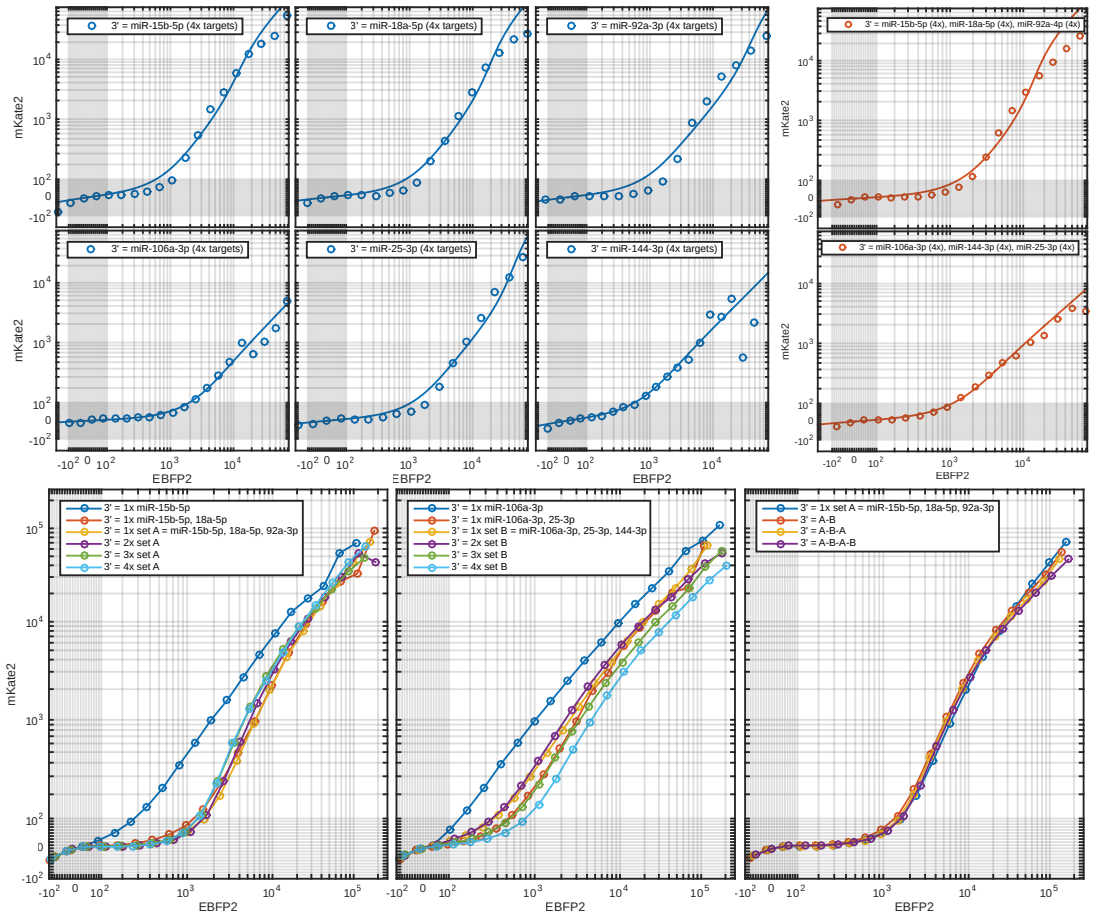


Figure 2-16: Antagonistic effects are maintained in miRNA sets containing distinct target sites.

To further explore the antagonistic activity contributions from sets of miRNA targets, we constructed miRNA sets composed of three different high activity miRNAs. Single input sensor data are plotted (upper left) and data for the corresponding 3-input classifier designs are plotted (upper right) for reference. We observed that concatenating up to four repeats of these target sets had negligible effect on observed activity for set A [miR-15b-5p, miR-18a-5p, miR-92a-3p] and minimal effect for set B [miR-106a-3p, miR-25-3p, miR-144-3p] (bottom). We also tested single repeats for sub sets of miRNAs in sets A and B. We found that at least two target sites were required to see significant repression. A minimal trend towards increased activity with increasing numbers of target sets exists for set B likely due to slightly mismatched activity, as miR-106a-3p has slightly higher activity than either miR-25-3p or miR-144-3p. The maintenance of antagonistic effects even when miRNA targets

are different from each other suggests that possible mechanism for antagonism could be strong cooperative effects between Ago molecules that is not dependent on miRNA sequence, which is saturable at high numbers of target sites. One possibility is Ago-TNRC6 complexes containing many Ago molecules allowing for cooperative binding of Agos bearing different miRNA. Further exploration into the extent of complexes containing multiple Agos may be warranted in future studies.

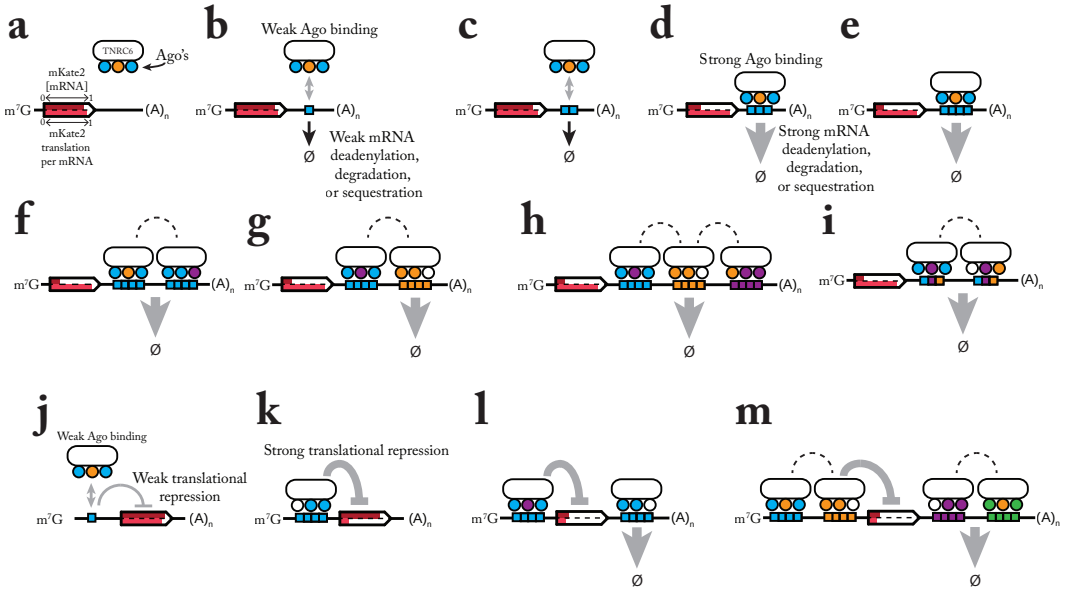


Figure 2-17: Speculative mechanistic model for miRNA repression observed in this study.

We present a speculative model that explains observations made in this study. Relative mKate reporter mRNA concentrations are shown above the coding sequence cartoon while relative translation levels per mRNA molecule are shown below. Arrow widths denote the amount of flux through a given step and gray arrows show the proposed rate limiting step for a given sensor. Argonaut molecules and miRNA target sites are colored according to miRNA loaded into the Ago molecule or bearing matching sequence to the target respectively. Dotted lines denote Agos that share the same repression machinery. We hypothesize that miRNA target sites within the 3' UTR mainly act to deadenylate, degrade, or sequester mRNA transcripts without affecting translation and vice versa for target sites within the 5' UTR. However the model does not strictly require assignment of these degradation pathways, only that repression mechanisms for the two UTRs be distinct enough that they do not significantly share resources and can separately saturated by addition of many miRNA target sites. Future models may swap or substitute repression mechanisms as further data becomes available. **a-e)** The model involves cooperative binding of Ago to miRNA target sites. This cooperativity is necessary to explain why little repression is observed for sensors bearing 1-2 miRNA target sites but significant repression for those bearing 3-4 or more targets (Fig. 2-10). The data suggest that at low numbers of targets, binding of Ago to reporter transcripts is the rate limiting step - resulting in low numbers of transcripts that can be deadenylated/degraded/sequestered. **f-i)**

Addition of up to 8 of the same miRNA target site or 4 each two or three different miRNA target sites in blocks or interleaved fashion all resulted in similar repression to sensors with only 4 strong miRNA targets (Fig. 2-2, 2-11, 2-16). These data suggest that inclusion of target sites above ∼4 repeats results in saturation of repression machinery and that degradation is not likely to be simple cleavage mediated by Ago2 as this would be unlikely to result in saturation at such low numbers of target sites. Instead, the data suggest that Ago binds to other factors that repress the transcript and these other factors can be saturated. **j-l)** We observed that miRNA targets within the 5' UTR repress independently (i.e. synergistically) targets within the 3' UTR. We propose that a separate repression mechanism exists for 5' UTR miRNA targets, here we show one possibility: translational repression either at initiation or elongation. With four miRNA targets within the 5' strong repression is observed and with the same set of 4x miRNA targets within the 5' and 3' UTRs a combined effect on repression is observed (Fig. 2-3). In our speculative model, the combined effect is a result of both reduced mKate2 mRNA concentration and translation levels per mRNA and that binding of Ago to one UTR cooperatively enhances binding to the other UTR. **m)** For our 4-input miRNA sensors, we observed antagonism within the UTRs and synergy across UTRs, again supporting the notion that miRNA target sites within the 5' and 3' UTRs repress by separate mechanisms and are independently saturable (Fig. 2-4).

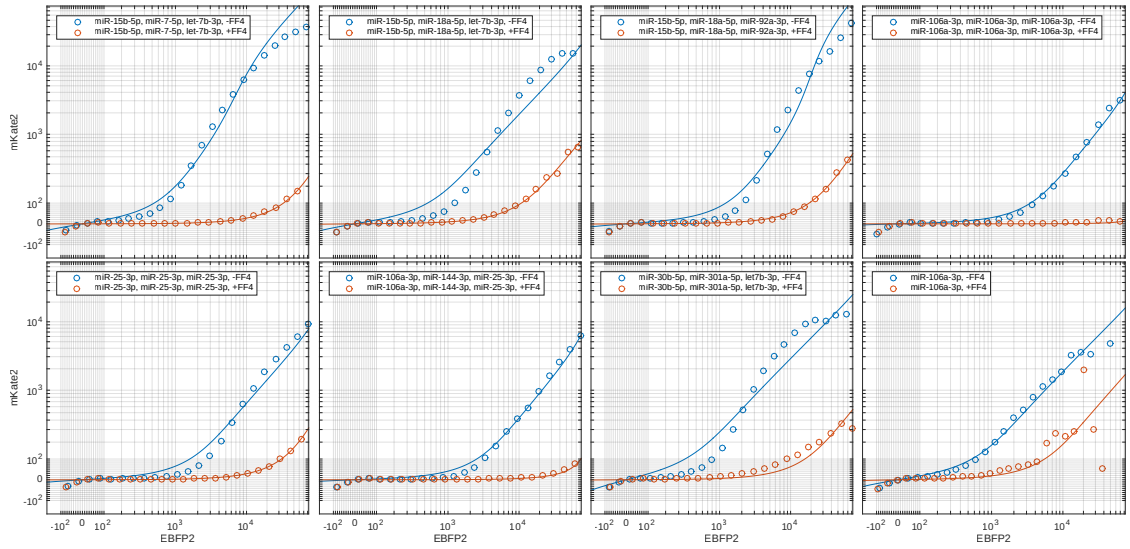
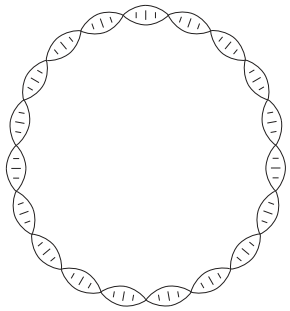


Figure 2-18: Antagonistic interactions not due to repression limits
Several variants of miRNA sensors bearing sets of 4x miRNA target repeats and also a set of 4x FF4 target sites within the 3' UTR were tested in HEK293FT cells with and without FF4 siRNA. In all cases, addition of FF4 siRNA was able to further knockdown mKate2 expression, showing that repression is not complete with three sets of high activity miRNA targets. Interestingly, FF4 knockdown seemed to take place in addition to knockdown from endogenous miRNAs (i.e. more synergistic rather than antagonistic), suggesting distinct mechanisms for reporter repression with siRNA and miRNA.



This page intentionally left blank to keep captions near figures.

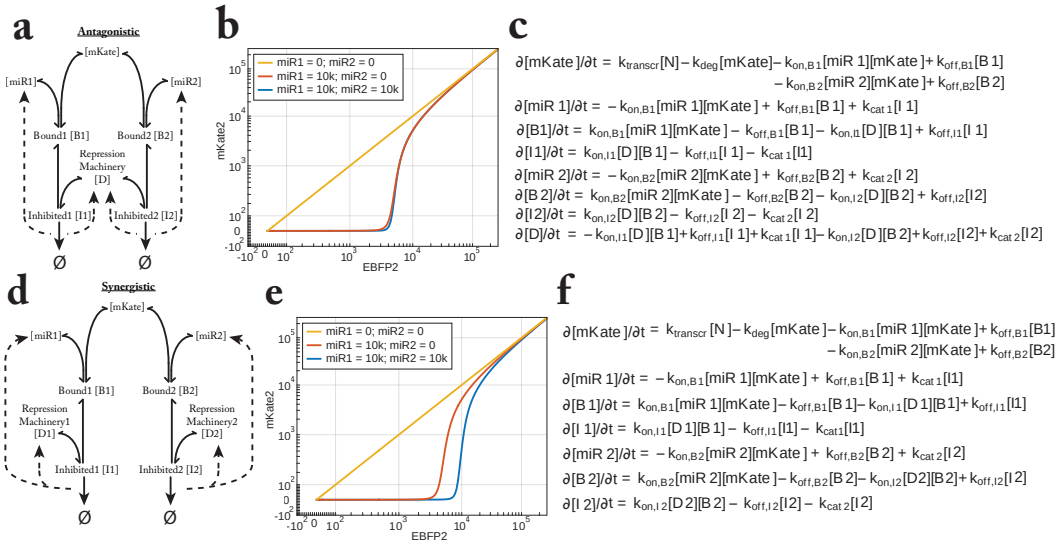


Figure 2-19: Speculative ODE model for antagonistic and synergistic repression.

We propose general ordinary differential equation (ODE) models describing repression from miRNAs when repression machinery is shared, generating an antagonistic interaction, or distinct, generating a synergistic interaction. **a)** The model for antagonistic interactions includes mKate transcripts that can be reversibly bound by different miRNA-RISC molecules [miR1] and [miR2] to form bound complexes. The bound complexes can then form inhibited complexes after reversible binding with a shared pool of repression machinery. Inhibited complexes can also be irreversibly degraded, eliminating the mKate transcript and regenerating the miRNA and repression machinery. **b)** Results from a numerical simulation of the model described in (a) with two similar miRNAs. miRNAs are present at concentrations of either 0 or 10,000 molecules per cell. As expected from an antagonistic interaction, the addition of miR2 results in minimal further repression compared to miR1 alone since the repression machinery is saturated at concentrations provided by miR1 itself. **c)** The ODEs used in (a) and (b) are listed here for reference. **d)** The model for synergistic interactions is similar to that for the antagonistic model, with the exception that the repression machinery for miR1 and miR2 are distinct, since in this case miR1 and miR2 would have target sites in opposite UTRs (in this example miR1 in the 5' UTR and miR2 in the 3' UTR). Note that initial D2 = D for the antagonistic model, since both D2 and D2 are concentrations for the repression machinery associated with the 3' UTR. For simplicity, we also set D1 = D since we generally observe that repression from 5' target sites is similar to that from 3' target sites, though this equivalence

is not required for a synergistic interaction to be observed. **e)** In the simulation corresponding to the model in (d), miR2 is able to contribute to further repression of mKate compared to miR1 alone, since repression machinery not shared between the two miRNAs. This behavior is descriptive of a synergistic relationship. **f)** The ODEs used in (d) and (e) are listed here. Equations are similar to (c) except that there are two terms for repression machinery [D1] and [D2].

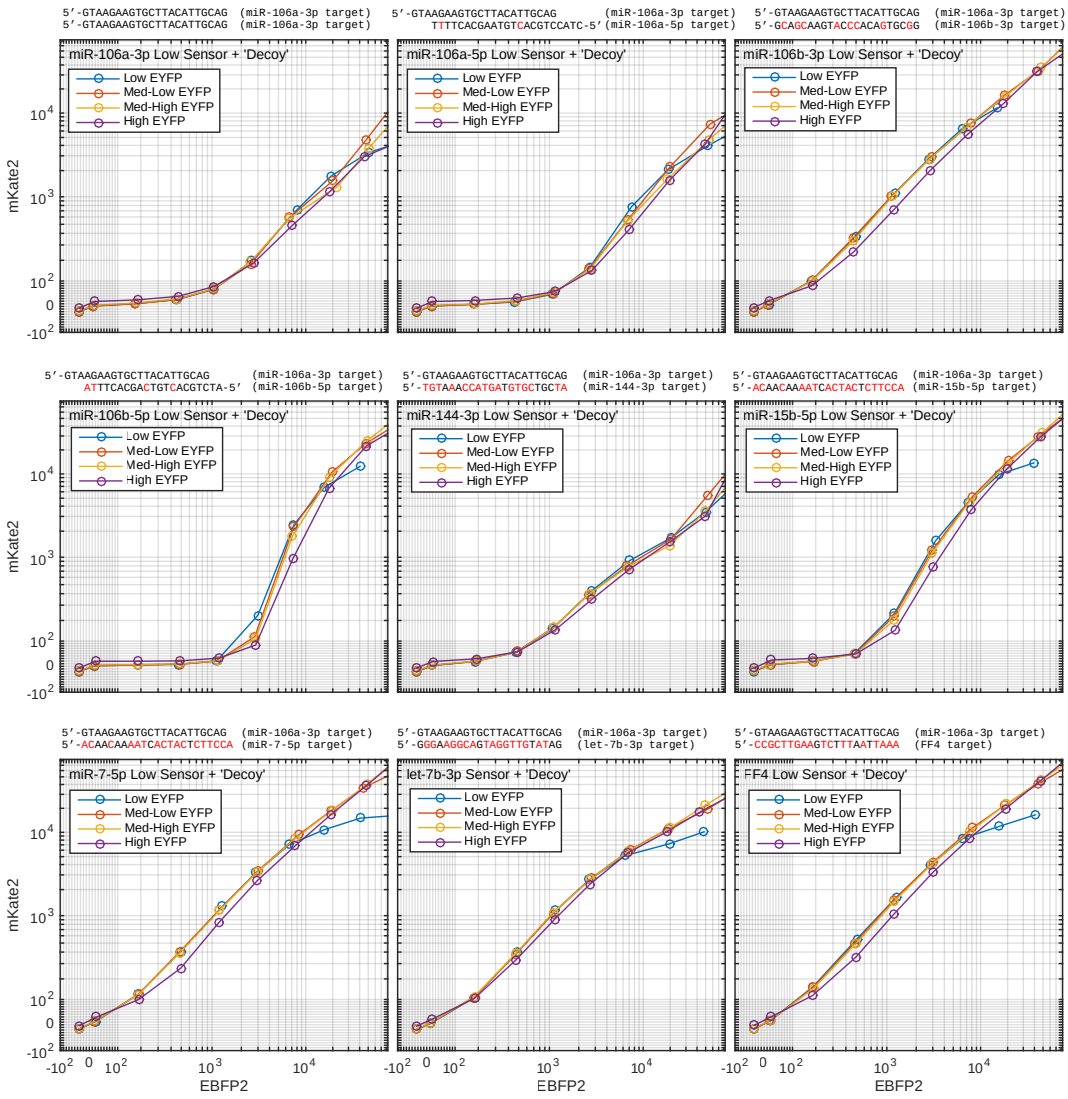


Figure 2-20: Minimal resource sharing between miRNA sensors of related and unrelated miRNAs.

To test whether transfection of miRNA sensors has measurable effect on repression of other miRNA targets, a 'decoy' sensor encoding miR-106a-3p target sites along with a separate hEF1a-EYFP transfection marker was built. Sequential transfection of the decoy sensor followed by one of nine low sensors was conducted in HEK293FT cells. Cells were binned from low to high EYFP expression - indicating low to high levels of decoy sensor - and traditional EBFP2 vs mKate2 plots were constructed. For all tested sensors, including related (miR-106a-3p, miR-106a-5p, miR-106b-3p, miR-

106b-5p) and unrelated (miR-144-3p, miR-15b-5p, miR-7-5p, let-7b-3p) sensors, no trend was seen across different decoy sensor levels. If resource sharing was apparent, less miRNA activity would be observed for higher levels of decoy sensor, but this was not the case. These results suggest that miRNA sensors impart low or undetectable levels of resource sharing on other targets.

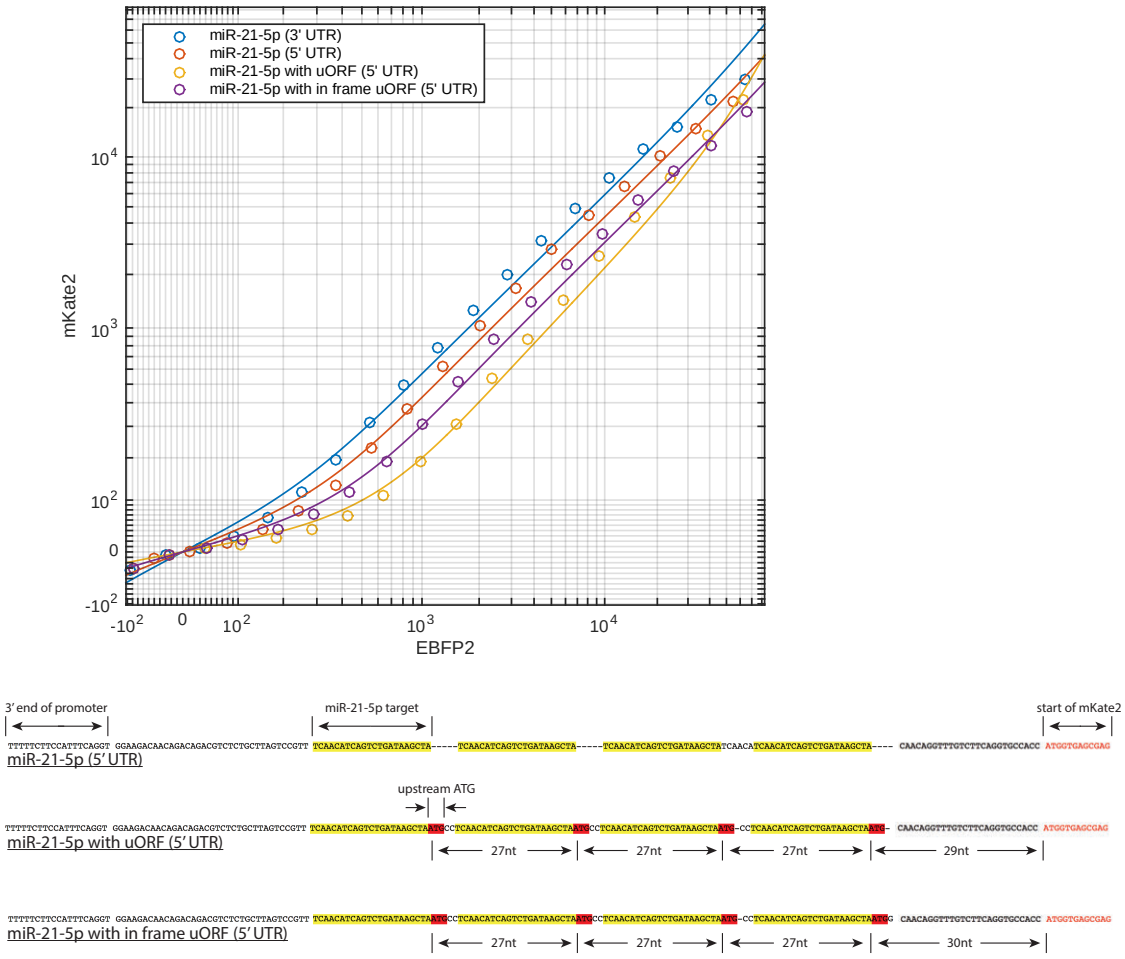


Figure 2-21: Repression of sensors bearing targets with and without uORFs.

miRNA low sensors were constructed bearing four repeats of miR-21-5p in the 3' UTR (blue circles), or 5' UTR (red, yellow, purple circles). Since miR-21-5p does not have ATGs, several ATGs were added between the miR-21-5p target sites in some sensors. Extra bases were added to separate the ATGs by a number of bases divisible by three. The distance between the last ATG and the true reporter start codon was either divisible by three (purple circles) or not (yellow circles). We observed a 2-fold drop in fluorescence when out of frame ATGs were added that was partially recovered when the ATGs were then shifted into frame. When using miRNA targets containing ATGs, altering the sensor design to keep them in frame may allow better measurements of miRNA activity.

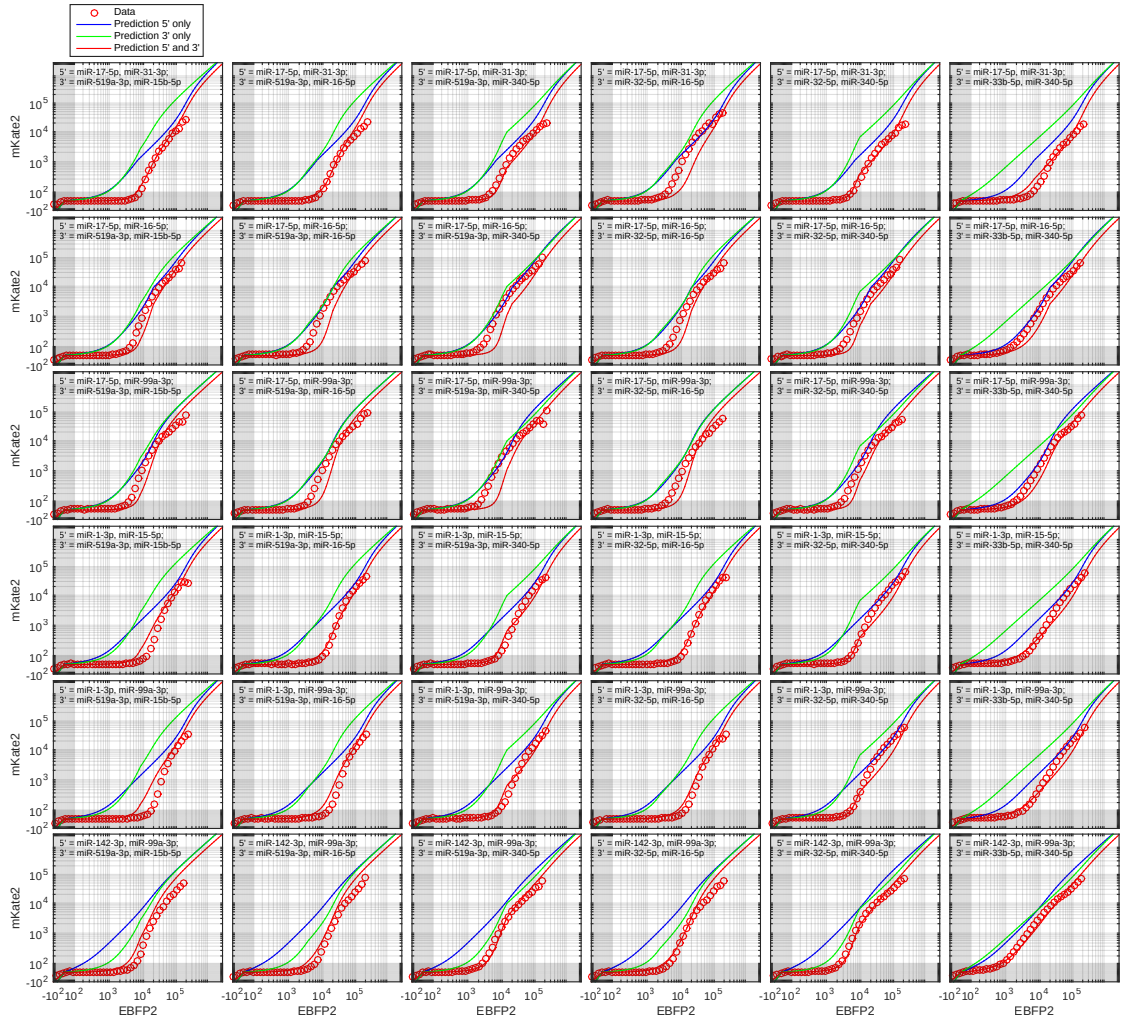


Figure 2-22: Predictions for Ant/Syn model

miRNA activity data and predictions using Ant/Syn model for 36 different 4-input classifiers. Predictions are shown for 5' UTR only (blue lines), 3' UTR only (green lines), combined prediction (red lines), and data (red circles). In general, predictions explain observed miRNA activity well.

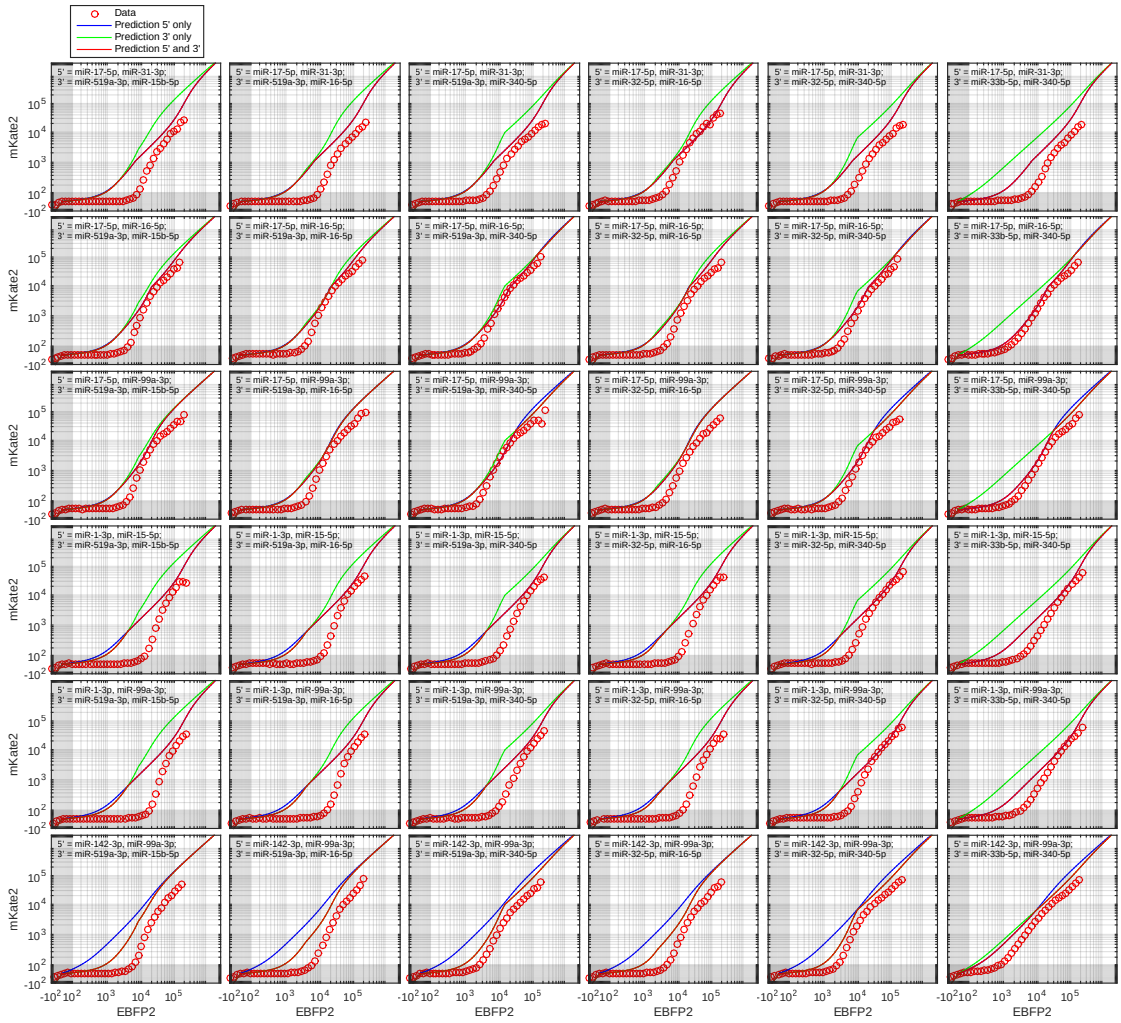


Figure 2-23: Predictions for antagonist-only model
miRNA activity data and predictions using antagonistic-only model for 36 different 4-input classifiers. Predictions are shown for 5' UTR only (blue lines), 3' UTR only (green lines), combined prediction (red lines), and data (red circles). In general, predictions generated by the antagonistic-only model underestimate observed miRNA activity.

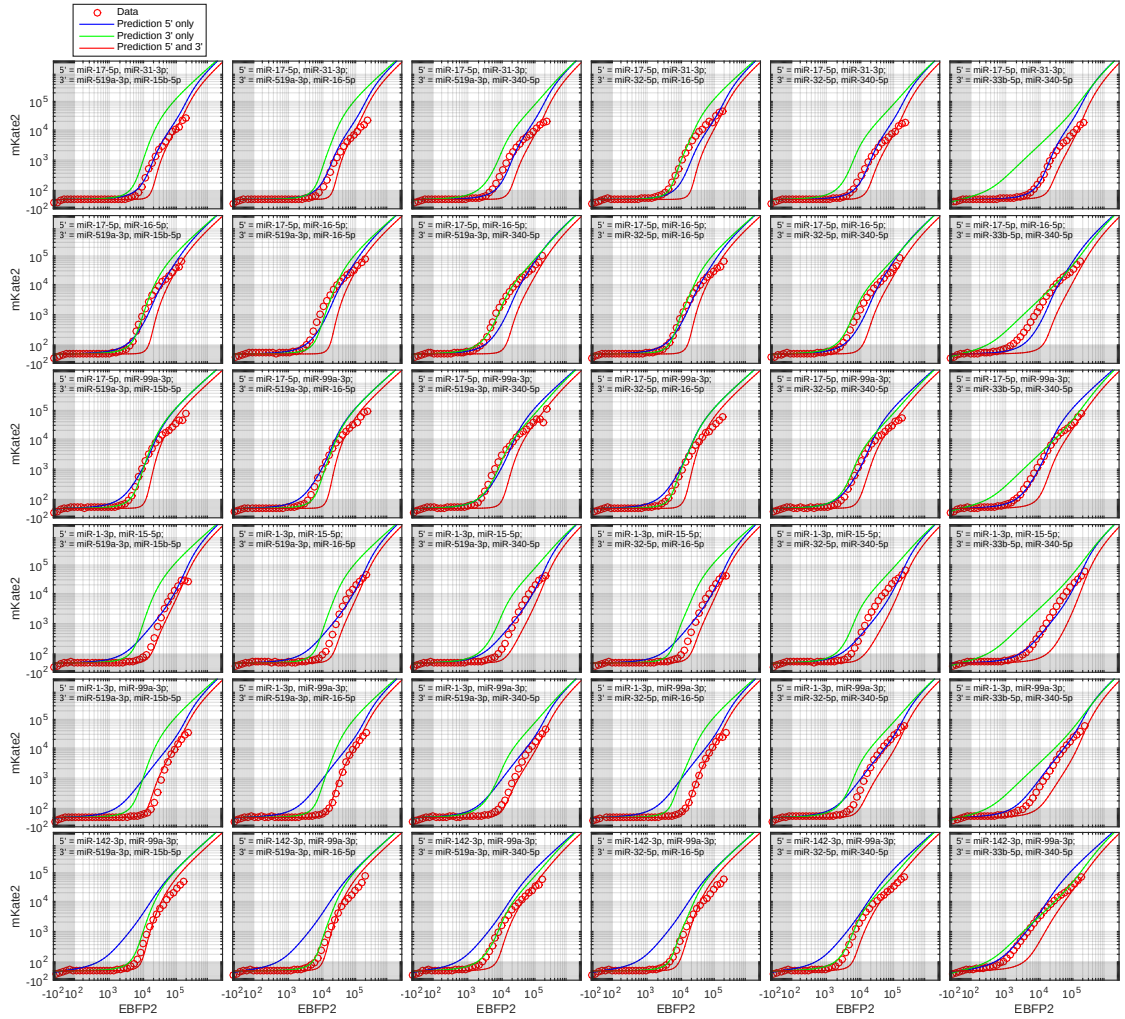


Figure 2-24: Predictions for the synergistic-only model
miRNA activity data and predictions using synergistic-only model for 36 different 4-input classifiers. Predictions are shown for 5' UTR only (blue lines), 3' UTR only (green lines), combined prediction (red lines), and data (red circles). In general, predictions generated by the synergistic-only model overestimate observed miRNA activity.

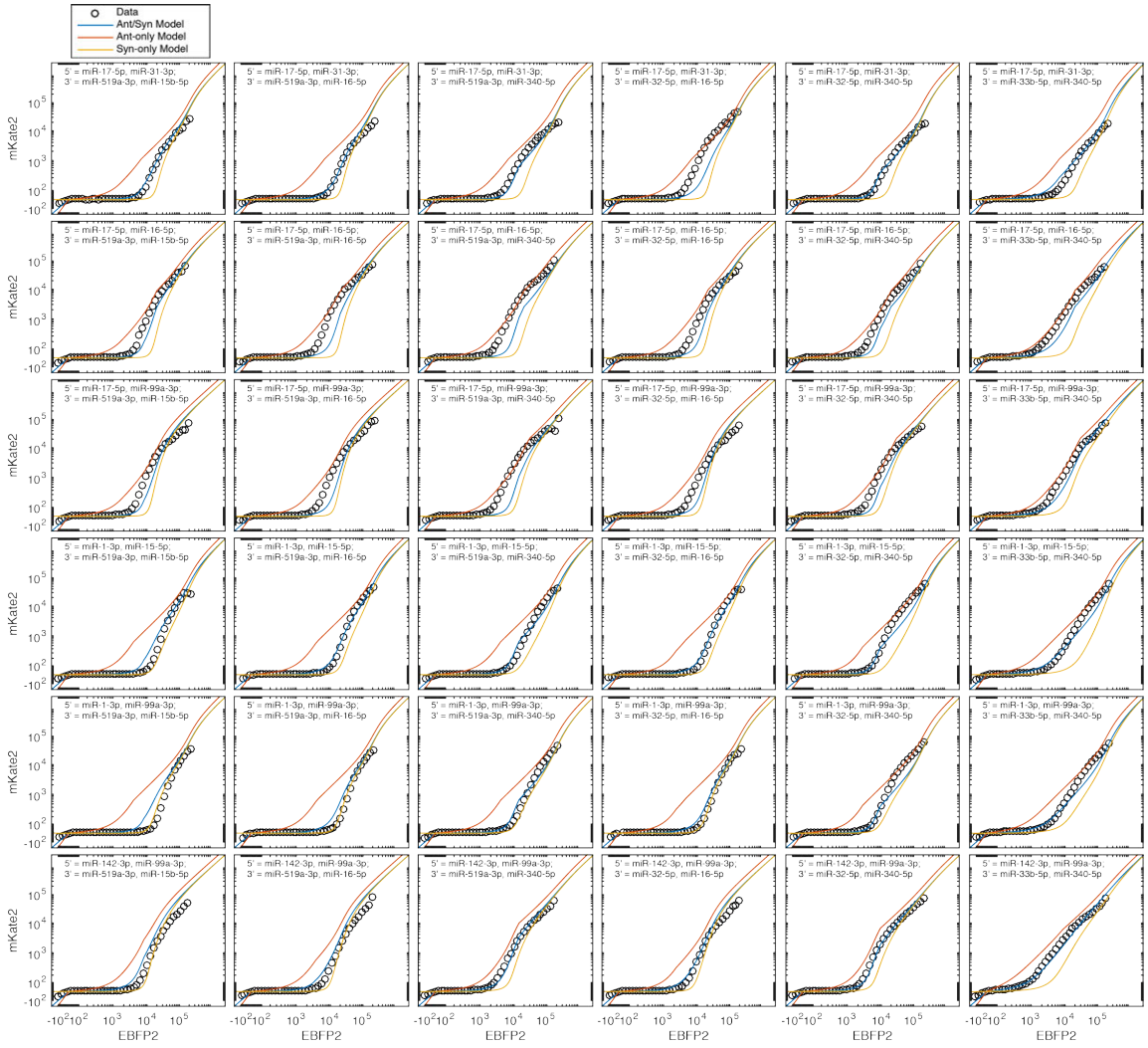


Figure 2-25: Comparison of predictions for all three models
Direct comparison between the three models for 36 different 4-input classifiers. In most cases, the Ant/Syn model best explains the data while the antagonistic-only model underestimates activity and the synergistic-only model overestimates activity.

HEK 293FT	HeLa	HepG2	miRNAs
High	Low	Low	miR-16-5p, miR-18a-5p, miR-519c-3p, miR-520c-3p
Low	High	Low	miR-27b-3p, miR-29b-3p, miR-98-5p
Low	Low	High	miR-16-5p, miR-10a-3p, miR-106b-3p, miR-142-5p
Low	High	High	miR-16-5p, miR-21, miR-23a-3p, miR-130a-3p, miR-29c-3p

Table 2.1: Best candidate miRNAs with specific activities among tested cell lines.

miRNAs listed here were considered for construction in HEK293FT classifiers. Classifiers with highest on/off predicted by miCAD were constructed and tested in cells.

Plasmid Name UTR Location GG Overhang	**miRNA 1** 5' UTR Q3-Q5	**miRNA 2** 5' UTR Q5-Q9	**miRNA 3** 3' UTR Q1-Q3	**miRNA 4** 3' UTR Q3-Q2
JG324	miR-17-5p (H)	miR-31-3p (H)	miR-519a-3p (H)	miR-15b-5p (H)
JG325	miR-17-5p (H)	miR-31-3p (H)	miR-519a-3p (H)	miR-16-5p (M)
JG326	miR-17-5p (H)	miR-31-3p (H)	miR-519a-3p (H)	miR-340-5p (L)
JG327	miR-17-5p (H)	miR-31-3p (H)	miR-32-5p (M)	miR-16-5p (M)
JG328	miR-17-5p (H)	miR-31-3p (H)	miR-32-5p (M)	miR-340-5p (L)
JG329	miR-17-5p (H)	miR-31-3p (H)	miR-33b-5p (L)	miR-340-5p (L)
JG330	miR-17-5p (H)	miR-16-5p (M)	miR-519a-3p (H)	miR-15b-5p (H)
JG331	miR-17-5p (H)	miR-16-5p (M)	miR-519a-3p (H)	miR-16-5p (M)
JG332	miR-17-5p (H)	miR-16-5p (M)	miR-519a-3p (H)	miR-340-5p (L)
JG333	miR-17-5p (H)	miR-16-5p (M)	miR-32-5p (M)	miR-16-5p (M)
JG334	miR-17-5p (H)	miR-16-5p (M)	miR-32-5p (M)	miR-340-5p (L)
JG335	miR-17-5p (H)	miR-16-5p (M)	miR-33b-5p (L)	miR-340-5p (L)
JG336	miR-17-5p (H)	miR-99a-3p (L)	miR-519a-3p (H)	miR-15b-5p (H)
JG337	miR-17-5p (H)	miR-99a-3p (L)	miR-519a-3p (H)	miR-16-5p (M)
JG338	miR-17-5p (H)	miR-99a-3p (L)	miR-519a-3p (H)	miR-340-5p (L)
JG339	miR-17-5p (H)	miR-99a-3p (L)	miR-32-5p (M)	miR-16-5p (M)
JG340	miR-17-5p (H)	miR-99a-3p (L)	miR-32-5p (M)	miR-340-5p (L)
JG341	miR-17-5p (H)	miR-99a-3p (L)	miR-33b-5p (L)	miR-340-5p (L)
JG342	miR-1-3p (M)	miR-16-5p (M)	miR-519a-3p (H)	miR-15b-5p (H)
JG343	miR-1-3p (M)	miR-16-5p (M)	miR-519a-3p (H)	miR-16-5p (M)
JG344	miR-1-3p (M)	miR-16-5p (M)	miR-519a-3p (H)	miR-340-5p (L)
JG345	miR-1-3p (M)	miR-16-5p (M)	miR-32-5p (M)	miR-16-5p (M)
JG346	miR-1-3p (M)	miR-16-5p (M)	miR-32-5p (M)	miR-340-5p (L)
JG347	miR-1-3p (M)	miR-16-5p (M)	miR-33b-5p (L)	miR-340-5p (L)
JG348	miR-1-3p (M)	miR-99a-3p (L)	miR-519a-3p (H)	miR-15b-5p (H)
JG349	miR-1-3p (M)	miR-99a-3p (L)	miR-519a-3p (H)	miR-16-5p (M)
JG350	miR-1-3p (M)	miR-99a-3p (L)	miR-519a-3p (H)	miR-340-5p (L)
JG351	miR-1-3p (M)	miR-99a-3p (L)	miR-32-5p (M)	miR-16-5p (M)
JG352	miR-1-3p (M)	miR-99a-3p (L)	miR-32-5p (M)	miR-340-5p (L)
JG353	miR-1-3p (M)	miR-99a-3p (L)	miR-33b-5p (L)	miR-340-5p (L)
JG354	miR-142-3p (L)	miR-99a-3p (L)	miR-519a-3p (H)	miR-15b-5p (H)
JG355	miR-142-3p (L)	miR-99a-3p (L)	miR-519a-3p (H)	miR-16-5p (M)
JG356	miR-142-3p (L)	miR-99a-3p (L)	miR-519a-3p (H)	miR-340-5p (L)
JG357	miR-142-3p (L)	miR-99a-3p (L)	miR-32-5p (M)	miR-16-5p (M)
JG358	miR-142-3p (L)	miR-99a-3p (L)	miR-32-5p (M)	miR-340-5p (L)
JG359	miR-142-3p (L)	miR-99a-3p (L)	miR-33b-5p (L)	miR-340-5p (L)

Table 2.2: miRNA target combinations for sensors with targets in the 5' UTR.

miRNA names, UTR location, and golden gate overhangs are listed. miRNA activities in HEK293FT cells are indicated as high (H), medium (M), or low (L)

Rate	Value	Unit (per cell)	Description	Reference
k_{trs}	7	mRNAs/hour	transcr rate	[89,90]
$k_{deg,m}$	0.5	1/hour	mRNA degr. rate	[91,92]
$k_{tln,EBFP2}$	5	proteins/mRNA/hr	EBFP2 transl. rate	[90]
$k_{tln,mKate2}$	5	proteins/mRNA/hr	mKate2 transl. rate	[90]
$k_{deg,EBFP2}$	0.5	1/hour	EBFP2 degr. rate	[93]
$k_{deg,mKate2}$	0.5	1/hour	mKate2 degr. rate	[93]
K_M	various	molecules	Michaelis constant	[94]
M	various	molecules	total effective miRNA concentration	[95]

Table 2.3: Rate constants used for miRNA repression model

Designation	Sequence	Description
Q1	GCTT	miRNA target overhang 1
Q2	CAAC	miRNA target overhang 2
Q3	CAGA	miRNA target overhang 3
Q4	TGTG	miRNA target overhang 4
Q5	GAGC	miRNA target overhang 5
Q6	AACG	miRNA target overhang 6
Q7	CTTC	miRNA target overhang 7
Q8	AGAC	miRNA target overhang 8
Q9	AGGT	miRNA target overhang 9
Qa	GGAG	overhang between backbone and insulator
Qb	TACT	overhang between insulator and promoter
Qc	CAGA	overhang between promoter and 5' UTR
Qd	AGGT	overhang between 5' UTR and gene
Qe	GCTT	overhang between gene and 3' UTR
Qf	CAAC	overhang between 3' UTR and poly A
Qg	CGCT	overhang between poly A and backbone

Table 2.4: Golden Gate overhangs used for assembly

2.7 Modeling

The model as described includes reactions listed in equations 2.1 to 2.5. In each cell, the number of plasmids is assumed to be a constant value such that the steady state approximation can be made.

$$k_{trs}N - k_{deg,m}[m_{EBFP2}] = 0 \tag{2.1}$$

$$k_{tln,EBFP2}[m_{EBFP2}] - k_{deg,EBFP2}[EBFP2] = 0 \tag{2.2}$$

$$k_{trs}N - k_{deg,m}[m_{mKate2}] - k_{on,miR}[m_{mKate2}][m_{miR}] + k_{off}[m_{mKate2,miR}] = 0 \tag{2.3}$$

$$k_{tln,mKate2}[m_{mKate2}] - k_{deg,mKate2}[mKate2] = 0 \tag{2.4}$$

$$k_{on,miR}[m_{mKate2}][m_{miR}] - k_{off}[m_{mKate2,miR}] - k_{cat}[m_{mKate2,miR}] = 0 \tag{2.5}$$

The total concentration of miRNA is the summation of both free (m_{miR}) and bound ($m_{mKate2,miR}$) species. The Michaelis constant is used as a lump parameter, fitting only two parameters to characterize miRNA activity minimizes the risk of overfitting.

$$M = [m_{miR}] + [m_{mKate2,miR}] \tag{2.6}$$

$$K_m = \frac{k_{cat} + k_{off}}{k_{on}} \tag{2.7}$$

Use of additional lumped parameters α to δ simplifies the form of equation 2.12. These lumped parameters have constant value for all miRNA sensors regardless of miRNA activity since they depend only on rates of transcription, translation, and degradation of unbound mKate2 and EBFP2 species.

$$\alpha = \frac{k_{trs}}{k_{deg,m}} \tag{2.8}$$

$$\beta = \frac{k_{cat}}{k_{deg,m}} \tag{2.9}$$

$$N = \frac{k_{deg,m}k_{deg,EBFP2}}{k_{trs}k_{tln,EBFP2}}[EBFP2] = \gamma[EBFP2] \tag{2.10}$$

$$[mKate2] = \frac{k_{tln,EBFP2}}{k_{deg,mKate2}}[m_{mKate2}] = \delta[EBFP2] \tag{2.11}$$

The solution for mKate2 concentration takes a quadratic form for a single-input miRNA sensor. The two parameters quantifying miRNA activity are M and K_m.

$$\begin{aligned}[mKate2] = \frac{\delta}{2}(\sqrt{(-\alpha\gamma[EBFP2] + \beta * M + K_m)^2 + 4\alpha\gamma K_m[EBFP2]}) \\ + \frac{\delta}{2}(\alpha\gamma[EBFP2] - \beta M - K_m)\end{aligned} \tag{2.12}$$

To obtain predictions for combined miRNA activity for two-input (or more) sensors, several approaches can be taken. While numerical simulation could be used, its implementation would be relatively slow here. We chose to use a more general

approach borrowed from combination of (drug) inhibitors where combinations are divided into three classes of interactions: additive - where effects from each inhibitor alone can be 'summed' to obtain the combined effect since inhibitors are assumed not to interact with each other, synergy - where the combined effect is greater than the sum of each inhibitor alone, and antagonism - where the combined effect is less than the sum of each inhibitor alone. General equations for combining miRNA activities using these three classes of interactions have been previously described[74] and are given as follows, where $m_{add.}$, $m_{synerg.}$, and $m_{antag.}$ refer to mKate2 concentrations resulting from the combination of i different miRNAs which individually repress mKate2 concentration to levels m_i. $m_{neg.}$ refers to the mKate2 concentration in the negative control case where the sensor contains no miRNA target sites. For the antagonistic case, the limit as the Hill coefficient (n) approaches zero is taken since that is the case of perfect antagonism, where combined activity simply reflects the highest activity present in the individual sensors. Values for $m_{add.}$, $m_{synerg.}$, and $m_{antag.}$ are predicted across the entire range of EBFP2 expression (which serves as a transfection marker) since repression of mKate2 is often threshold-like and dependent on transfection levels. Example predictions using the three equations are shown in Figure 2-2.

$$\frac{1}{m_{add.}} = \sum_{i=1}^{n} \frac{1}{m_i} - \frac{n-1}{m_{neg.}} \tag{2.13}$$

$$\frac{m_{syn.}}{m_{neg.}} = \prod_{i=1}^{n} \frac{m_i}{m_{neg.}} \tag{2.14}$$

$$\lim_{n\to 0} \left(\frac{1 - m_{ant.}}{m_{ant.}} \right)^{1/n} = \lim_{n\to 0} \sum_{i=1}^{k} \left(\frac{1 - m_i}{m_i} \right)^{1/n} \tag{2.15}$$

Equation 2.15 should be equivalent to taking the maximum miRNA activity (minimal mKate2 expression) for constituent miRNAs at each level of EBFP2 expression, such that:

$$\frac{m_{ant.}}{m_{neg.}} = min\left\{\frac{m_1}{m_{neg.}}, \ldots, \frac{m_k}{m_{neg.}}\right\} \tag{2.16}$$

2.8 Design considerations

One design consideration to note is that miRNA sensors bearing target sites in the 5' UTR can present challenges for a subset of miRNA target sites bearing an 'ATG' sequence (e.g. 176 of our 620 high confidence miRNA target sites contain an 'ATG'). In these cases, introduction of upstream open reading frames (uORFs) could result in reporter mKate2 knockdown that is not due to true miRNA activity. We tested several variants of sensors bearing miR-21-5p target sites (which do not contain a natural 'ATG' sequence) in the 5' UTR, and also added either no uORFs, out of frame uORFs, or in frame uORFs in order to determine the possible contributions to knockdown (Fig. 2-21). By introducing additional nucleotides between miRNA target sites to place the uORFs in frame with mKate2 we observed that about half of the lost fluorescence could be recovered. In certain applications, this recovery may allow for miRNA targets bearing ATGs to be placed within the 5' UTR with small alterations.

References

1. Brown, B. D. *et al.* Endogenous microRNA Can Be Broadly Exploited To Regulate Transgene Expression According To Tissue, Lineage and Differentiation State. *Nature Biotechnology* **25,** 1457–1467 (2007).

3. Kozomara, A., Hunt, S., Ninova, M., Griffiths-Jones, S. & Ronshaugen, M. Target repression induced by endogenous microRNAs: large differences, small effects. *PloS one* **9,** e104286 (2014).

11. Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods* **9,** 840–846 (2012).

12. Kozomara, A. & Griffiths-Jones, S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42,** 68–73 (2014).

18. Cloonan, N. Re-thinking miRNA-mRNA interactions: Intertwining issues confound target discovery. *BioEssays* **37,** 379–388 (2015).

22. Eichhorn, S. W. *et al.* MRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular Cell* **56,** 104–115 (2014).

25. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics* **16,** 421–433 (2015).

26. Guo, Z. W. *et al.* MtiBase: A database for decoding microRNA target sites located within CDS and 5'UTR regions from CLIP-Seq and expression profile datasets. *Database* **2015,** 1–9 (2015).

27. Lee, I. *et al.* New class of microRNA targets containing simultaneous 5' -UTR and 3'-UTR interaction sites. *Genome Research* **19,** 1175–1183 (2009).

28. Fang, Z. & Rajewsky, N. The impact of miRNA target sites in coding sequences and in 3' UTRs. *PLoS ONE* **6,** 1–6 (2011).

29. Lytle, J. R., Yario, T. a. & Steitz, J. a. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Pro-*

ceedings of the National Academy of Sciences of the United States of America **104,** 9667–9672 (2007).

33. Aeschimann, F. *et al.* LIN41 Post-transcriptionally Silences mRNAs by Two Distinct and Position-Dependent Mechanisms. *Molecular Cell* **65,** 476–489.e4 (2017).

34. Gurtan, A. M. & Sharp, P. A. The role of miRNAs in regulating gene expression networks. *Journal of Molecular Biology* **425,** 3582–3600 (2013).

35. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333,** 1307–1311 (2011).

36. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).

37. Leidinger, P. *et al.* A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biology* **14,** R78 (2013).

38. Ikeda, S. *et al.* Altered microRNA expression in human heart disease. *Physiological Genomics* **31,** 367–373 (2007).

39. Miki, K. *et al.* Efficient Detection and Purification of Cell Populations Using Synthetic MicroRNA Switches. *Cell Stem Cell* **16,** 699–711 (2015).

40. Sayeg, M. K. *et al.* Rationally Designed MicroRNA-Based Genetic Classifiers Target Specific Neurons in the Brain. *ACS Synthetic Biology* **4,** 788–795 (2015).

41. Bloom, R. J., Winkler, S. M. & Smolke, C. D. A quantitative framework for the forward design of synthetic miRNA circuits. *Nature Methods* **11,** 1147–53 (2014).

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

44. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature genetics* **43,** 854–9 (2011).

45. Mukherji, S. & van Oudenaarden, A. Synthetic biology: Understanding biological design from synthetic circuits. *Nature Reviews Genetics* **10,** 859–871 (2009).

55. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS ONE* **4** (2009).

62. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110 (2001).

64. Golden, R. J. *et al.* An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* **542,** 197–202 (2017).

65. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4,** 1–38 (2015).

66. Grey, F. *et al.* A viral microRNA down-regulates multiple cell cycle genes through mRNA 5'UTRs. *PLoS Pathogens* **6** (2010).

67. Rouya, C. *et al.* Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. *RNA* **20,** 1398–1409 (2014).

68. Hon, L. S. & Zhang, Z. The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. *Genome Biology* **8,** 1–18 (2007).

69. Wu, S. *et al.* Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* **29,** 2302–2308 (2010).

70. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).

71. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).

72. Mintz, P. J. *et al.* MicroRNA-181a* Targets Nanog in a Subpopulation of CD34(+) Cells Isolated From Peripheral Blood. *Mol Ther Nucleic Acids* **1,** e34 (2012).

73. Wang, W.-C. *et al.* miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics* **10,** 328 (2009).

74. Chou, T.-C. & Talalay, P. Generalized Equations for the Analysis of Inhibitions of Michaelis-Menten and Higher-Order Kinetic Systems with Two or More Mutually Exclusive and Nonexclusive Inhibitors. *European Journal of Biochemistry* **115,** 207–216 (1981).

75. Broderick, J., Salomon, W., Ryder, S., Aronin, N. & Zamore, P. Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *Rna* **17,** 1858–1869 (2011).

76. Elkayam, E. *et al.* Multivalent Recruitment of Human Argonaute by GW182. *Molecular Cell* **67,** 646–658.e3 (2017).

77. Denzler, R. *et al.* Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Molecular Cell* **64,** 565–579 (2016).

78. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature reviews. Genetics* **12,** 99–110 (2011).

79. Fabian, M. R. & Sonenberg, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature Structural and Molecular Biology* **19,** 586–593 (2012).

80. Lam, J. K., Chow, M. Y., Zhang, Y. & Leung, S. W. siRNA versus miRNA as therapeutics for gene silencing. *Molecular Therapy - Nucleic Acids* **4** (2015).

81. Xuan, M., Xiaofeng, C., Beixin, M. & Xuemei, C. Trip to ER: MicroRNA-mediated translational repression in plants. *RNA Biology* **10,** 1586–1592 (2013).

82. Gu, S. & Kay, M. A. How do miRNAs mediate translational repression? *Silence* **1,** 11 (2010).

83. Valencia-Sanchez, M. A., Liu, J., Hannon, G. J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes and Development* **20,** 515–524 (2006).

84. Zhou, H. & Rigoutsos, I. MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *RNA (New York, N.Y.)* **20,** 1431–9 (2014).

85. Siciliano, V. *et al.* MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nature Communications* **4,** 1–7 (2013).

86. Moore, W. A. & Parks, D. R. Update for the logicle data scale including operational code implementations. *Cytometry Part A* **81 A,** 273–277 (2012).

87. Zhu, X. P. *et al.* Microarray analysis of the aberrant microRNA expression pattern in gliomas of different grades. *Oncology Reports* **34,** 318–324 (2015).

88. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).

89. Darzacq, X. *et al.* In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14,** 796–806 (2007).

90. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).

91. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology* **4,** 1707–1719 (2006).

92. Dar, R. D. *et al.* Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 17454–9 (2012).

93. Halter, M., Tona, A., Bhadriraju, K., Plant, A. L. & Elliott, J. T. Automated live cell imaging of green fluorescent protein degradation in individual fibroblasts. *Cytometry Part A* **71,** 827–834 (2007).

94. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature structural and molecular biology* **11,** 599–606 (2004).

95. Bosson, A. D., Zamudio, J. R. & Sharp, P. A. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Molecular Cell* **56,** 347–359 (2014).

Chapter 3

miRNA gene tags for measuring transcription

Biologists are increasingly making use of diverse types of measurements (e.g. steady state levels or transcription rates of RNA, steady state levels or translation rates of protein, and modifications, accessibility, or higher order conformation of DNA). At the same time, synthetic biologists are trying to use these measurements as inputs to live cells in order to actuate appropriate responses. In mammalian cells, the ability to sense and respond to endogenous expression levels in a modular fashion would enable complex behavior that is more temporal in nature. These 'embedded' circuits could serve as an interface between different biological processes, for instance sensing when iPS cells are have undergone differentiation to definitive endoderm, and inducing transcription factors to promote downstream differentiation to hepatocytes. A key part of these embedded circuits that is under active research is what signals can be reliably sensed and how to do so. Established examples of sensing include the expression of luciferase to gauge deliver and expression of nucleic acids delivered *in vivo*,

in situ hybridization, and the MS2-MCP system for labeling RNA.[96] Active areas of research include toehold-mediated RNA strand displacement[97] and modular protein sensing to activate transcription.[98] An ideal sensing technology would allow modular sensing of several genes simultaneously and result in a change of RNA and/or protein levels (if using transcriptional or translational regulation respectively). However, many of the above examples are not strictly modular and can only actuate fluorescent signals (e.g. ISH, MS2-MCP) while others are difficult to design and may require significant tuning for each sensor (e.g. RNA strand displacement, protein sensors).

Instead, we propose using miRNA sensors - which are easy to design and highly modular - to sense gene expression. To accomplish this, genes are tagged at their genomic locus with an artificial miRNA sequence, such that when a gene is transcribed, the miRNA is also transcribed and can be measured with a miRNA sensor (Fig. 3-1). In theory, the miRNA sensors can sense multiple miRNA sensors at the same time, using a combination of previously discussed miRNA high sensors and the Ant/Syn repression model. Additionally, the miRNA sensor output can be chosen to provide easy single-cell *in vitro* measurements (e.g. fluorescent proteins), *in vivo* measurements (e.g. luciferase), or actuation of downstream signals (e.g. a transcription factor). Here we show a proof of concept for how these 'miRNA gene tags' can provide useful measurements of gene activity in live cells.

3.1 Background and challenges

miRNA gene tags require several technologies for implementation, including miRNA sensors, design of artificial miRNAs, and reliable integration of miRNA tags into endogenous genes. Since we discuss miRNA sensors at length in Chapter 2 and

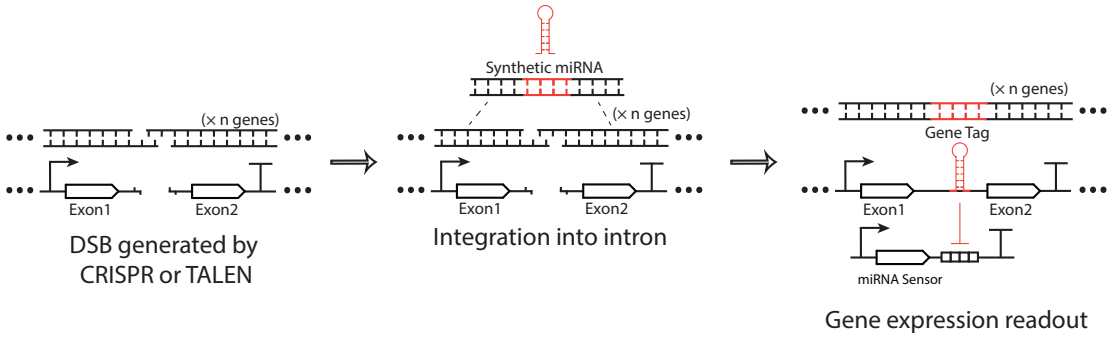


Figure 3-1: miRNA gene tag overview
miRNA gene tags generally involve the introduction of artificial miRNA sequences in the introns or near the 3' end of endogenous genes (to avoid nonsense mediated decay). In general this is achieved by generating double strand breaks (DSB) in a target way using CRISPR or TALENs, integration into the genome using repair templates of single- or double-stranded DNA with homology arms flanking a sequence for the miRNA hairpin (red), then using a miRNA sensor to sense the activity of the artificial miRNA, which should relate to transcription of the endogenous gene. Several genes (n) could be tagged with different miRNA sequences, allowing several distinct miRNA sensors or classifiers to sense many endogenous genes in a highly modular way.

design of artificial miRNAs in Chapter 4, we focus on the latter challenge here.

Segments of DNA can be inserted into the genome of human cells in several ways. Whereas methods like lentivirus and piggyBac transposons integrate their payload randomly, tagging individual genes requires targeted integration. Targeted integration has most commonly been achieved using nucleases based on zinc fingers, TALEs, and most recently CRISPR/Cas9. These nucleases introduce double-strand breaks at designated sequences which are repaired with homologous recombination (HR), given a repair template containing the sequence to be inserted flanked by homologous regions. However, HR is less efficient than alternative repair mechanism non-homologous end joining (NHEJ), meaning that most cells will not contain

the desired insertion (in some measures efficiency is 0.05-0.3% for HR compared to 0.6-1.3 for NHEJ).[99] As a result, cells bearing the edit must be selected (e.g. with antibiotic selection) or clonally sorted. While some efforts to knock down NHEJ have increased the efficiency of HR,[100] targeted editing still remains a significant problem for integrating genetic circuits into the genome. Efforts in our lab to create genomic landing pads in common cell lines have made integration of large payloads into predefined loci significantly easier.[101] However, for miRNA gene tags, payloads will often need to be integrated near endogenous genes without the use of antibiotic selection, since introduction of enhancer and promoter elements may disrupt endogenous gene regulation.* In the work that follows, we screened many different single cell colonies after gene editing to obtain cells with the targeted integration, though in the future hybridization and other approaches may allow efficient sorting of edited cells.

A final challenge involves combining the miRNA gene tag and miRNA sensor technologies, along with activation or repression of endogenous genes, to show that transcription of endogenous genes can be measured through the gene tag system. In this work we show that this function is possible in Sox2 and RHOXF2 genes with a strong FF4 artificial miRNA.

3.2 Initial design of miRNA gene tag

The original gene tag cell line was constructed starting from a HEK293-based cell line with genomically integrated expression of an FF4 sensor. This starting cell line designated YQ4035 was then modified to include FF4 artificial miRNA expression

*Demonstrated by collaborators at Thermo Fisher, Jon Chesnut (Jon.Chesnut@thermofisher.com) and Sanjay Kumar (Sanjay.Kumar2@thermofisher.com)

immediately downstream of the Sox2 locus, followed by constitutive expression of zerocin resistance from a CMV promoter.* This modification was carried out by introducing double strand breaks with Cas9 then repairing with an HR template with homology arms of 500bp. Expression of zeocin was flanked by LoxP sites so that after antibiotic selection and expression of Cre recombinase, the zeocin cassette could be excised, resulting in final cells bearing Sox2 tagged with FF4 and also expressing the associated FF4 sensor (Fig. 3-2).† Integration of FF4 with zeocin selection resulted in a cell line with approximately 10-fold reduced mKate levels compared to YQ4035 cells, presumably due to upregulation of Sox2 from activating elements from the CMV promoter driving zeocin. Upregulation of Sox2 was confirmed with qPCR. At this stage, no change in BFP fluorescence was observed, suggesting that changes in mKate were not due to error originating from the FF4 sensor. Following excision of zeocin selection by Cre, mKate fluorescence rose to levels approximately one third of that from YQ4035. After excision, Sox2 expression assayed by qPCR showed decrease back to original levels. It is possible that mKate levels did not recover completely to original ones due native expression of Sox2 repressing mKate with the FF4 gene tag.

Following construction of the Sox2 gene tag cell line, we at the Weiss lab sought to test whether measurements of perturbations of Sox2 expression could be measured from the FF4 sensor. We initially tested the use of TALE fusions to KRAB and VP64 guided to Sox2, but these fusions did not produce detectable changes in mKate expression of the sensor.‡ We then tested fusions with dCas9 (KRAB, VP160, VPR) and found that the FF4 responded to activation and repression of Sox2 with down or

*Construction of initial cell line was performed by collaborators at Thermo Fisher
†Sox genetag cell line is designated as line 2-7
‡Data can be found in `/jgam/Data/flow/2014-10-11_Gene_tag_initial`

upregulation of mKate by approximately 2-fold, which is the trend expected from the genetag circuit (Fig. 3-2).* Negative controls without gRNA did not show effects on mKate fluorescence, again as expected. It should be noted that the positive control with FF4 siRNA showed greater repression of mKate at approximately 10-fold, indicating that expression of FF4 from endogenous promoters was not as great as may be desired for maximum dynamic range.

3.3 Gene tag to measure RHOXF2 transcription

Following successful results with Sox2, we sought to refine gene tags in several ways. First we wanted to eliminate the need for antibiotic selection and Cre excision, since for instance in iPS cells it would be undesirable to inadvertently change expression of transcription factors being tagged. We also wanted to expand the range of genes that have been tagged and test the possible dynamic range of miRNA gene tags. Therefore we attempted to tag genes (OCT4, ASCL1 and RHOXF2) that have been shown to be activated by a dCas9-VPR fusion (as much as 18,000-fold activation).[102] We built backbones JG193, JG194, and JG195 that contain 500bp homology arms and which also contain Golden Gate assembly sites so that any miRNA (or combinations of miRNAs) can be assembled between the homology arms. We assembled FF4 expression into these repair template vectors and used PCR to generate linear DNA, before using Cas9 to edit YQ4035 cells with the gene tags. Following editing we sorted cells into single wells and allowed them to recover (reaching confluency in a 96-well plate) before splitting and lysing a portion of the cells in preparation for genomic PCR. Following genomic PCR, we identified a clone (2-10) appearing to have

*Data can be found in `/jgam/Data/flow/2016-03-11_GenTag_KRAB_new_param`

heterozygous integration the RHOXF2 gene tag as indicated by higher mass bands on an agarose gel.* Unfortunately, the other genes did not show clones with apparent insertions of the gene tag, indicating that future improvements in the efficiency of homologous recombinations will be necessary for efficient production of gene tag cell lines.

With the RHOXF2 gene tag we were able to show that activation of RHOXF2 with dCas9-VPR produced as much as a 10-fold decrease in mKate2 but only in a subpopulation of cells, suggesting that activation with dCas9-VPR may be more stochastic compared to more traditional transcriptional activators (Fig. 3-3). Repression of RHOXF2 with dCas9-KRAB showed minimal effect perhaps because RHOXF2 is minimally expressed such that repression does not significantly decrease transcription, an observation that would support previous findings that genes with greatest potential for activation tend to have minimal background expression.[102]

3.4 Discussion

Here we have shown that miRNA gene tags can be used to modulate expression from a reporter in response to changes in endogenous transcription. With future optimizations, miRNA gene tags could allow for robust temporal programs that both measure and module endogenous biological processes more closely. With enough orthogonal miRNA gene tags for many different genes, entire biological pathways could be monitored simultaneously. However several challenges remain for miRNA gene tags to be incorporated in useful circuits. The on/off ratio would likely need to be improved to increase the dynamic range for downstream regulation. This may be achieved to

*`jgam/Data/Gels/2016-10-12`

some degree by optimizing the design of gene tags, including the hairpin structure, position within UTRs or introns and the distance from the endogenous gene, and the artificial miRNA sequence itself. Alternatively, construction of an 'operational amplifier' in cells could be used to amplify smaller signals from gene tags into larger changes in the downstream part of the circuit. Though operational amplifiers would be a significant research undertaking in it's own right, and be applicable to many areas within synthetic biology. Gene tags also require highly efficient and targeted integration into the mammalian genome ideally without antibiotic selection, which remains an active area of research within biology. Technology like CRISPR has made targeted edits easier, but not efficient enough to the point of allowing techniques like Multiplex Automated Genomic Engineering (MAGE) due in part to the long growth times for mammalian cells. As such, large-scale genome-wide integration of miRNA gene tags will likely need to wait until targeted modifications become more efficient. In the near term, improvements may come in the form of overexpression of genes involved in homologous recombination like RS-1 or RAD51[103] or knockdown of NHEJ components.[100] Nevertheless, miRNA gene tags should enable study and manipulation of biological processes in the near term and, combined with other synthetic biology technology, organisms in the future.

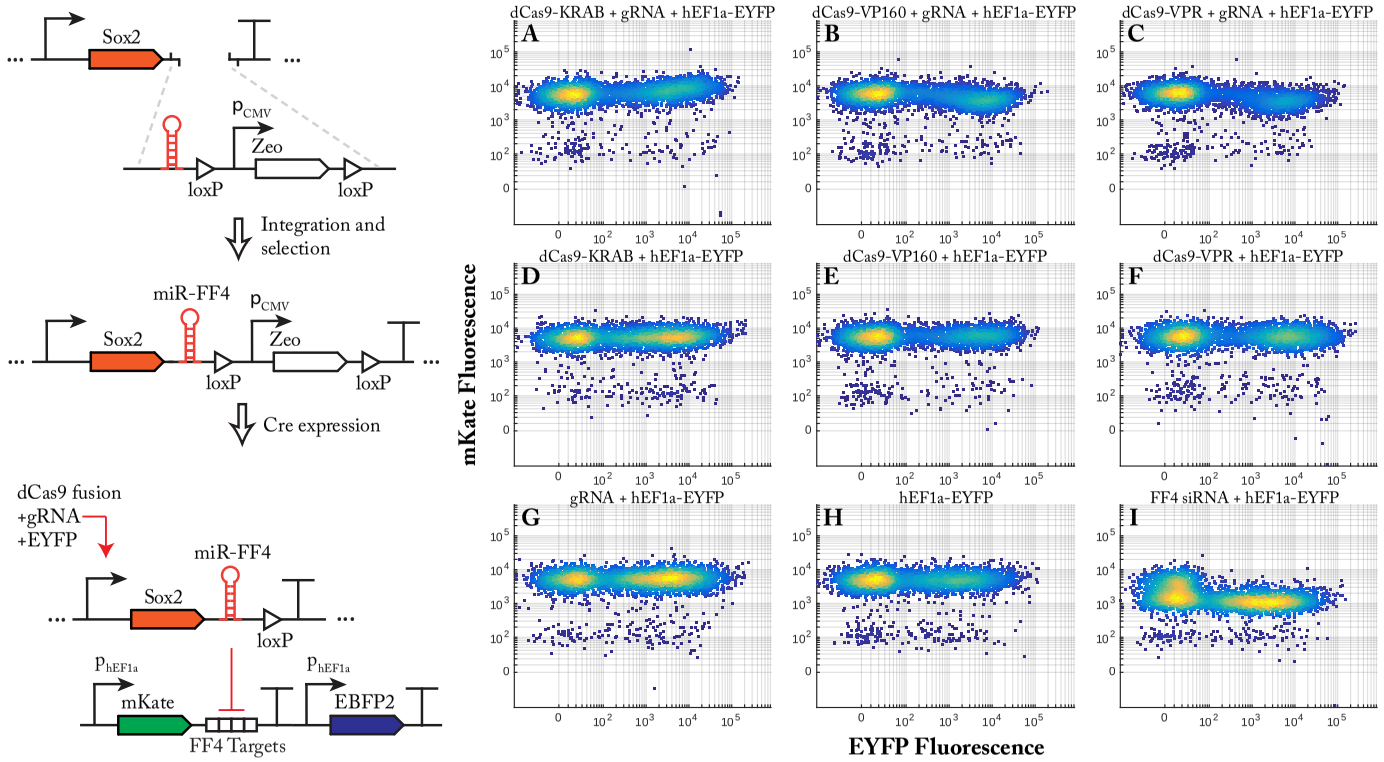

Sox2
P_{CMV}
Zeo
loxP
loxP
Integration and selection
miR-FF4
Cre expression
dCas9 fusion
+gRNA
+EYFP
P_{hEF1a}
mKate
EBFP2
FF4 Targets
dCas9-KRAB + gRNA + hEF1a-EYFP
dCas9-VP160 + gRNA + hEF1a-EYFP
dCas9-VPR + gRNA + hEF1a-EYFP
gRNA + hEF1a-EYFP
hEF1a-EYFP
FF4 siRNA + hEF1a-EYFP
mKate Fluorescence
EYFP Fluorescence

Figure 3-2: Gene tag measurements from endogenous Sox2.
(left) Diagram of Sox2 gene tag integration process. The repair template contains an FF4 hairpin and CMV-Zeo cassette flanked by LoxP sites. Double strand breaks were introduced with CRISPR-Cas9 and the repair template was integrated into a fraction of cells with homologous recombination. Cells were selected with zeocin to increase the proportion of cells with successful integrations and clonal population of cells isolated, then the zeocin cassette was removed by introducing Cre into the cells. Since cells initially contained integrated expression of an FF4 sensor, the final cells contain all components necessary for sensing transcription with gene tags. **(A-I)** Density plots of miRNA gene tag output with transcriptional activation and repression or controls. **(A)** Transcriptional repression with targeted dCas9-KRAV of Sox2 results in production of less FF4, resulting in ∼2-fold increased levels of mKate output in the transfected cells (EYFP positive cells). **(B)** Transcriptional activation of Sox2 using targeted dCas9-VP160 increases the production of FF4 which reduces mKate levels by ∼2-fold. **(C)** Transcriptional activation of Sox2 with targeted dCas-VPR increases the production of FF4 which reduces mKate levels also ∼2-fold. **(D-H)** Exclusion of gRNA, a dCas9-fusion, or both, results in minimal changes in mKate output in EYFP positive cells. **(I)** A positive control with delivery of FF4 siRNA with hEF1a-EYFP results in large reduction in mKate output.

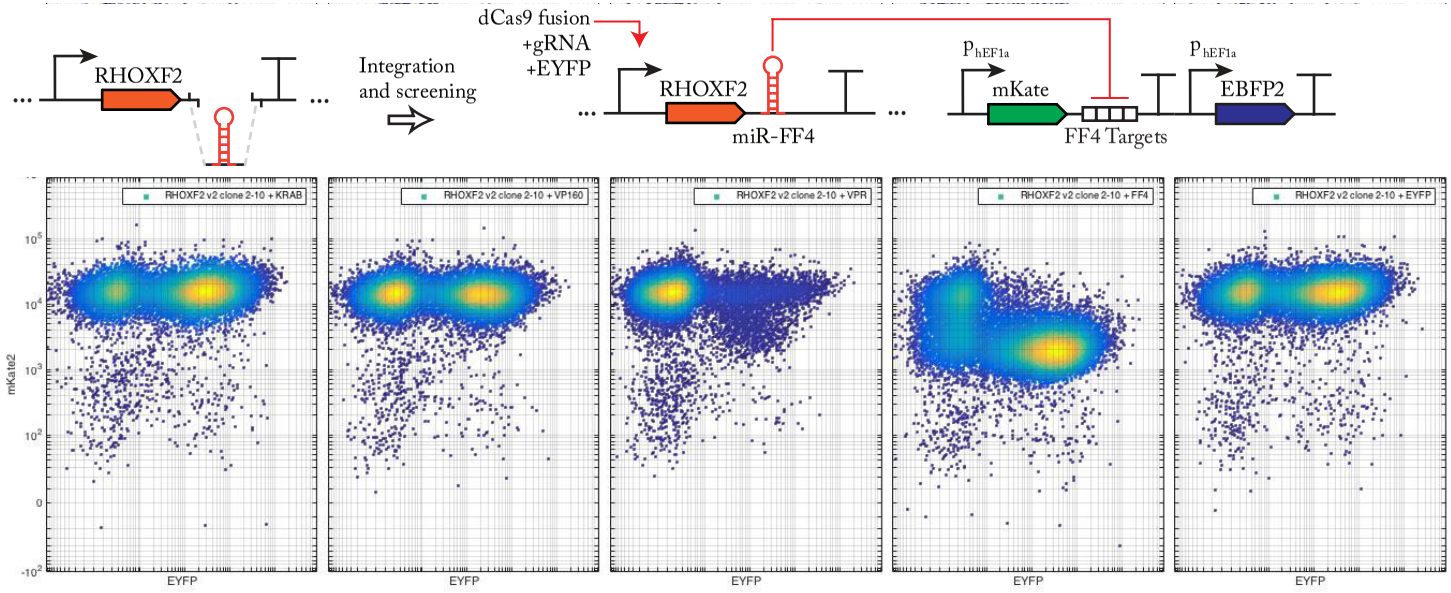

RHOXF2
Integration
and screening
dCas9 fusion
+gRNA
+EYFP
RHOXF2
miR-FF4
P hEF1a
mKate
FF4 Targets
P hEF1a
EBFP2
RHOXF2 v2 clone 2-10 + KRAB
RHOXF2 v2 clone 2-10 + VP160
RHOXF2 v2 clone 2-10 + VPR
RHOXF2 v2 clone 2-10 + FF4
RHOXF2 v2 clone 2-10 + EYFP
mKate2
EYFP

Figure 3-3: miRNA gene tag in RHOXF2.
(top) Integration of the RHOXF2 genetag without antibiotic selection. A similar repair template for RHOXF2 was synthesized and assembled, in which the homology arms for RHOXF2 surround a LacZ cassette with Golden Gate overhangs so any artificial miRNA hairpin can be inserted. For simplicity we inserted FF4 so that the same cells with integrated FF4 sensor could be used. Following CRISPR and repair of cells, clones were isolated and screened for insertion using genomic PCR. **(bottom)** Flow cytometry data with dCas9-KRAB-mediated repression, dCas9-VP160 activation, dCas9-VPR activation, FF4 siRNA, and no effector samples (from left to right). Each sample contains EYFP as a transfection marker. RHOXF2 gene tag shows minimal response to repression with KRAB, perhaps because baseline expression of RHOXF2 is minimal. Response to activation by VP160 was also minimal, but interestingly a subpopulation of cells showed large decreases in mKate expression with VPR, perhaps because VPR strongly activates transcription from endogenous genes. Controls with FF4 siRNA and just EYFP show the expected behavior.

References

96. Weil, T. T., Parton, R. M. & Davis, I. Making the message clear: Visualizing mRNA localization. *Trends in Cell Biology* **20,** 380–390 (2010).

97. Sulc, P., Ouldridge, T. E., Romano, F., Doye, J. P. & Louis, A. A. Modelling toehold-mediated RNA strand displacement. *Biophysical Journal* **108,** 1238–1247 (2015).

98. Siciliano, V. *et al.* Engineering modular intracellular protein sensor-actuator devices. *Nature Communications* **9,** 1881 (2018).

99. Seluanov, A., Mao, Z. & Gorbunova, V. Analysis of DNA Double-strand Break (DSB) Repair in Mammalian Cells. *Journal of Visualized Experiments,* 1–7 (2010).

100. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology* **33,** 543–548 (2015).

101. Gaidukov, L. *et al.* A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Research* **46,** 1–15 (2018).

102. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12,** 326–8 (2015).

103. Song, J. *et al.* RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nature Communications* **7,** 1–7 (2016).

Chapter 4

One-pot optimization of genetic circuits

4.1 Abstract

Recent advances in single-cell and pooled approaches are greatly enhancing our ability to build and test genetic systems for studying and manipulating biology. However, since existing approaches are often complex, time-consuming and expensive, we devised a 'poly-transfection' method that yields multi-dimensional datasets using simple modifications to inexpensive transfection protocols. Using poly-transfection, we demonstrate rapid optimization of useful genetic systems including gene activation by dCas9-fusion and a miRNA classifier for distinguishing cell types.

4.2 Introduction

Projects in synthetic biology typically use a design-build-test-learn (DBTL) cycle to obtain a genetic circuit with the desired behavior.[104,105] Initially, a candidate circuit is designed either using computer-aided design tools or in an ad hoc manner, depending on whether parts have been fully characterized in the context of interest. With each iteration of the cycle, performance of the current circuit is observed and potential improvements to the next generation of circuits are proposed based on suspected failure modes. However, as circuits become increasingly sophisticated, the number of possible failure modes increases and it becomes more difficult to determine the source and nature of failure mode(s) for a given circuit.[106] Moreover the time to complete each DBTL cycle can take on the order of weeks to months depending on the model organism and experimental setup,[107,108] while the gradient descent-like behavior of the DBTL cycle can result in a final design reflective of a local maximum.[109,110] These limitations necessitate methods that quickly explore the full design space possible to a genetic circuit, in order to determine which tuned circuit variants meet the desired specifications, ideally in a single DBTL iteration.[111–113] Current and developing methods to fully explore design spaces are often technically complex, requiring large screens,[114,115] mutagenesis,[88] or preparation of combinatorial libraries for screening[112,116–118]. Ideally researchers would be able to efficiently explore the whole design space using readily available genetic constructs and using simple techniques performed by virtually all life-science labs. In this study, we introduce poly-transfections (separately mixed transfection mixtures to decrease covariance in expression) and computational analysis techniques, which combine to drastically increase the amount of information that can be obtained through transfections.

Our approach - which we term one-pot characterization and optimization - serves

as a way to determine how a given circuit performs with any combination of circuit component concentrations, all in a single tube. Circuit behavior for many possible circuit variants can be derived from a single one-pot dataset and used to make informed design choices for direct testing in an application of interest. We also developed DNA assembly schemes and used machine learning to accelerate the build and learn stages of the DBTL cycle. Our one-pot method, in concert with modern design tools, will accelerate all stages of genetic circuit development.

4.3 Building a miRNA classifier using traditional design processes

To provide a point of reference so that one-pot optimization approaches can be compared to traditional approaches, we provide examples of our initial attempts to optimize a miRNA classifier. We initially tried to improve upon previous classifier designs by using better components than those used previously, namely using transcriptional repressors (e.g. TALER14 and BM3R1) with stronger ability to repress compared to LacI (Fig. 4-1). To test this design, we used a Gateway-Gibson like cloning approach to make classifiers that were encoded on single plasmids. We chose to use a single plasmid design in order to reduce the possibility for variance and error introduced in co-transfections, at the cost of increased cloning time. However, the performance of the new classifier was lacking, with low dynamic range even across a wide range of Dox and TALER14 levels (Fig. 4-2)

We then tested many different variants,* most of which are not interesting enough

*notable flow cytometry files at `/jgam/Data/flow/2015-02-27HSBFF4`; `/jgam/Data/flow/2015-04-15_HSB2_HEK`;

to cover here. The final designs (Fig. 4-3) did show significant improvements in dynamic range compared to the initial TALER14-based classifier, with up to 100-fold differences between HEK293FT and HeLa cells in some cases (Fig. 4-4 and 4-5). However, it is important to note that the time for designing, planning, testing, and iterating took approximately half of a grad student's time over the course of over a year. And even with that significant dedication of time, a functional design is by no means guaranteed and with each iteration it is not known which modifications could yield better designs and which may be dead ends. As a result, whenever a new genetic circuit is conceived, the highly iterative design-build-test-learn cycle is slow and cumbersome and the information learned in each of the iterations may not be very applicable to designing future unrelated circuits. Researchers have recently noted how a reduction in the number of iterations can have drastic effects on accelerating development times,[119] and more fundamental work on infrastructure for forward engineering of genetic circuits via accurate predictions has been shown.[120] Therefore, work on in-depth characterization of parts and accurately modeling their composition into systems will continue to be an important part of synthetic biology. The one-pot approaches we present here represent an accelerated way to perform deep characterization in a single sample and also a method for optimizing whole circuits at once. Using one-pot optimization, we gathered data for optimizing a miRNA classifier that performs better than the iterations shown in this section, all in a single afternoon session of flow cytometry. As a result, future optimization of genetic

---

`/jgam/Data/flow/2015-04-19_HSB1_HEK`;
`/jgam/Data/flow/2016-03-08_JG164-JG170`;
`/jgam/Data/flow/2016-06-05_HS_JG165_199_200`;
`/jgam/Data/flow/2016-06-19_HS_variants`. Plasmid maps correspond to: HSB1; HSB2; JG164; JG165; JG166; JG199; JG200;JG206; JG207; JG208; JG209.

circuits may take as little as 2 weeks including cloning steps, compared to the one year we took using the traditional iterative approach.

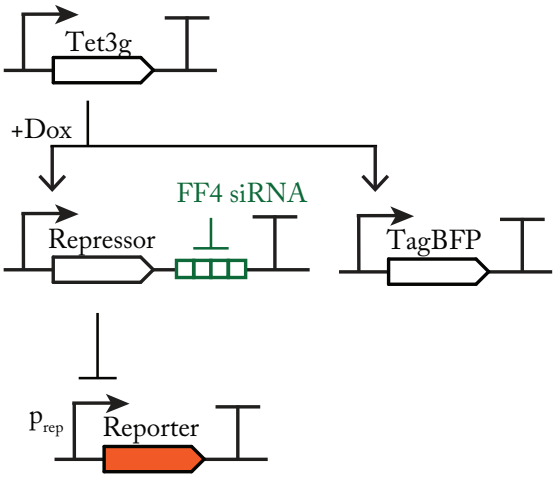


Figure 4-1: Circuit diagram for the initial miRNA high classifier. Having previously demonstrated the proof of concept for multi-input miRNA-based classification of cells in Xie et al.[35], we sought to improve the design by replacing the weak LacI repressor with stronger repressors and also using brighter fluorescent proteins for output (mKate2). In this design, the repressor chosen was TALER14. We also used a Dox responsive transfection marker (TagBFP) in order to better visualize changes in Dox concentration rather than just transfection efficiency. In these experiments we focused on optimizing the miRNA high sensor part of the circuit as we believed that had the greatest potential for improvement.

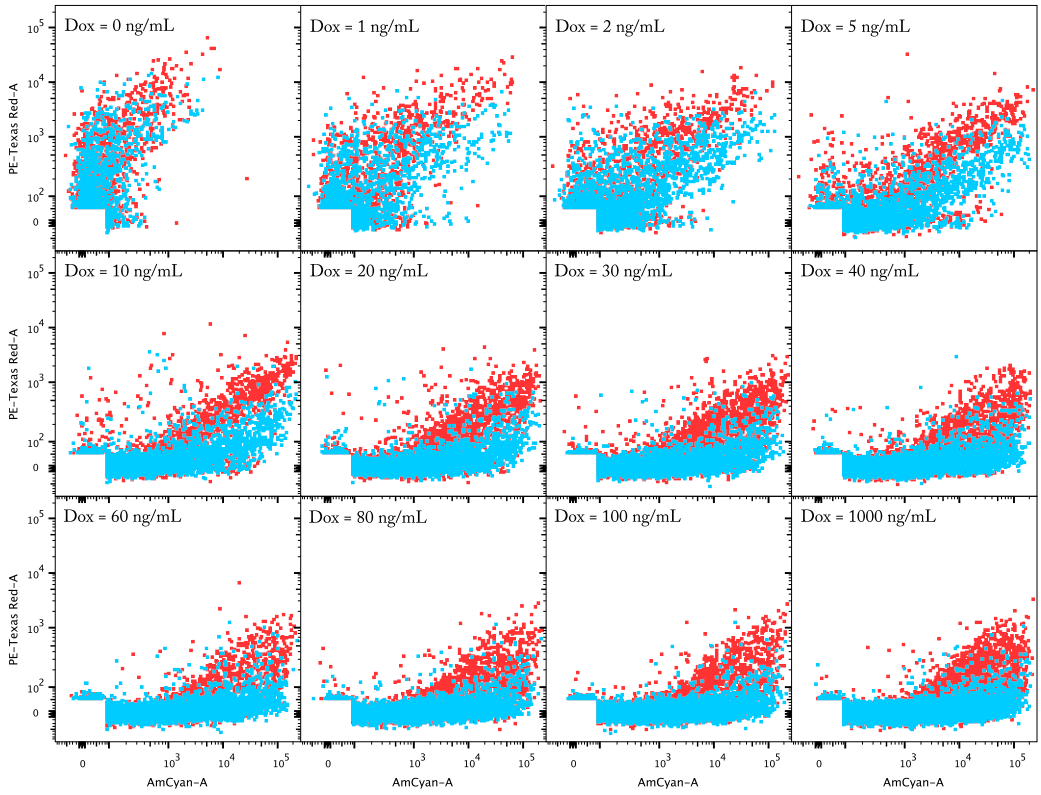


Figure 4-2: Tests of initial miRNA high classifier with TALER14 repressor.

Fluorescence plots of miRNA classifier using TALER14 that senses a single high input (FF4 siRNA codelivered to cells). Blue indicates samples where no input FF4 was delivered and red indicates samples with FF4. We performed a Dox titration, observing the expected increase in TagBFP fluorescence with increasing Dox. Addition of Dox also increased expression of the TALER14, reducing expression of mKate2 to a degree. Though we did observe higher output in the samples with FF4 added, the dynamic range for this classifier was not high enough for use as a practical classifier. This lack of dynamic range was observed across all Dox concentrations, suggesting that little optimization could be obtained by only varying TALER14 expression. We also looked towards using other easier to use and better performing repressors in the future.

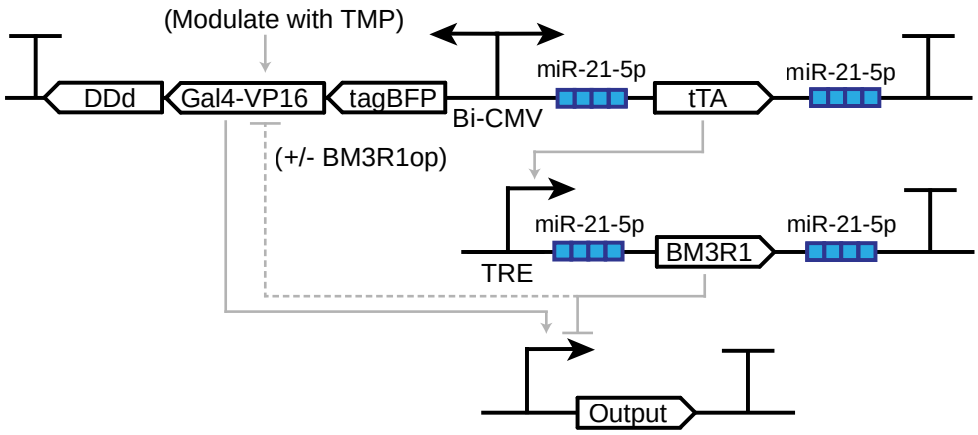


Figure 4-3: Design of miRNA classifier variants.
Following other previous experiments with miRNA classifiers, we tried a design with many variations in topology and expression levels. The design used many more miRNA target sites than previous designs, with targets in the 5' and 3' UTRs of both the BM3R1 repressor and tTA activator, since the Ant/Syn model indicates that this design should increase dynamic range. We used Gal4-VP16 to activate the output, and modulated Gal4-VP16 with the presence (or absence) of a degradation domain (DDd) that can be stabilized by the addition of trimethoprim (TMP). Also, the expression of Gal4-VP16 may be modulated in certain designs that have a BM3R1 operon (BM3R1op) upstream.

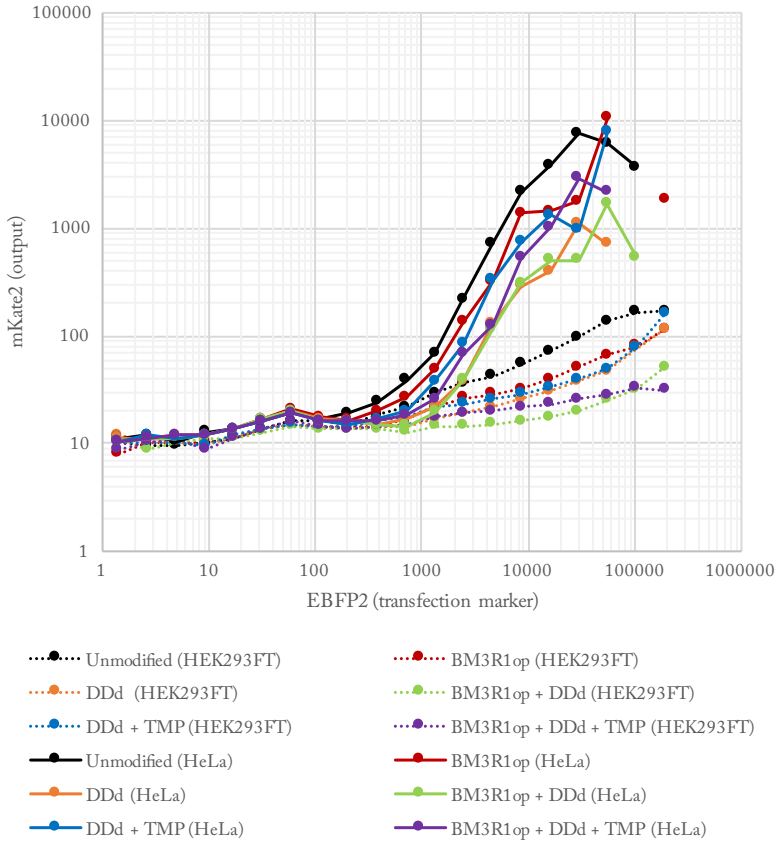


Figure 4-4: Output from miRNA classifier variants.
Output (mKate2) as a function of TagBFP shows much better dynamic range compared to the previous TALER14 design, especially given the fact that the classifier is sensing an endogenous miRNA rather than an siRNA in this case. The classifier senses differences in activity of miR-21-5p, which is high in HeLa cells and low in HEK293FT cells. Background expression in HEK293FT cells (dotted lines) vary across the different variants as expected, with the unmodified classifier showing highest expression and other variants having lower expression, presumably due to the decrease in Gal4-VP16 expression either from the DDd or BM3R1op. The trend in expression in HeLa cells (solid lines) is similar however, meaning that the variants have similar dynamic range and are merely shifting up or down expression in both cell lines.

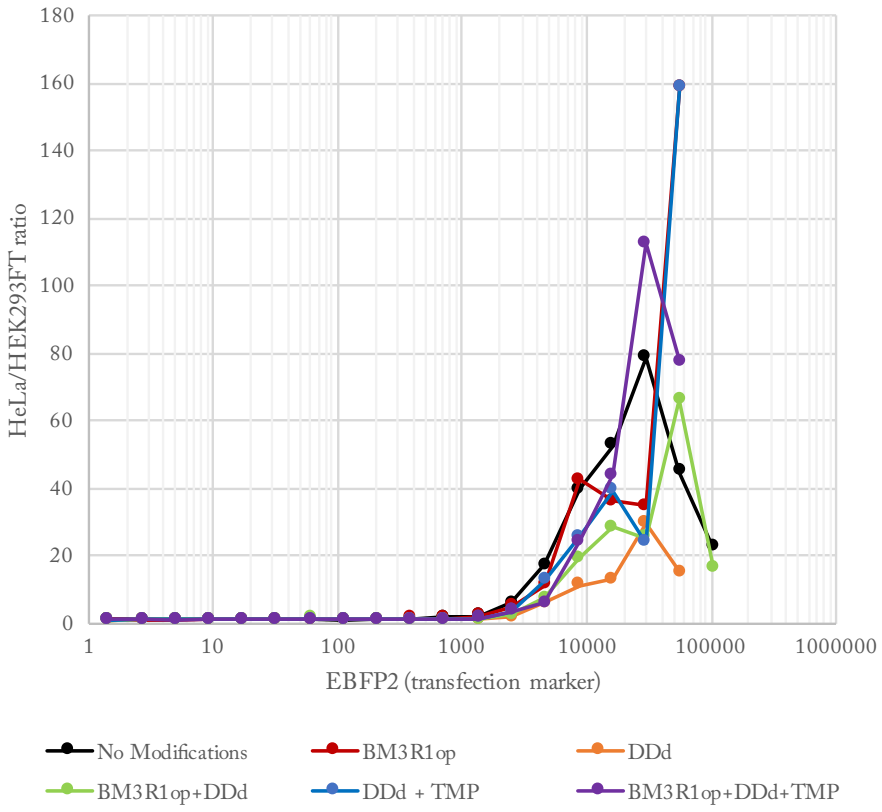


Figure 4-5: Ratios in output from classifier variants.
Comparison of dynamic range as a function of transfection marker shows that most variants have similar dynamic range. The DDd only and BM3R1op+DDd variants appeared to have slightly lower dynamic range. Though dynamic range was improved compared to the TALER14 design, we believed further optimizations could be obtained if expression levels for all components in the classifier could be tuned properly.

4.4 Results

Biologists frequently use transient co-transfections to introduce exogenous DNA, RNA, or protein components into cultured cells to modify cellular signaling or behav-

ior. In many cases, multiple genetic components must be delivered simultaneously. For instance, simultaneous expression of several transcription factors has been used to reprogram cell fate,[121] and both guide RNA (gRNA) and Cas9 are needed for targeted nuclease activity.[122] Often, the concentrations of each component and their relative ratios are important for function; for example, specific ratios of reprogramming transcription factors improve reprogramming efficiency by several fold.[123,124] However, the ratios for each gene-coding component are often chosen based on intuition, trial and error, or coarse optimization, meaning the operation of the system as a whole is likely to be suboptimal. For instance, popular plasmids for CRISPR/Cas9 (e.g. pX330[125]) encode constitutive high expression of both gRNA and Cas9, but recent results show significant gains in editing efficiency after optimizing the ratio of gRNA to Cas9.[126] For even larger and more complex systems, the number of possible stoichiometries grows exponentially, which drastically reduces the likelihood that subjectively chosen expression levels will yield optimal or even functional behavior.[127] Therefore, optimizing these systems currently involves complex and expensive pooled methods[111,112,128] or time-consuming manual approaches.[129] Pooled experiments in mammalian cells include many steps, each of which requiring significant expertise, including: design of pooled DNA, assembly into a uniform library, delivery of the library into cells using virus, and preparation and analysis of high throughput sequencing experiments.[118,130] On the other hand, optimization is more commonly achieved by manually varying component levels in many different independent co-transfection samples, which requires an exponential increase in samples for each component within the system ($\sim 10^{n-1}$ co-transfections for n components; Fig. 4-6). Thus co-transfection experiments are often tedious for smaller systems and infeasible for larger systems such as those being developed in synthetic biology.

Here we introduce a one-pot 'poly-transfection' method which enables broad ex-

ploration of a genetic system's behavior across a wide range of component stoichiometries and in a single sample. Poly-transfections employ the same reagents and genetic constructs as co-transfections, but differ in both the protocol and subsequent data analysis. In this study, we focus on components in the form of DNA plasmids each encoding a particular gene in the system, though poly-transfections can also be applied to other types of molecules like RNA or proteins and possibly those that encode multiple genes. For reference, in a typical co-transfection, plasmids are mixed at a predetermined ratio before the transfection reagent is added. Unfortunately, this protocol results in highly correlated delivery of each plasmid and data that covers only a single stoichiometry per transfected well (Fig. 4-6b,c).[131,132] Moreover, inherent variability in delivery results in expression that varies over several orders of magnitude, making it difficult to make quantitative comparisons unless a constitutively expressed transfection marker is added for normalization.[133] For poly-transfections, we instead mix each plasmid separately with transfection reagent before addition to cells, yielding transfection complexes that each have varying amounts of a single plasmid species. Transfected cells uptake multiple complexes, resulting in decorrelated delivery of each component plasmid (Fig. 4-6d,e).[131,132] Thus for poly-transfections the variability in transfection efficiency is actually advantageous and enables poly-transfections to broadly sample the component 'concentration space'. In order to measure how much of each component is delivered to a cell, we include a distinct transfection markers with each plasmid, such that each fluorescence intensity relates to the concentration of one component in the system (Fig. 4-6a,b,d). When combined with single-cell analysis methods such as flow cytometry, each poly-transfected cell provides an independent measurement of how the system behaves at a specific combination of component concentrations. Since poly-transfections simplify experimental planning and execution to a single sample, significant savings in active experiment

time can be obtained (Fig. 4-6f).

We first determined whether poly-transfections can recapitulate findings from previous studies and co-transfection experiments. For a dCas9-VPR/gRNA transcriptional activation system, we found that increasing gRNA concentration contributes to graded increases in activation, whereas dCas9-VPR exerts most of its effects at a relatively low threshold (Fig. 4-6g). These results are consistent with our co-transfection experiments (Fig. 4-8) and other studies where efficiency of Cas9 editing depended more on gRNA concentration than Cas9 mRNA concentration.[126] Therefore activation efforts with dCas9-VPR (and presumably editing applications) should benefit more from improved delivery of DNA coding for gRNA rather than Cas9. We also used poly-transfection to analyze input-output relationships for several genetic systems, which would enable easier forward engineering of more complex systems like cascades and oscillators in mammalian cells. Genetic systems we tested include translational repression by L7Ae, transcriptional activation by Tet3G and Gal4-VP16, and post-transcriptional repression by artificial miRNAs (Fig. 4-9-4-14). We found good correlation between data from co- and poly-transfections, demonstrating that poly-transfections provide reliable data ($R > 0.93$ in all cases; Fig. 4-9-4-14).

Next, we used poly-transfections to rapidly engineer a difficult-to-optimize genetic system for discriminating cancerous from noncancerous cells.[35] Cancer cells often express the biomarker miRNA-21-5p highly;[134] thus, we built and optimized a cell classifier which produces a genetic output only in the presence of miR-21-5p (i.e. a single-miRNA high classifier). Our classifier is composed of three DNA components and responds to miR-21-5p in the following way: (i) at high miR-21-5p activity, the miRNA degrades the transcript encoding the BM3R1 transcriptional repressor, (ii) degradation of BM3R1 allows transcription from the promoter driving

mKO2 reporter, and (iii) a Gal4-VP16 transcriptional activator is needed for output expression (Fig. 4-7a). Constraining this system to accurately produce output only in miR-21-5p expressing cells requires balancing the ratios of plasmid concentrations for BM3R1, Gal4-VP16, and reporter.

We optimized transfection-based classification of cells via the miR-21-5p classifier by using poly-transfection to identify an optimal DNA ratio of the three classifier components that would specifically distinguish HeLa from HEK293FT cells, which have high and low miR-21-5p activity respectively. The comprehensive nature of poly-transfection data allowed us to easily sample various component stoichiometries in order to better understand how each component contributes to the performance of the classifier. We evaluated classification accuracy at different component ratios by subsampling the 3-dimensional poly-transfection data, designating ratiometric 'trajectories' through component concentration space (Fig. 4-7b), and including only cells that were in close proximity to the trajectory (Fig. 4-15). Subsampling in this manner provided reliable measurements of classification metrics including specificity, sensitivity, and accuracy (Fig. 4-14). Using an optimization method , we identified a well-performing ratio of 10.9 : 1.5 : 1 for Gal4-VP16 : output : BM3R1 plasmids at which HeLa and HEK293FT cells were distinguished with 91% specificity, 62% sensitivity, and 77% accuracy (Fig. 4-7c, Fig. 4-14). We verified this prediction by performing a co-transfection at the corresponding ratio of components and found that the specificity, sensitivity, and accuracy metrics agreed with the values computed from the subsampled poly-transfection (99%, 68%, 84% respectively; Fig. 4-7d). Further comparisons between poly- and co-transfections at 27 different DNA ratios representing titrations of each individual component showed good correlation for specificity, sensitivity, and accuracy. Analysis showed that high classification accuracy was possible with Gal4-VP16 expressed at a wide range of high levels, and

with BM3R1 and output within a narrow band of lower concentrations (Fig. 4-7e). Overall, these data indicate that subsampled data from single poly-transfection samples can be used to quickly evaluate system performance, reducing or eliminating the need to iterate through different physical designs (Fig. 4-14).

Since many future applications of cell-type classifiers will require single-plasmid designs for packaging into viruses or integration into the genome, we applied our findings to construct optimized single-plasmid versions of the miR-21-5p classifier (Fig. 4-7f). Given that relative DNA copy numbers are fixed in a single-plasmid system, we instead tuned component expression ratios to more optimized levels identified above by inserting a repressive upstream open reading frame (uORF) into the 5'UTR of the mKO2 reporter and truncating the CMV promoter driving BM3R1 (Fig. 4-16, 4-17).[135,136] The miR-21-5p classifier optimized in this way showed higher classification accuracy in transfections compared to both an unoptimized classifier (high expression of all components) and a poorly optimized classifier (low Gal4-VP16 expression via truncated CMV) (Fig. 4-7g). Additionally, when we replaced the fluorescent output with the apoptosis regulator Bax, the optimized design showed greater ability to selectively induce apoptosis in HeLa cells compared to the other classifier variants (Fig. 4-7g), and demonstrated similar selectivity compared to other more complex miRNA classifiers.[35,137] Note that we optimized the classifier design according to classification accuracy, though we could potentially optimize according to other combinations of specificity and sensitivity (i.e. points on the ROC curve), depending on if the application is more lenient towards false positives or negatives. This ability to take a single poly-transfection dataset and optimize based on parameters deemed important (even those not initially considered when designing the circuit and experiment) contributes to greater flexibility of our one-pot approach.

Poly-transfections bridge the gap between simple but low information co-transfections

and complex high-information pooled approaches (Fig. 4-18). By taking advantage of innate variations in transfection efficiency, poly-transfections yield high-information data similar to modern single-cell analysis techniques, while also being substantially simpler. Specifically, the increased cost and complexity of FlowSeq methods[128] and droplet microfluidics[127] makes them unsuitable for small- to medium-scale systems (2-4 components) commonly used in biology. Furthermore, our poly-transfection method can be readily extended to newer technologies such as spectral analyzers and mass cytometers (e.g. CyTOF) that enable even higher dimensional datasets than are possible with the 5-color flow cytometers used in this study. Likewise, future work could further extend the number of components that can be measured with poly-transfections by using a combination of RNA barcoding and single-cell sequencing.

Overall our poly-transfection method represents a convenient and powerful one-pot approach for evaluating concentration-dependent relationships for diverse experimental systems, from characterization of basic components to optimization of complex multi-component designs. In addition to Cas9 and classifier systems shown here, we anticipate that poly-transfections will aid biologists in extracting deeper information from genomic perturbations, designing reporters with higher signal to noise ratio, and building tools to more robustly modulate gene expression, among other applications. Therefore, while poly-transfections are immediately applicable to fields such as synthetic biology where optimization of complex multi-component systems is already essential, we envision that the ease of our method will make such one-pot strategies more feasible to a wide range of fields within biology.

4.5 Poly-transfection characteristics

We ran preliminary experiments to determine the characteristics of poly-transfections, namely whether the covariance typically observed in co-transfection could be reduced in order to obtain cells with more disparate transfection levels of different DNA molecules. For the purposes of this study, we examined whether poly-transfection behavior would be suitable for up to five-dimensional circuit characterization, which would allow for up to 4 circuit parts in addition to one output. We chose to examine up to five dimensions since the laser/filter settings for most available flow cytometers limits the use of fluorescent proteins to around five, further fluorescent proteins would result in significant spectral overlap (Fig. 4-23).[138] We observed significantly reduced covariance with poly-transfections compared to co-transfections (Fig. 4-6e-f) sufficient for at least 4-component circuits to be analyzed with one-pot methods.

We estimated experiment times for co- and poly-transfections to determine how much time could possibly be saved through one-pot methods. For co-transfections, estimates of active experiment time (time spent planning, performing experiments, and analyzing data; incubation periods not included) were calculated based on the following times: 1hr + 1.25min/sample (planning), 1hr + 1.25min/sample (transfection), 1 hr + 2min/sample (flow cytometry), 1 hour + 0.5 min/sample (analysis). With 10^{n-1} samples for n circuit components we obtain $4 + 0.083 \times 10^{(n-1)}$ hr. For poly-transfections the following times were used: 1hr (planning), 1hr (transfection), 1hr + $s^{(n-4)}$hr (flow cytometry), 1hr (analysis). Where s is the scale factor for amount of time needed per dimension for flow cytometry, which is dependent on transfection efficiency and distribution and sparsity of sampling tolerated. The ideal poly-transfection will have $s = 10$ similar to co-transfections, while in the worst case 100-fold more cells would need to be collected per dimension. With sparse sampling

$s < 10$ could be possible with the example shown in Figure 4-6d at s=3. We anticipate that recovering information from sparsely sampled data could prove essential for testing large genetic circuits, though this is an area of research we have reserved for future work. We used n-4 in the exponent since we empirically found that one hour of time was required for the 4-part circuit and significantly less for the smaller circuits. The final equation is then $4 + s^{(n-4)}$ hr.

The inherent bias in poly-transfection expression distributions (which are also present in co-transfections) could hamper efforts to sufficiently and evenly sample higher order parameter spaces. We have also tested modifications to increase the efficiency of poly-transfections. We examined whether poly-transfection distribution could be influenced by the amount of each DNA complex transfected. Indeed in a two-component circuit, we could shift the location of double-positive cells by reducing the amount of DNA in one of the complexes and replacing with non-expressing DNA (Fig. 4-20). This change could allow more efficient exploration of parameter space by using separate poly-transfections with different distributions that complement lower density regions that might be present in a single poly-transfection. For example, a 3-input 1-output system could utilize four poly-transfections, considering that there are three lower density regions in the 3D input space. Such a compromise would result in linear increases in the number of required samples per experiment (n samples required for n dimensions), still a marked improvement compared to the exponential increase with co-transfections ($\sim 10^{n-1}$ samples required). Alternatively, we propose that the dimensionality of the experiment could be reduced by pre-mixing some parts together before the poly-transfection. This could be helpful when previous small scale characterizations of sub-circuits have already indicated the ideal ratio for some of the parts to be used. Thus experiments could be hierarchical; different parts of the circuit could be tested in parallel with separate poly-transfections to determine ideal

ratios for some of the parts, then a final poly-transfection with those parts pre-mixed in order to determine the best relative ratios for the final circuit.

We also tested whether the poly-transfection method was flexible in the types of transfection reagents that could be used, given that many different reagents and protocols exist. We saw that in general most transfection reagents give similar distributions covering the available parameter space with some minor variation [that could perhaps be attributed to variations in transfection efficiency] (Fig. 4-22). We also tested whether sequential transfection utilizing a mixture of electroporation followed by chemical transfection. Again we saw that parameter space was explored with this method, though with more bias towards the electroporated plasmid, likely due to higher transfection efficiencies achieved by that method (Fig. 4-21).

4.6 Activation by Gal4-VP16

To demonstrate the advantages of one-pot method in characterization of commonly used genetic parts, we characterized several regulatory modules that act at various levels of the central dogma. We characterized non-inducible (Gal4-VP16) and inducible (Tet3G, Clontech) transcriptional activation modules, mRNA degradation modules (synthetic miRNAs) and a translational repression module (L7Ae). Each module characterization required a single sample (rather than multiple titration curves), and two transfection complexes.

We began by characterizing the behavior of Gal4-VP16, a transcriptional activator of gene expression.[139] In our system (Fig. 4-27a), Gal4-VP16 is produced constitutively by a CMV promoter. iRFP720 is produced constitutively on the same plasmid as Gal4-VP16 and serves as a transfection marker indicating the relative concentration of Gal4-VP16 delivered to a cell. On a separate plasmid, an mKO2 fluorescent

output is placed under control of a UAS-containing promoter that is activated by Gal4-VP16. This second plasmid contains a constitutively expressed mNeonGreen transfection marker, which indicates the concentration of reporter plasmid delivered to each cell. Input parameter space, the plane defined by Gal4-VP16 (∼iRFP720) and reporter (∼mNeonGreen), was sufficiently explored in a single poly-transfection (Fig. 4-27b). Output expression as a function of each tunable parameter can be given as a 2D heatmap, with colors defined by output fluorescence at each input bin (Fig. 4-27c). The heatmap indicates that, as would be expected from a well-behaved activator/promoter pair, output is only expressed when both plasmids are present at high concentration. In other words, minimal leakiness in output is observed when reporter concentration is high but Gal4-VP16 concentration is low, a common failure mode for inducible promoters. We produced corresponding surface and line plots, where different concentrations of reporter plasmid are indicated by colors (Fig. 4-27d). Across all reporter bins, we observed a sigmoidal response of output in response to increases in Gal4-VP16, indicative of typical transcriptional activation. When the data are instead grouped by Gal4-VP16 amount, the out transfer curve shows a more linear response to reporter amount but only at high Gal4-VP16 concentration (Fig. 4-27e). These results demonstrate that in our system, when there is sufficient Gal4-VP16 for activation, output can be further increased by increasing the concentration of reporter plasmid. We conclude that the concentration of reporter can be used to 'set' the maximal amount of output to that desired in a given application, while Gal4-VP16 can be used to switch between low and high output levels.

4.7 Characterization of Tet3G system

Tetracycline-based induction of transcriptional activation/repression by the *E. coli* protein TetR and its derivatives has been widely used to tune gene expression both in output level (by choosing different tetracycline concentrations) and in time. Dox works by changing the conformation of TetR to abolish (wild-type TetR) or enable (reversed TetR, or rTetR[140]) binding to its target DNA operator (TetO). In mammalian cells, rTetR can be fused to a *trans*-activation domain to drive Doxycycline-induced gene activation from promoters with the TetR-responsive element (TRE), an array of TetO binding sites upstream of a minimal promoter; this system is referred to as the reversed tet *trans*-activator (rtTA). Given rtTA's use in various biological studies ranging from initiating cell reprogramming, differentiation, and specific gene expression, we wanted to test how the amount of rTetR protein and its target promoter affected the observed Dox induction curve. This would help to understand how changing cellular conditions might propagate to affect the Dox curve, and thus the ability to predictably control gene expression.

To study rtTA's activation function and compare the poly- and co-transfection methods, we performed side-by-side experiments with paired experimental conditions. First, HEK-293FT cells we poly-transfected with two complexes, (1) hEF1a-driven Tet3G (Clontech) with a hEF1a-driven mNeonGreen reporter and (2) TRE-driven mKate2 with a hEF1a-driven TagBFP reporter (Fig. 4-10a). In parallel, we co-transfected various wells of HEK-293FT cells with the same paired plasmids at various ratios, keeping the same ratio between the reporters and the concomitant TetR/TRE plasmids. Both poly- and co-transfected samples were analyzed with flow cytometry.

Data from the co-transfected samples were pooled together to create a similar

two-dimensional data space comparable to poly-transfections, which we could then bin in the exact same manner for both methods (Fig. 4-10b). Similar to the Gal4-VP16 experiment, the 2D grid of bins allows delineation between an increase in the output plasmid (here marked by TagBFP) and an increase in the Tet3G protein level (here marked by mNeonGreen). Median mKate2 expression levels in each bin were calculated and plotted in Fig. 4-10c. The 2D profiles for the poly- and co-transfected samples were overall very similar, with the most noticeable differences being at lower TetR/mNeonGreen levels.

We then focused in and compared the effect of increasing Tet3G on the Dox-induction curves at various selected TagBFP bins (Fig. 4-10d). Overall, the poly- and co-transfection data agreed very well for all bins except at high-TagBFP and low-mNeonGreen levels, possibly due to high sensitivity in the system. We found that relatively little Tet3G is needed to maximize protein expression. However, we noticed that increasing expression of Tet3G did not monotonically increase promoter output, but rather caused a decline after reaching a peak output level. We suspect that this can be attributed to self-squelching by the *trans*-activation domain in Tet3G[141]. These results demonstrate the utility of understanding how all components to a system contribute to knowledge of function. With the given data, we can now understand how to maximize gene expression and fold-inducibility from the rtTA system.

4.8 Characterization of artificial miRNA for use as synthetic biology parts

We tested the ability for poly-transfections to simplify the characterization of orthogonality between different parts. Here we analyzed a two-part system where a production of an artificial miRNA sequence[142] represses output of corresponding miRNA sensor - a fluorescent reporter bearing miRNA target sites (Fig. 4-11a).[1,35] EYFP served as a marker for relative concentration of artificial miRNA, while EBFP2 served as a marker for the sensor concentration. We determined that input parameter space (EBFP2 vs. EYFP) was sufficiently covered by the poly-transfection (Fig. 4-11b). Plotting sensor output fluorescence (∼mKate2) as a function of miRNA concentration (∼EYFP) shows that, as expected, output mKate2 is reduced at high miRNA concentration relative to EBFP2 which contains no target sites (Fig. 4-11c). A graph of poly-transfection data binned by miRNA concentration then binned by transfection marker demonstrates the performance of the miR-FF5 sensor at different miRNA concentrations (Fig. 4-11d). As miRNA concentration is increased, mKate2 fluorescence is decreased further compared to EBFP2 and in a threshold-like manner as previously observed from similar sensor constructs.[44]

We then demonstrated that detailed information about miRNA orthogonality could be obtained from poly-transfection data, at a reduced number of transfections compared to co-transfections. We conducted poly-transfections using all possible combinations of eleven different miRNA and sensor sequences, resulting in 11*11=121 separate poly-transfections (Fig. 4-11e). To obtain somewhat similar data using co-transfections would have required approximately 10*11*11=1210 transfections, since around 10 co-transfections are required to adequately cover a 2 dimensional

parameter space (Fig. 4-6d). In this poly-transfection experiment, samples were prepared in only two 96-well plates, allowing both transfections and flow cytometry to be conducted in fewer than two hours each with the appropriate high throughput equipment. Of note is that only $11 + 11 = 22$ lipix-DNA complexes need to be mixed for poly-transfections, compared to $11 \times 11 = 121$ complexes for co-transfections.

The results from the orthogonality test of artificial miRNAs show that as miRNA concentration is increased from bin 1 to bin 6, repression of the corresponding sensor is increased with minimal effects on sensors bearing other miRNA target sites. Specifically miR-FF5 and miR-FF6 showed the greatest degree of repression, indicating that these artificial miRNAs are the best choices when trying to maximize dynamic range (Fig. 4-11e and Fig. 4-13). There was one example of significant off-target effects, where miR-SHC007 repressed the sensor for FF6, though on further examination we found that the FF6 target contained a match for the seed sequence of miR-SHC007, since both were derived from the same luciferase sequence. We also compared co-transfection and poly-transfection data by combining all data from co-transfections into a poly-like dataset and analyzing both datasets in a similar fashion. For both tested miRNAs miR-FF4 and miR-FF5 the behavior measured from co-transfections and poly-transfections was similar (Fig. 4-12)

4.9 Characterization of translational repressor L7Ae

We evaluated poly-transfection for its ability to characterize other RNA-level modules such as translational repression by L7Ae. L7Ae has been used in genetic circuits to tightly bind to its cognate k-turn RNA motif and block downstream translation.[143] We therefore test a two-part system where L7Ae expression represses expression of a reporter containing 2xkturn encoded in its 5' UTR. L7Ae was expressed constitu-

tively along with mKO2 as a marker, while the reporter (2xkturn-mNeonGreen) was expressed constitutively with tagBFP as marker (Fig. 4-9a). Fig. 4-9b shows output mNeonGreen fluorescence of cells within this two-dimensional parameter space and the corresponding surface plot is depicted in Fig. 4-9c. As expected, we see that fluorescence increases in response to increasing amounts of reporter but decreases in response to increasing amount of L7Ae. Also, when compared to a concatenated set of 11 co-transfected samples that explored the same parameter space, the poly-transfection data were similar.

4.10 Building up to a cell-type classifier responding to miR-21-5p

We scaled up our efforts towards characterizing miRNA high sensors which contain either three or four tunable parameters and one output. Our high sensors are based on previous designs, with inclusion of a stronger transcriptional repressor and miRNA target sites in both the 5' and 3' UTRs.[35] In the circuit design, an endogenous miRNA represses a repressor - in this case BM3R1 - which represses the output, such that output is produced when miRNA activity is high. We also included a Gal4-VP16 activation of output in order to more easily tune circuit behavior by modulating Gal4-VP16 concentration. This is in contrast to a miRNA low sensor where output is produced when miRNA activity is low (Fig. 4-11). We gradually built up toward a miRNA high sensor starting from the Gal4-VP16 module (Fig. 4-27, 4-25a-b) and adding the BM3R1 transcriptional repressor (Fig. 4-25c-d) and miRNA target sites (Fig. 4-25e-g). We tested designs in two cell types: HEK293FT which exhibits low activity of miR-21-5p, and HeLa which exhibits high activity. Thus our

immediate goal was to identify areas in parameter space where output was minimal in HEK293FT cells and high in HeLa cells. Starting with the Gal4-VP16 activation of the output, output fluorescence as a function of the two input parameters was high when both the reporter and Gal4-VP16 concentration are high (Fig. 4-25b). This behavior was observed for both HEK293FT and HeLa cells, as shown by a low HeLa/HEK293FT ratio in output across all of parameter space. Addition of BM3R1 to transcriptionally repress the output added another dimension to the parameter space, such that output was only expressed when reporter was high, Gal4-VP16 was high, and BM3R1 was low (Fig. 4-25d). Again we observed that output was similar in HEK293FT and HeLa cells. Finally, addition of miR-21-5p target sites to BM3R1 allowed for cell-type specific expression (Fig. 4-25e). HEK293FT cells exhibit negligible miR-21-5p activity so the output behavior in those cells was similar to the previous case lacking target sites (Fig. 4-25d,f) On the other hand since miR-21-5p has high activity in HeLa cells, the region of parameter space where HeLa expressed high output was expanded, resulting in a band of parameter space where output was high in HeLa but not HEK293FT cells (Fig. 4-25f). Furthermore, calculation of specificity, sensitivity, and accuracy showed a similar result, where specificity was high where HEK293FT output was low, sensitivity was high where HeLa output was high, and accuracy was high in a band of parameter space where both were true (Fig. 4-25g). Therefore, high sensor designs should be tuned such that expression of each of the inputs is located around the band where accuracy is highest. We also tested a 4-component design where BM3R1 was activated by the transcriptional activator tTA, though minimal benefit in performance was observed for that variant so further testing was performed in the simpler 3-plasmid circuit (Fig. 4-26-4-28).

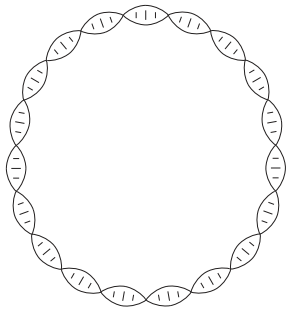
For the 3-component design, optimization of expression ratios for all components was initially performed using the Nelder-Mead simplex algorithm (fminsearch

in MATLAB) with subsampling the poly-transfection and calculating classification accuracy, resulting in a ratio of 435ng: 227ng : 45ng for Gal4-VP16 : output : BM3R1 and an accuracy of 72%. Then we manually tuned the DNA amounts to find a more robust region of parameter space (435ng : 60ng : 40ng) with better classification accuracy of 77%. With a more advanced optimization strategy, it is likely that robust classification could be optimized without manual intervention. Such advances will likely be required in the future with larger more complex systems.

4.11 Additional analysis

We anticipate several challenges in applying poly-transfections to increasingly larger genetic circuits. First, the transfection biases (which are inherent to all transfections) result in data that is not uniformly sampled across the parameter space. This bias towards singly- and doubly-transfected cells can be corrected for by subsampling the data to normalize the transfection spread (Fig. 4-24) Second, with greater dimensionality, datasets become more challenging to analyze and interpret into a set of design rules. Consequently, we anticipate that techniques like machine learning will play a role in simplifying results into sets of design rules that are intelligible to designers. We tested the ability for random forest regression to determine the most important parameters in the 4-component miRNA high sensor circuit. We chose random forests due to their simplicity and ability to generate feature importances. Our analysis indicated that Gal4-VP16 was the most important parameter for obtaining functioning high sensors, followed by the reporter and ratios between repressing (tTA, BM3R1) and activating (Gal4-VP16, reporter) circuit components (Fig. 4-29). Third, the broad excitation and emission spectra of fluorescent proteins limit the number of components that can be simultaneously measured to the

number of lasers in current flow cytometers. One potential solution is to instead use spectral analyzers, which enable better deconvolution of overlapping spectra by measuring at many wavelengths simultaneously. Additionally, mass cytometry (e.g. CyTOF) paired with epitope tagging of circuit components could offer another approach, though staining will introduce further varability to the measurements. A final limitation is given by the number of cells that can be measured per unit time. Current cytometers can measure a few thousand events per second, but as the dimensionality of an experiment increases, the number of events measured must similarly increase to maintain the same level of coverage (Fig. 4-6e) A combination of undersampling (and subsequent data reconstruction) and advances in flow cytometry instrumentation may be necessary to keep pace with large circuit testing.



This page intentionally left blank to keep captions near figures.

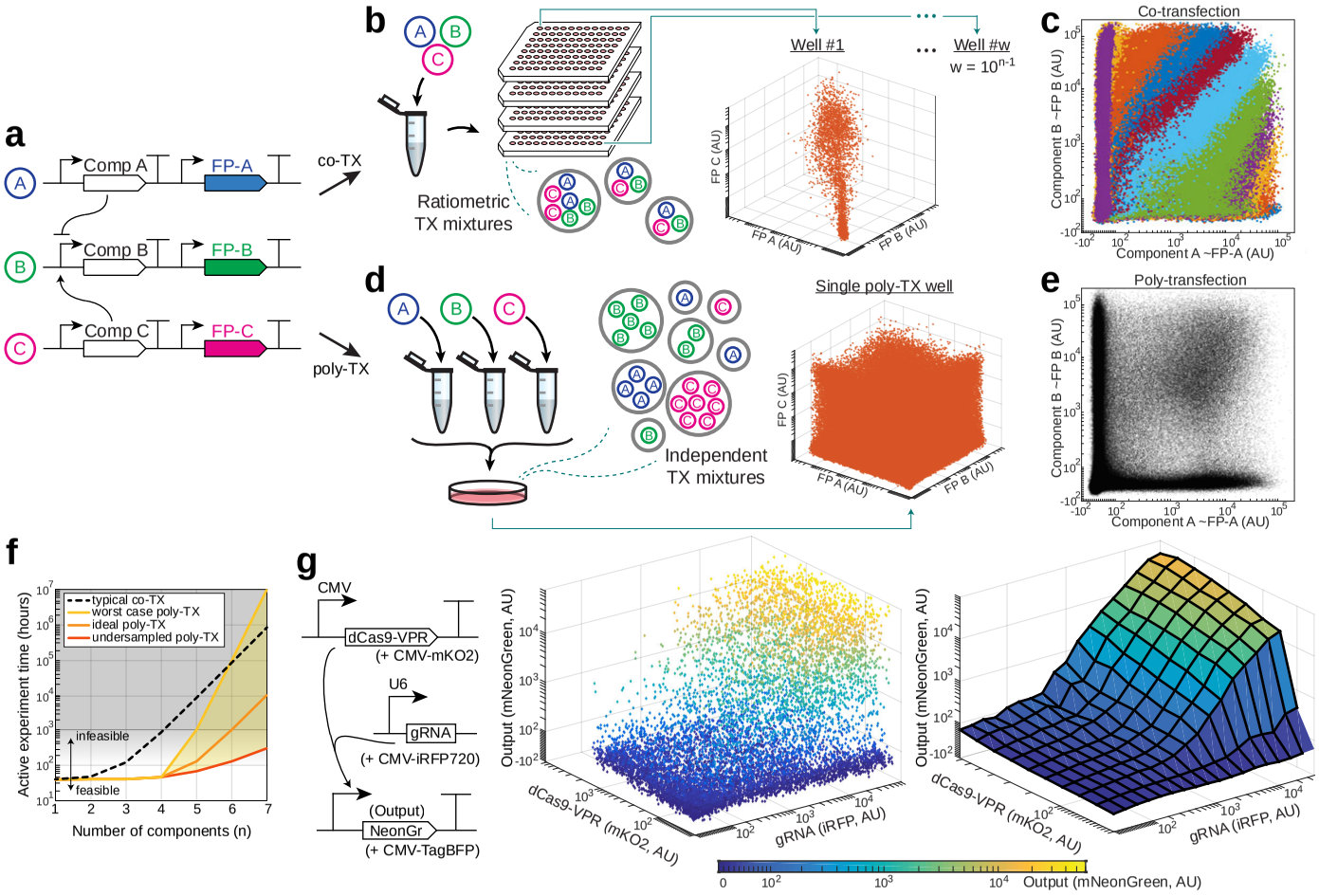
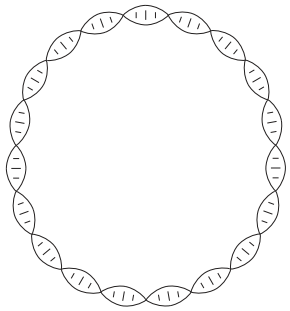

a
Comp A
FP-A
Comp B
FP-B
Comp C
FP-C
co-TX
poly-TX
b
Ratiometric
TX mixtures
Well #1
Well #w
w = 10^n-1
FP C (AU)
FP A (AU)
FP B (AU)
c
Co-transfection
Component B ~FP B (AU)
Component A ~FP-A (AU)
d
Independent
TX mixtures
Single poly-TX well
e
Poly-transfection
f
typical co-TX
worst case poly-TX
ideal poly-TX
undersampled poly-TX
infeasible
feasible
Active experiment time (hours)
Number of components (n)
g
CMV
dCas9-VPR
(+ CMV-mKO2)
U6
gRNA
(+ CMV-iRFP720)
(Output)
NeonGr
(+ CMV-TagBFP)
Output (mNeonGreen, AU)
dCas9-VPR (mKO2, AU)
gRNA (iRFP, AU)

Figure 4-6: Comparison of co-transfection and poly-transfection methods. **(a)** Diagram of genetic system with three DNA plasmid components (Comp A-C). This example corresponds to a system for miRNA-based cell classification, where Comp A, B, and C are BM3R1 with miR-21-5p target sites, mKO2, and Gal4-VP16 respectively (Fig. 4-7). In a typical co-transfection, constitutive expression of only a single transfection marker fluorescent protein (FP) is needed, though here we include three (FP A-C) so that coverage of 'concentration space' can be visualized for both co- and poly- transfections. **(b)** Co-transfection (co-TX) workflow. Each DNA component is premixed at defined ratios before transfection reagent is added, forming complexes containing DNA and transfection reagent (gray circles). Since DNA is premixed, all cells within a well are transfected with a fixed ratio of each component. While the number of total plasmids delivered to each cell varies, the ratio of each remains similar. Therefore only a single ratio of concentration space is explored per well, as indicated by the scatter plot in 3-dimensional concentration space. **(c)** Number of co-transfection samples required to sample concentration space. With a system containing two-components (e.g. L7Ae repression, Fig. 4-9), a scatter plot of 2-component co-TX samples with DNA premixed at varying ratios indicates that $\sim 10^{n-1}$ samples are required for a circuit containing n parts. Different colors indicate samples with different DNA ratios. **(d)** Poly-transfection (poly-TX) workflow. Each component is mixed with transfection reagent separately to form distinct transfection complexes. Transfected cells take up multiple complexes and express varying amounts of each component, filling concentration space in a single sample as shown in the 3D scatter plot. **(e)** Coverage of 2-dimensional concentration space with single poly-transfection sample. **(f)** Estimates of active researcher experiment time needed for co- and poly-transfections. Co-transfections require a fixed setup time plus hands-on/analysis time for each sample. Since the number of samples increases exponentially with the number of components within a genetic system, experimental time also increases exponentially. As poly-transfections require only one sample, experiment time depends mostly the number of cells to be collected, resulting in more efficient use of time. With greater than four circuit components, experiment time may scale depending on transfection efficiency and tolerance for undersampling; examples are indicated by orange lines and the range is indicated by yellow shading. Feasible and infeasible experiment times are indicated by white and gray backgrounds respectively. Equations for estimated experiment time can be found in section 4.5. **(g)** Characterization of transcriptional activation of a fluorescent reporter by dCas9-VPR. Contributions to activation by dCas9-VPR and gRNA can be measured simultaneously with our one-pot method. Shown are the circuit

diagram, subsampled scatter, and a surface plots of output as a function of gRNA and dCas9-VPR at intermediate TagBFP transfection marker levels (10^3 to 10^4 AU). Surface plots indicate the medians of mNeonGreen within each 2D bin and data at all TagBFP levels can be found in Fig. 4-30.



This page intentionally left blank to keep captions near figures.

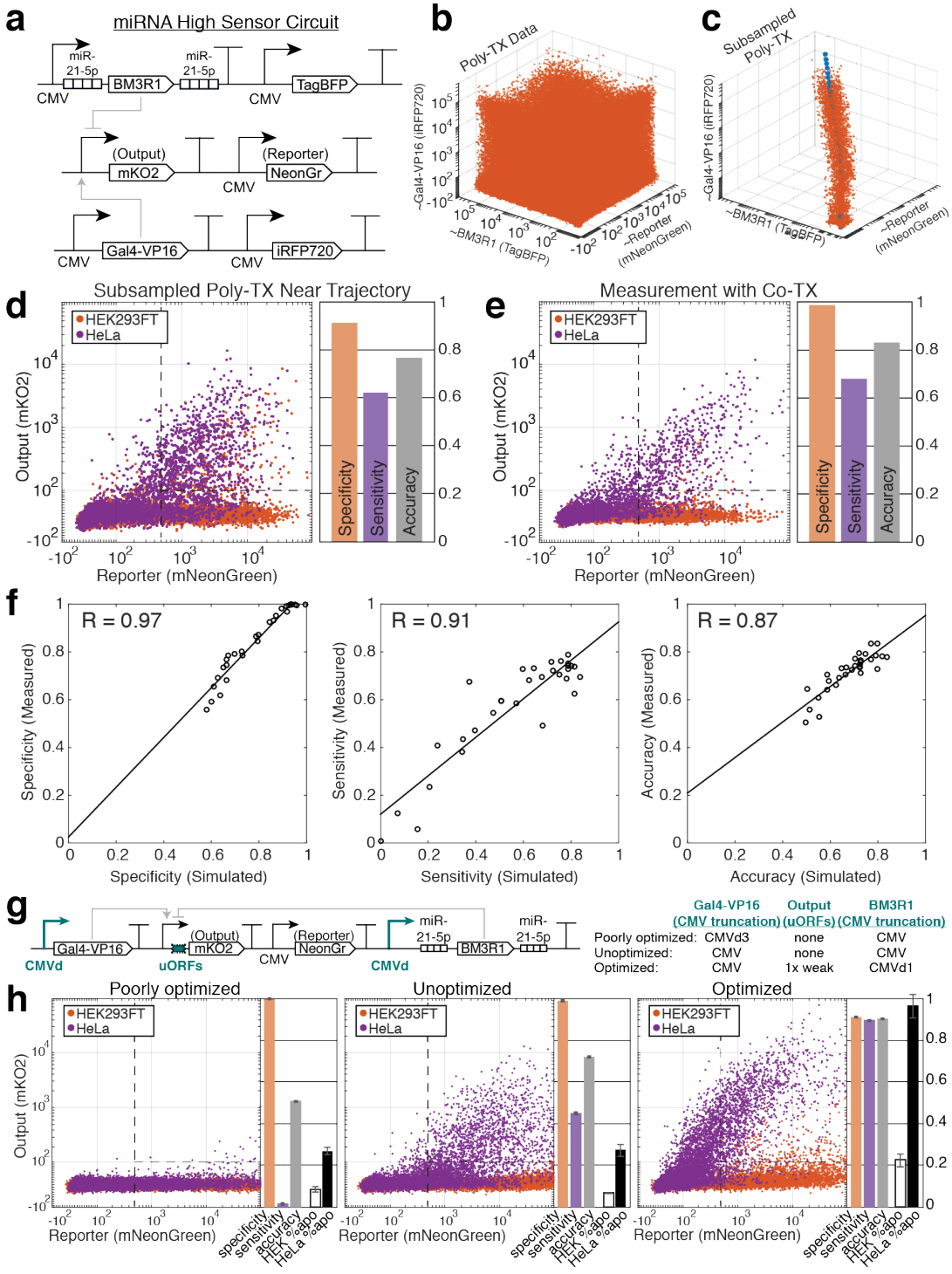
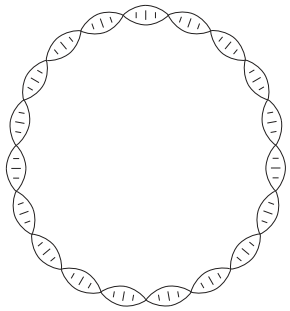

a
miRNA High Sensor Circuit
miR-21-5p
BM3R1
CMV
TagBFP
(Output)
mKO2
(Reporter)
NeonGr
Gal4-VP16
iRFP720
b
Poly-TX Data
~Gal4-VP16 (iRFP720)
~BM3R1 (TagBFP)
~Reporter (mNeonGreen)
c
Subsampled Poly-TX
d
Subsampled Poly-TX Near Trajectory
HEK293FT
HeLa
Output (mKO2)
Reporter (mNeonGreen)
Specificity
Sensitivity
Accuracy
e
Measurement with Co-TX
f
R = 0.97
Specificity (Measured)
Specificity (Simulated)
R = 0.91
Sensitivity (Measured)
Sensitivity (Simulated)
R = 0.87
Accuracy (Measured)
Accuracy (Simulated)
g
CMVd
uORFs
Gal4-VP16 (CMV truncation) Output (uORFs) BM3R1 (CMV truncation)
Poorly optimized: CMVd3 none CMV
Unoptimized: CMV none CMV
Optimized: CMV 1x weak CMVd1
h
Poorly optimized
Unoptimized
Optimized
specificity
sensitivity
accuracy
HEK %apo
HeLa %apo

Figure 4-7: Rapid optimization of a miRNA classifier using poly-transfection.
(a) miRNA classifier to be optimized. In cell lines where miR-21-5p activity is high (e.g. HeLa), BM3R1 is repressed by the miRNA, derepressing output mKO2 in the presence of Gal4-VP16. **(b)** Poly-transfection (poly-TX), co-transfection (co-TX), and subsampled poly-TX data in three component dimensions. Poly-TX data shows sufficient coverage in three input dimensions such that it can be subsampled according to a ratio trajectory (blue) to yield data similar to co-TX. For poly-transfection, at least 1.5 million cells were assayed. **(c)** Classification in HEK293FT and HeLa cells based on the subsampled poly-TX. Scatter diagram of mKO2 output is plotted as a function of reporter marker fluorescence (mNeonGreen). Vertical dotted lines designate the threshold for cells designated as transfected (mNeonGreen = 5×10^2 AU) and horizontal lines designate the threshold for determining if cells express high or low output (mKO2 = 10^2 AU). Poly-TX data was subsampled according to a trajectory corresponding to Gal4-VP16 = 435 ng of DNA, reporter = 60 ng, and BM3R1 = 40 ng. For HEK293FT cells (red), output remained low in most cells across all reporter levels, while in HeLa cells (purple) a majority of transfected cells expressed high output (specificity = 91%, sensitivity = 62%, accuracy = 77%). **(d)** Cell classification in a co-TX experiment at same component ratio as in (c). Similar to the subsampled poly-TX, output was low in HEK293FT cells and high in HeLa cells (specificity = 99%, sensitivity = 68%, accuracy = 84%). **(e)** Classification accuracy as a function of component amounts. For either co-TX or subsampled poly-TX data, accuracy showed similar dependency on BM3R1, reporter, and Gal4-VP16 amounts. DNA amounts correspond to those in (c) with one of the components varied at a time. For co-transfections, the 27 different ratios used are listed in Table 4.2 **(f)** Circuit diagram for single-plasmid miRNA high sensors. Tunable parts include either upstream open reading frames (uORFs) or CMV truncations (CMVd), both of which tune down expression to a defined degree (Figs. 4-16, 4-17).[135,136] Also given is the table of CMV truncations and uORFs used in single-plasmid constructs. The optimized circuit encodes high expression of Gal4-VP16 and reduced levels of output and BM3R1, the unoptimized circuit encodes high expression of all three, while the poorly optimized circuit encodes low expression of Gal4-VP16 and high expression of output and BM3R1. **(g)** Performance of an optimized single-plasmid high sensor compared to unoptimized and poorly optimized variants. Scatter plots of output (mKO2) as a function of transfection marker (mNeonGreen) in the poorly optimized high sensor (left) show low output in both HEK293FT and HeLa cells, resulting in high specificity (100%) but inferior sensitivity (2%) and classification accuracy (51%). The unoptimized high sensor (middle) shows low output in HEK293FT

(specificity = 99%), but only high output in a fraction of HeLa cells (sensitivity = 45%), and an overall lowered ability to classify cells (accuracy = 72%). For the optimized high sensor (right), output remained low in most transfected HEK293FT cells (specificity = 91%) and high in most transfected HeLa cells (sensitivity = 90%), resulting in a generally high classification accuracy (accuracy = 90%). Results from the optimized classifier with Bax as output show much higher degree of killing in on-target HeLa cells (black) with low killing in off-target HEK293FT cells (white). Apoptosis percentages are relative to positive controls with constitutively expressed Bax. Error bars for bar charts indicate standard deviations for technical triplicates.



This page intentionally left blank to keep captions near figures.

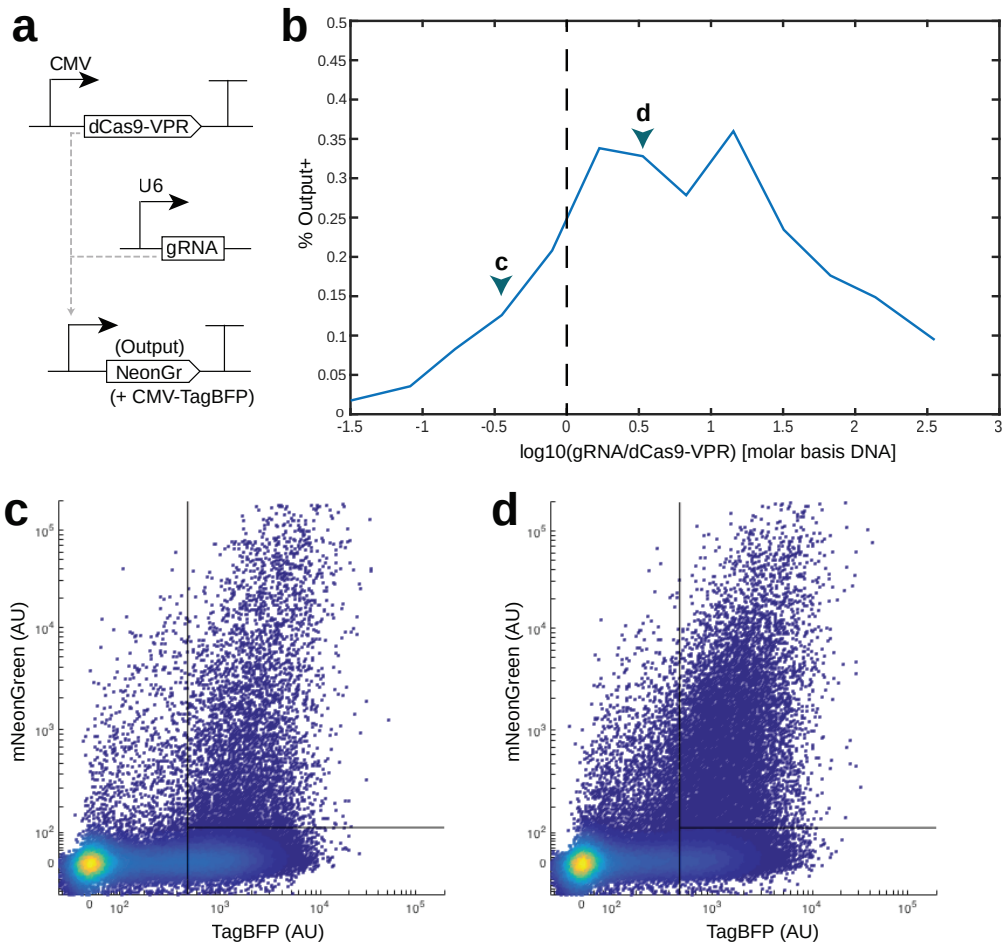


Figure 4-8: Co-transfection experiment for dCas9-VPR transcriptional activation.

(a) dCas9-VPR activation system in co-transfection format. Similarly to the poly-transfection experiment (Fig. 4-6f), dCas9-VPR and gRNA are constitutively expressed and activate a TRE promoter which is targeted by the gRNA. Since all components are premixed before making transfection complexes, only a single transfection marker is needed (TagBFP). We transfected HEK293FT cells in 24-well format with 60 ng of the output plasmid, 60 ng of the transfection marker CMV-TagBFP, and a total of 210 ng payload split between dCas9VPR and gRNA plasmids at various ratios. **(b)** We calculated the fraction of transfected cells expressing the output above a threshold (100 AU) for each ratio of payload. The samples in which 1-10 fold more gRNA plasmid was delivered relative to dCas9-VPR showed higher frequency

of activation. This result is in agreement with the poly-transfection experiment (Fig. 4-6f). Two samples where gRNA DNA amount was below (c) or above (d) dCas9-VPR are highlighted. **(c)** Scatter plot showing amount of output (mNeonGreen) and transfection marker (TagBFP) along with the thresholds for determining whether a cell was transfected (TagBFP = 500 AU) or expressing output (mNeonGreen = 100 AU). For this sample, gRNA DNA was ∼3-fold lower than that for dCas9-VPR. **(d)** Similar scatter plot to (c) but where gRNA dNA was ∼3-fold higher than dCas9-VPR. Note that in this co-transfection protocol, we used DNA amounts based on previous studies[102]. However, our corresponding poly-transfection experiment has indicated that greater activation can be achieved at altered amounts (Fig. 4-6f)

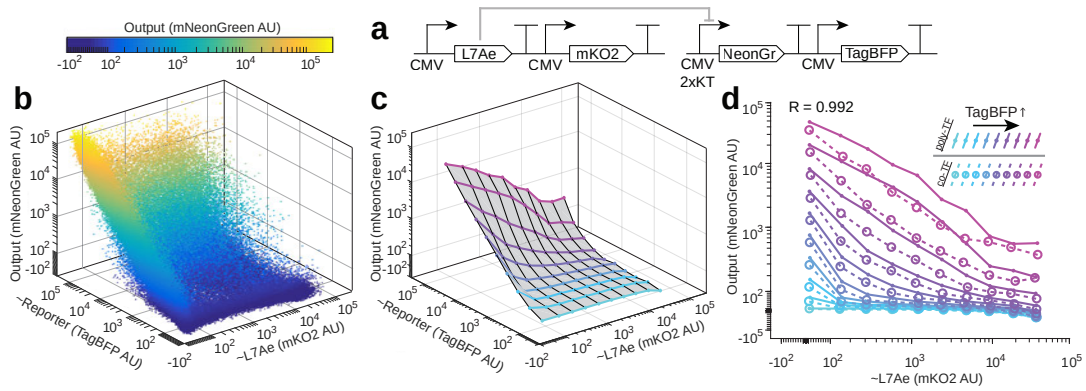


Figure 4-9: One-pot characterization of L7Ae translational repression. **(a)** Circuit diagram for L7Ae repression. The first plasmid expresses L7Ae and mKO2 constitutively. The second plasmid expresses tagBFP and the mNeonGreen output reporter that contains two k-turn motifs (2xKT) in its 5' UTR and is repressed by L7Ae. **(b)** 3D scatter plot for raw poly-transfection data. mNeonGreen fluorescence is is plotted and colored for the the two-dimensional parameter space define by reporter amount (tagBFP fluorescence) and L7Ae amount (mKO2 fluorescence). **(c)** Surface plot of the poly-transfection data. Data was binned on the fluorescence of the two input dimensions (tagBFP and mKO2) and points show median mNeonGreen in each bin. Transfer curve lines are colored corresponding to reporter bin. **(d)** Comparison of poly-transfection and co-transfection transfer curves. The data from 11 co-transfection samples for different ratios of L7Ae-containing plasmid and reporter-containing plasmid were concatenated, binned, and as in (c). Transfer curve lines are colored corresponding to reporter bin. Both poly-transfection data (solid) and co-transfection data (dashed) show similar behavior.

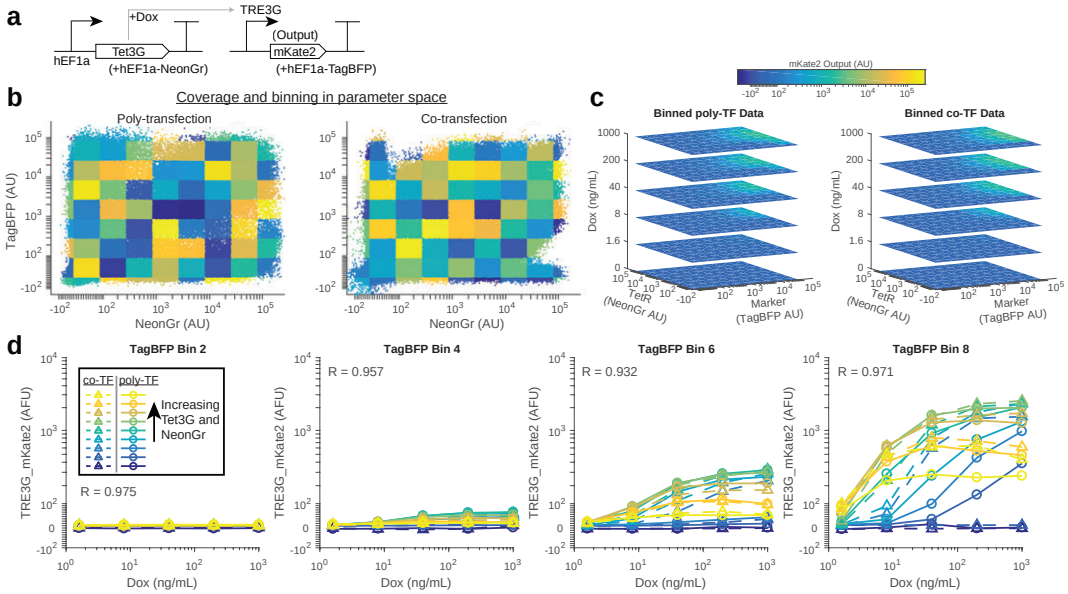


Figure 4-10: One-pot characterization of Doxycycline-induced Tet3G activation.

(a) Circuit diagram for Tet3G activation. The first transfection complex contains constitutively expressed Tet3G and mNeonGreen. The second complex contains a constitutive TagBFP reporter and a Dox-induced, Tet3G-driven mKate2 output. **(b)** Binning scheme for comparing poly- and co-transfections. Poly- and co-transfected data were divided into evenly-spaced bins determined by the fluorescence level of mNeonGreen (X-axis) and TagBFP (Y-axis), which report the amount of Tet3G and the output promoter, respectively. For comparability, co-transfection data were pooled into one collective sample prior to binning. Bin colors were randomly assigned for visualization purposes. **(c)** Median mKate2 fluorescence in each bin at each level of Dox induction for both the poly-transfection and pooled co-transfection data. **(d)** Direct comparison of Dox-induction curves at different Tet3G and output levels. Each panel shows the Dox-induction curve for one slice of data corresponding with the indicated TagBFP bin. Each line corresponds with one mNeonGreen bin. Dashed lines represent co-transfection data and solid lines represent poly-transfection data. The two methods show very similar behavior, except at very high output plasmid and low Tet3G levels.

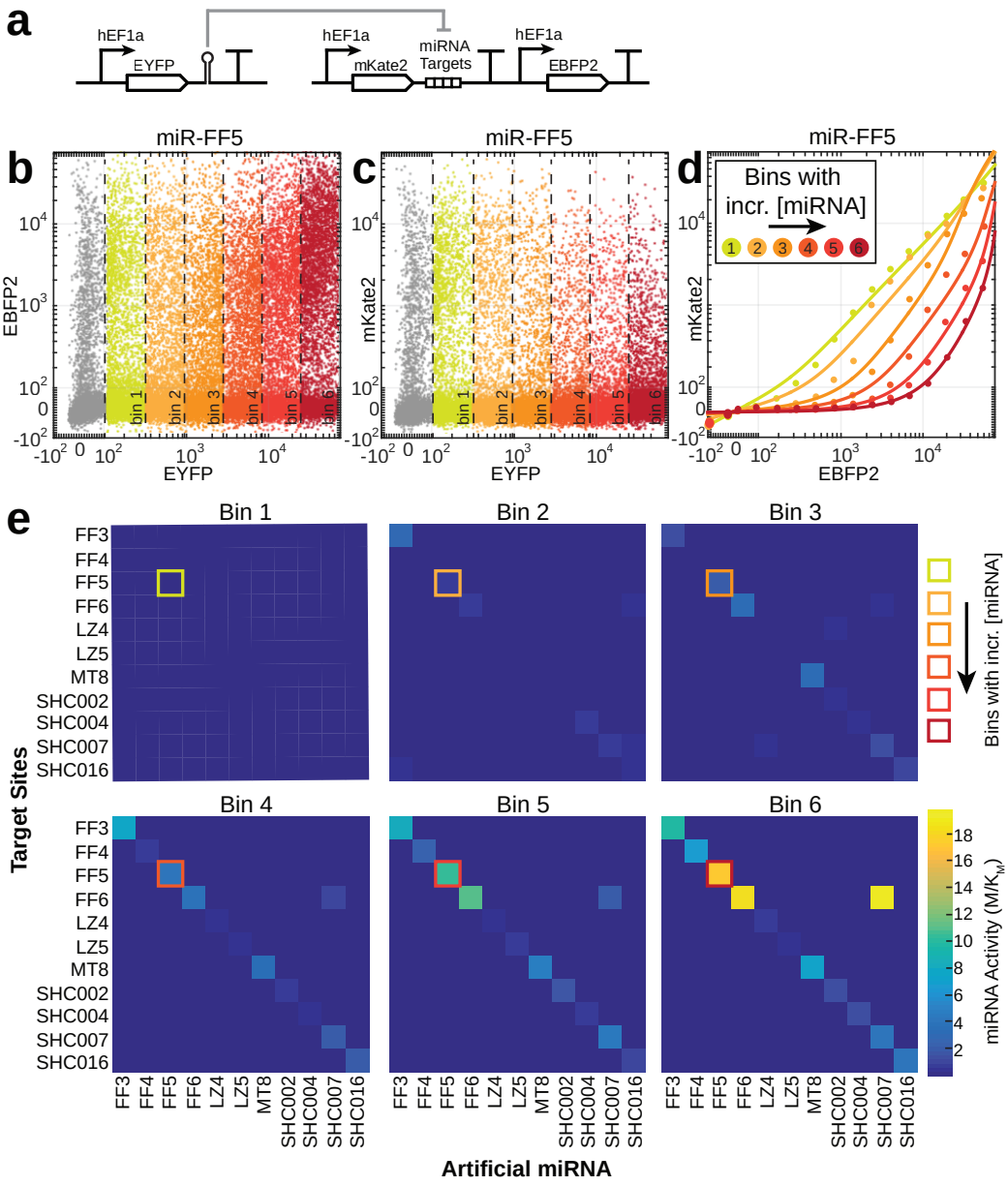

a
hEF1a
EYFP
hEF1a
mKate2
miRNA
Targets
hEF1a
EBFP2
b
miR-FF5
EBFP2
EYFP
bin 1
bin 2
bin 3
bin 4
bin 5
bin 6
c
miR-FF5
mKate2
EYFP
d
miR-FF5
Bins with
incr. [miRNA]
mKate2
EBFP2
e
Bin 1
Bin 2
Bin 3
Bin 4
Bin 5
Bin 6
Target Sites
FF3
FF4
FF5
FF6
LZ4
LZ5
MT8
SHC002
SHC004
SHC007
SHC016
Artificial miRNA
Bins with incr. [miRNA]
miRNA Activity (M/K_M)

Figure 4-11: One-pot characterization of artificial miRNA activity and orthogonality.
(a) Genetic circuit for testing artificial miRNA behavior. The circuit consists of two plasmids where the first expresses a synthetic miRNA based on a miR-155 expression platform and also an EYFP marker serving as an indicator of miRNA concentration. The second plasmid encodes a miRNA low sensor where the mKate2 fluorescent protein contains miRNA target sites in the 3' UTR which mediate repression of mKate2 in the presence of the complementary miRNA. An EBFP2 transfection marker indicates how much of this sensor plasmid is delivered to each cell. **(b)** Scatter plot of poly-transfection input parameter space for miR-FF5. A plot of miRNA concentration (∼EYFP) and sensor concentration (∼EBFP2) shows that most of parameter space is covered by the poly-transfection. Data are colored according to EYFP fluorescence, with yellow to red marking the lowest to highest EYFP bins respectively. **(c)** Scatter plot of output as a function of miR-FF5 concentration. In contrast to EBFP2, mKate2 is decreased when miR-FF5 concentration (∼EYFP) is increased, demonstrating that miR-FF5 is able to repress the FF5 sensor. **(d)** Sensor performance at different miR-FF5 concentrations. Since sensors for endogenous miRNAs are often plotted with output as a function of transfection marker, we binned data in each of the EYFP bins further using EBFP2 and calculated the mKate2 medians (circles). Lines indicate fits to a miRNA repression model.[44] As miRNA concentration was increased, its ability to repress the sensor was increased as indicated by further decreases in mKate relative to EBFP2. **(e)** Orthogonality of artificial miRNAs and sensors. We generated orthogonality matrices for each EYFP bin where rows indicate which target sites were present on the sensor and columns indicate the artificial miRNA introduced. Colors indicate the miRNA activity as fit using a miRNA repression model in order to compress the EBFP2 and mKate2 dimensions into a single measure for simplified visualization. Squares corresponding to the miR-FF5 data displayed in (d) are outlined with the appropriate EYFP bin color. Almost all miRNA/sensor pairs are highly orthogonal, as illustrated by miRNa activity along the diagonal at high miRNA concentration (bin 6). 11*11=121 poly-transfections were needed to generate the five-dimensional data shown here.

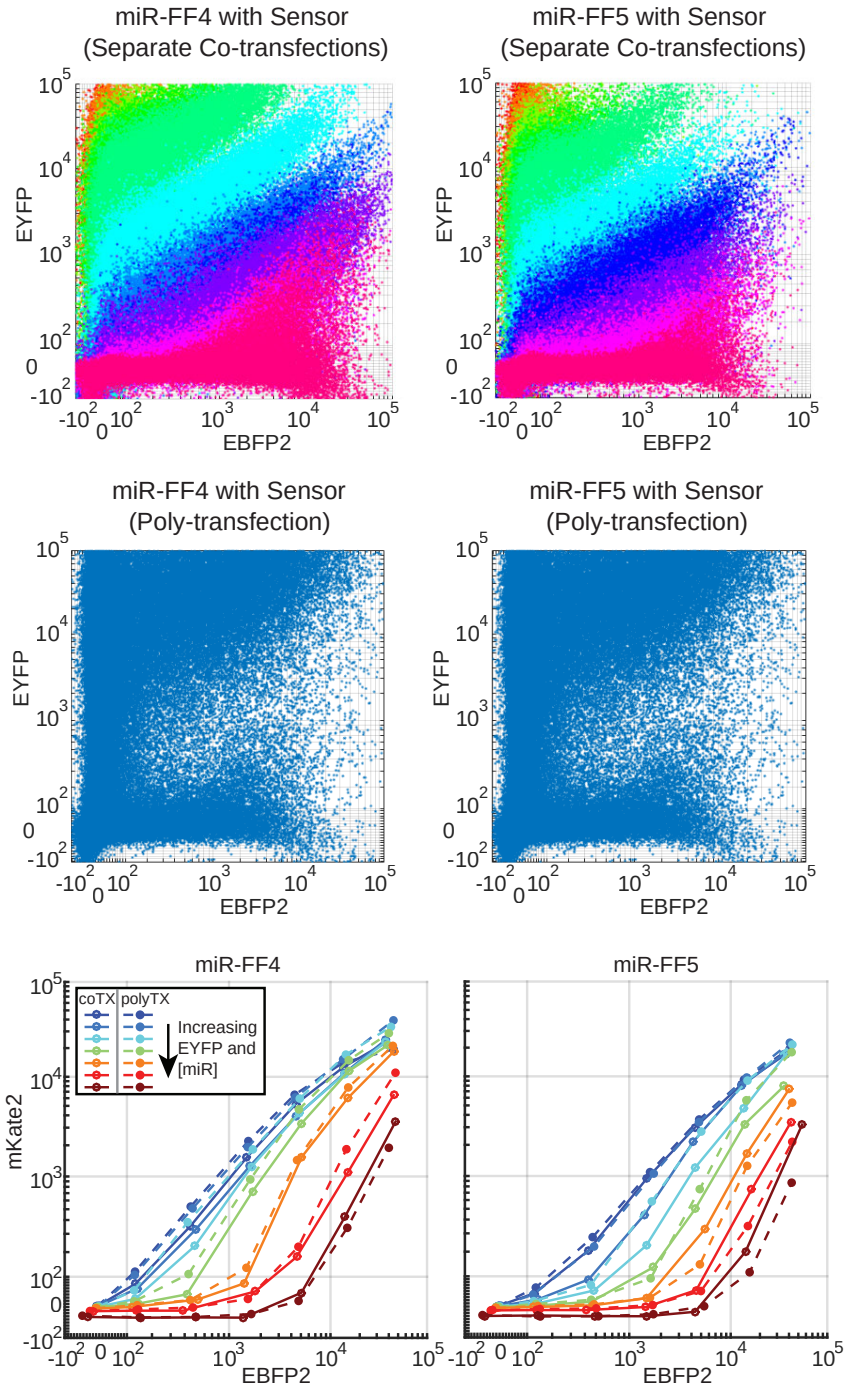

miR-FF4 with Sensor
(Separate Co-transfections)
miR-FF5 with Sensor
(Separate Co-transfections)
miR-FF4 with Sensor
(Poly-transfection)
miR-FF5 with Sensor
(Poly-transfection)
EYFP
EBFP2
miR-FF4
miR-FF5
mKate2
coTX
polyTX
Increasing
EYFP and
[miR]

Figure 4-12: Comparison of co- and poly-transfections for artificial miRNA experiment.
For FF4 (left) and FF5 (right) we used separate co-transfections (top) or a poly-transfection (middle) to obtain cells spanning the two-dimensional input parameter space. After concatenating data together from the separate co-transfections we analyzed data by binning along EYFP then EBFP to plot miRNA activity as amount of transfected miRNA increased (bottom). Both co-transfections (solid lines) and poly-transfections (dashed lines) exhibited similar sensor behavior in response to miRNA.

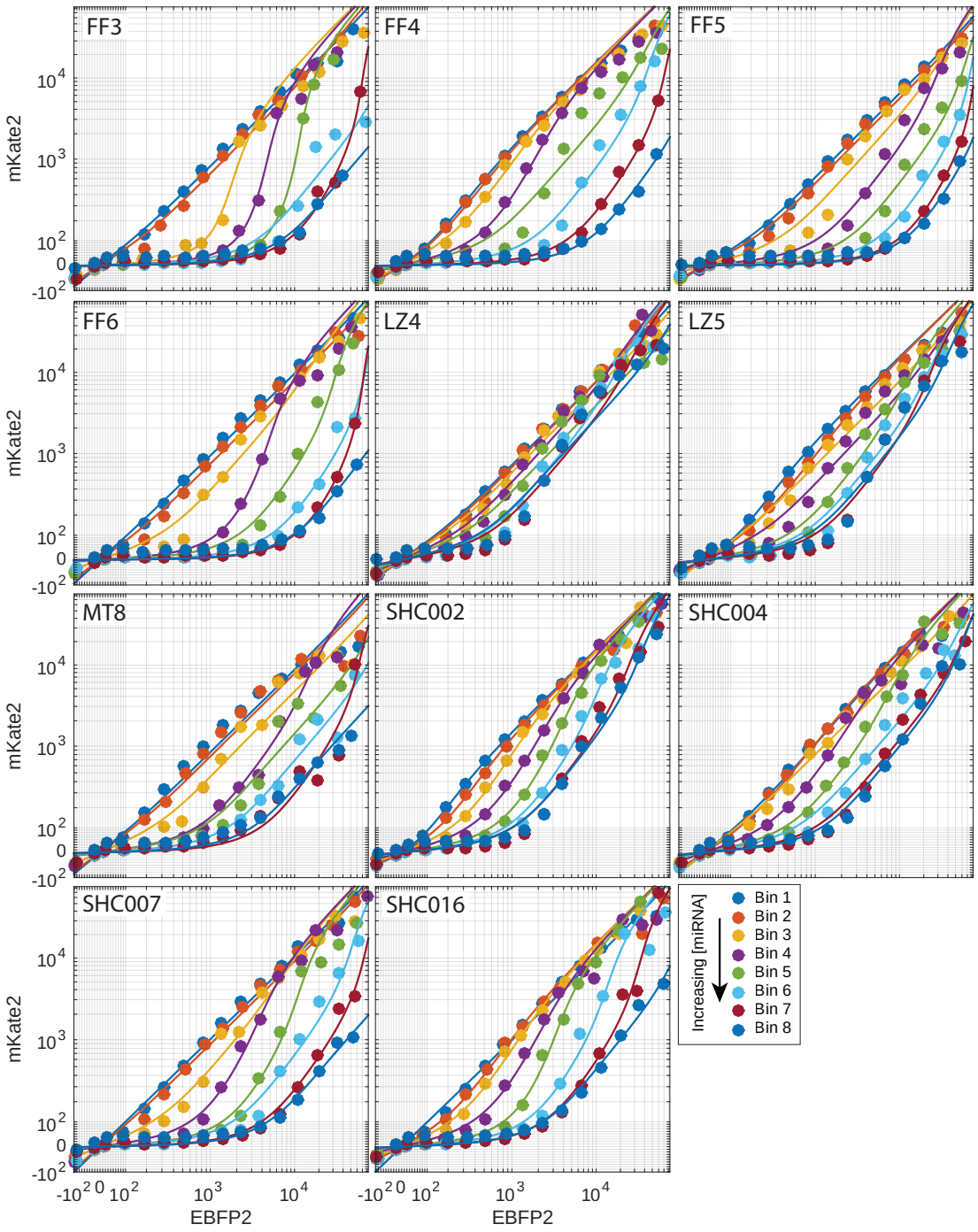

FF3
FF4
FF5
FF6
LZ4
LZ5
MT8
SHC002
SHC004
SHC007
SHC016
mKate2
EBFP2
Increasing [miRNA]
Bin 1
Bin 2
Bin 3
Bin 4
Bin 5
Bin 6
Bin 7
Bin 8

Figure 4-13: miRNA activity for each artificial miRNA tested.
Though each sensor showed a repression when the corresponding miRNA was added, the degree of repression and thus the activity of each miRNA varied. Curves were fit using a miRNA repression model similar to those previously published[44] and fit parameters were used as indicators for miRNA activity shown in Fig. 4-11. The FF miRNA showed the highest activity in general, followed by the SHC miRNAs derived from Sigma Aldrich sequences, and then the LZ and MT miRNAs. These results indicate that poly-transfections can serve as part of a simplified strategy to characterize miRNA and shRNA strengths.

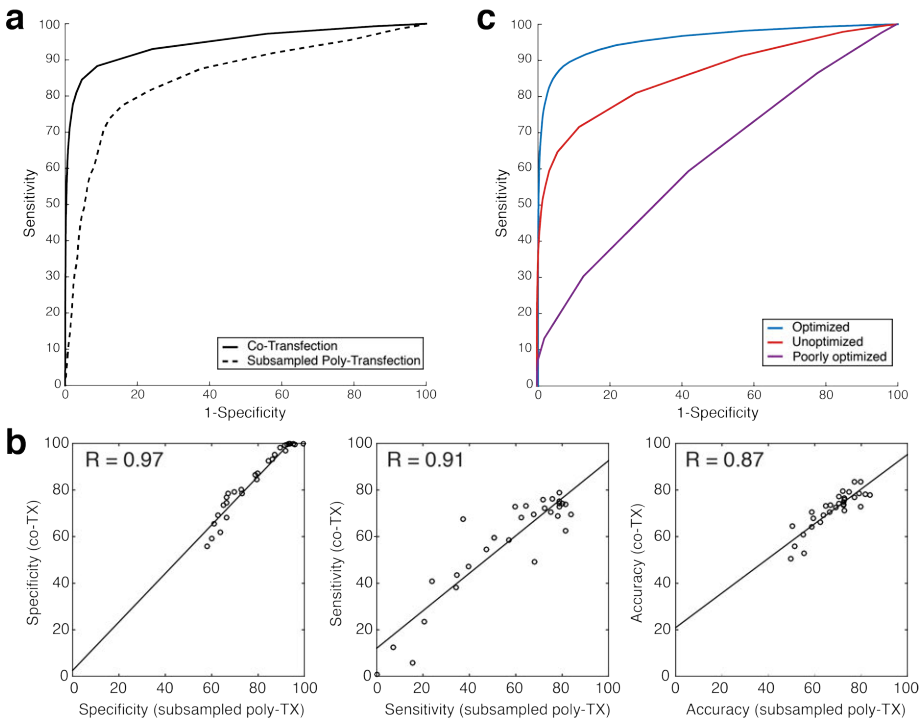
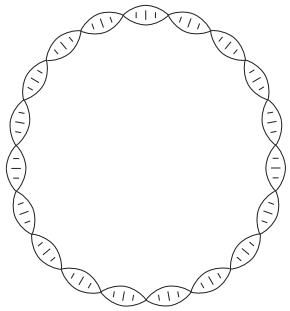


Figure 4-14: Additional information for the miR-21-5p classifier.
(a) Receiver operating characteristic (ROC) curves for discrimination of HeLa and HEK293FT cells with the optimize. Comparison of ROC curves for either a co-transfection or subsampled poly-transfection show similar behavior at DNA amounts of 40ng:60ng:435ng for BM3R1:reporter:Gal4-VP16. Higher accuracy and AUC were obtained with the co-transfection, perhaps due to increased noise with poly-transfections. AUC was 94.7% and 84.3% for co-transfection and subsampled poly-transfection respectively. **(b)** Correlation between subsampled poly-transfection and co-transfections. Sensitivity, specificity, and accuracy data corresponding to Figure 4-7e were plotted and correlation coefficient measured. For each metric, the subsampled poly-transfections correlated well with the actual measurements from co-transfection, indicating the potential for poly-transfections to circuit/system design, in this case for a miR-21-5p classifier. **(c)** ROC curves for variants of the miR-21-5p classifier. ROC curves show progressively increasing accuracy and AUC as variants transition from poorly optimized, to unoptimized, to optimized. AUC was 62.4%, 86.0%, and 95.8% respectively.



This page intentionally left blank to keep captions near figures.

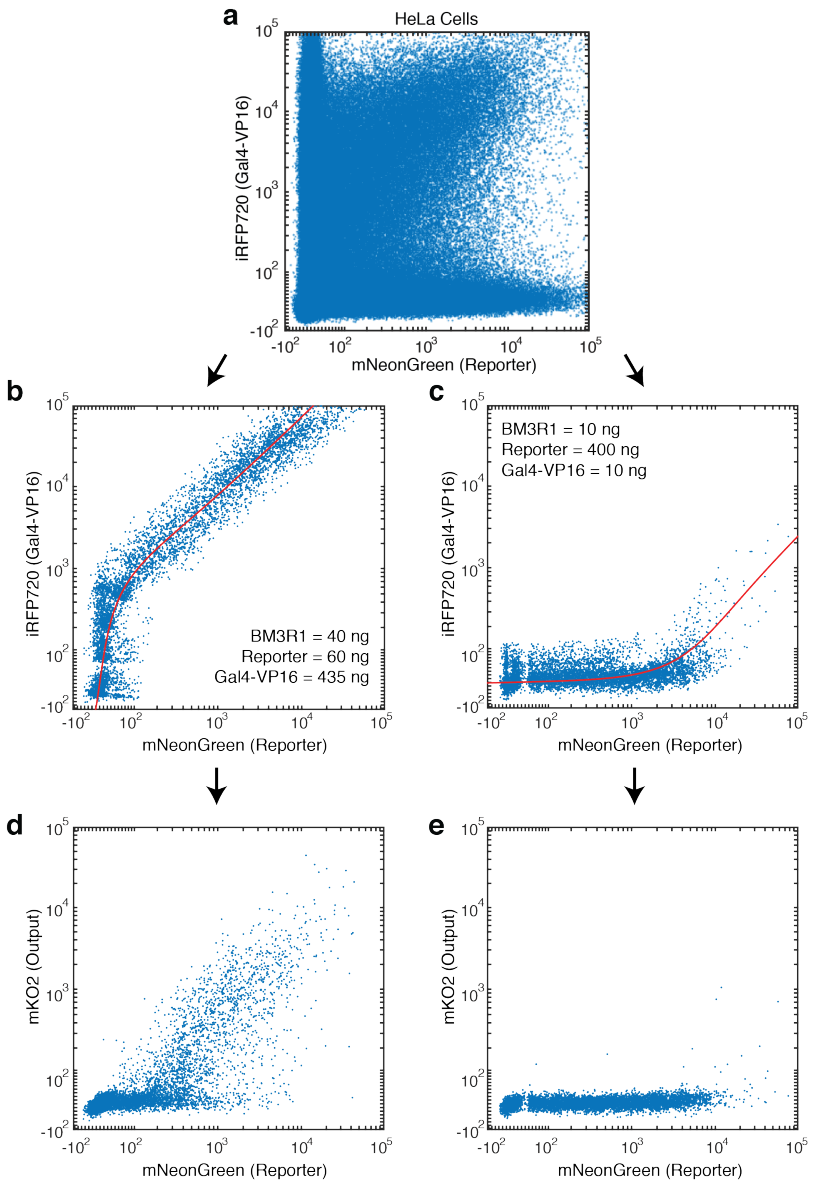
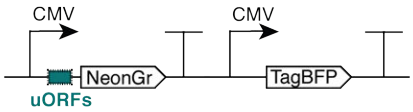
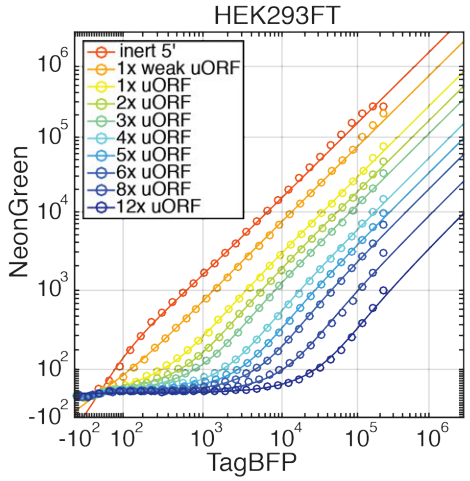
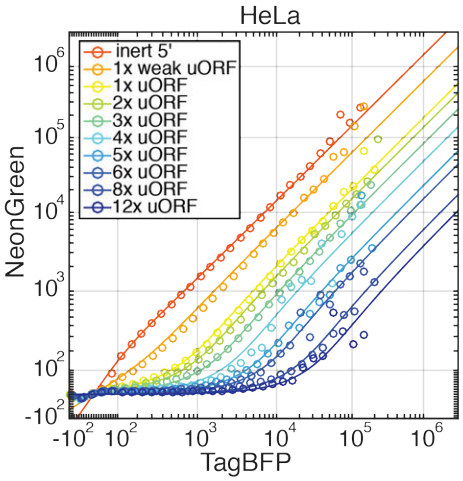


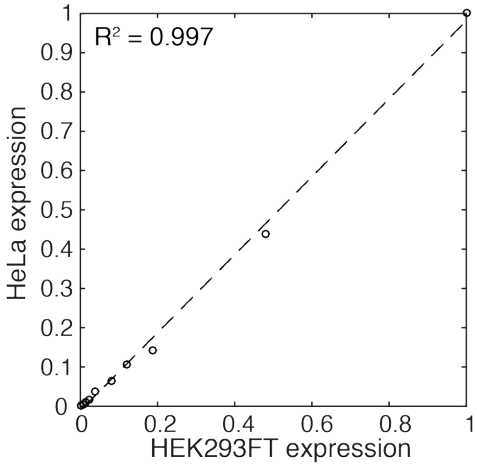
Figure 4-15: Demonstration of subsampling for poly-transfections. Since often the genetic circuits (or systems) used for therapeutic or other applications will be co-expressed at some ratio, it may be advantageous to subsample poly-transfection in a way that mimics expression at that ratio, giving a valuable estimate of how that genetic circuit will perform. In order to subsample, we defined 'trajectories' that correspond to a specific ratio of DNA as indicated in red (which can

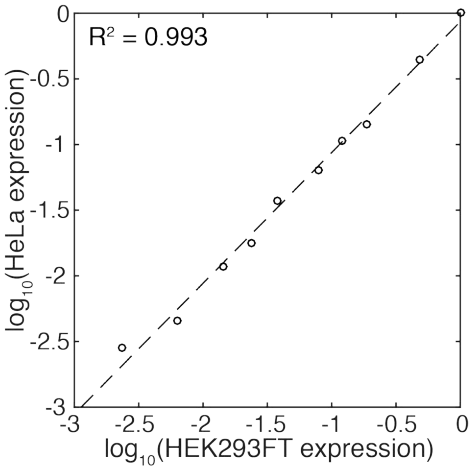
also be defined by DNA amounts). Poly-transfection data 'near' the trajectory are included for plotting in blue. Whether a given data point is included is determined by the distance of that point from a point in the trajectory. In our subsampling, we included the first 312 cells within 0.1 distance units of the trajectory point, 161 cells within 0.2 distance, 98 cells within 0.3 distance, 50 cells within 0.4 distance, and 18 cells within 0.5 distance. Distance was in logicle transformed space and the sampling distribution was based on a log-normal distribution within a three-dimensional parameter space. The genetic system shown here is for the miRNA high sensor shown in Fig. 4-7. We show only the data for two of the input parameters for simplicity, though the subsampling was performed in the full 3D parameter space. After subsampling, output can be plotted as a function of any of the input components, in this case the amount of reporter plasmid. Two examples in HeLa cells with different ratios of components are shown. Note that the example with high Gal4-VP16 showed higher output in HeLa cells as we indicated previously, while the example with low Gal4-VP16 showed minimal ability to provide a positive output.











uORF	Avg. Fold KD	uORF	Avg. Fold KD
1x weak	2.17	5x	48.4
1x	6.07	6x	76.6
2x	8.87	8x	183
3x	13.9	12x	388
4x	26.9		

Figure 4-16: Tuning gene expression using upstream open reading frames (uORFs).
We tested whether uORFs could be used to tune gene expression where the CMV truncations are not suitable, for instance when using an inducible or tissue-specific promoter. In our test constructs, mNeonGreen was placed under control of a CMV promoter with uORFs upstream in between promoter and gene, while TagBFP was used as a transfection marker. Thus effects from the uORF should knockdown mNeonGreen expression but not TagBFP. We observed this behavior in both HEK293FT and HeLa and to an increasing degree as more uORFs were added. Analysis was conducted similarly to the CMV truncations, with binned data fit to a ratiometric model to determine the degree of knockdown. Knockdown was consistent across cell lines with R-squared values above 0.99 for both linear and log plots. Knockdown values averaged across cell lines are shown, with up to 388-fold knockdown possible with the 12x uORF variant.

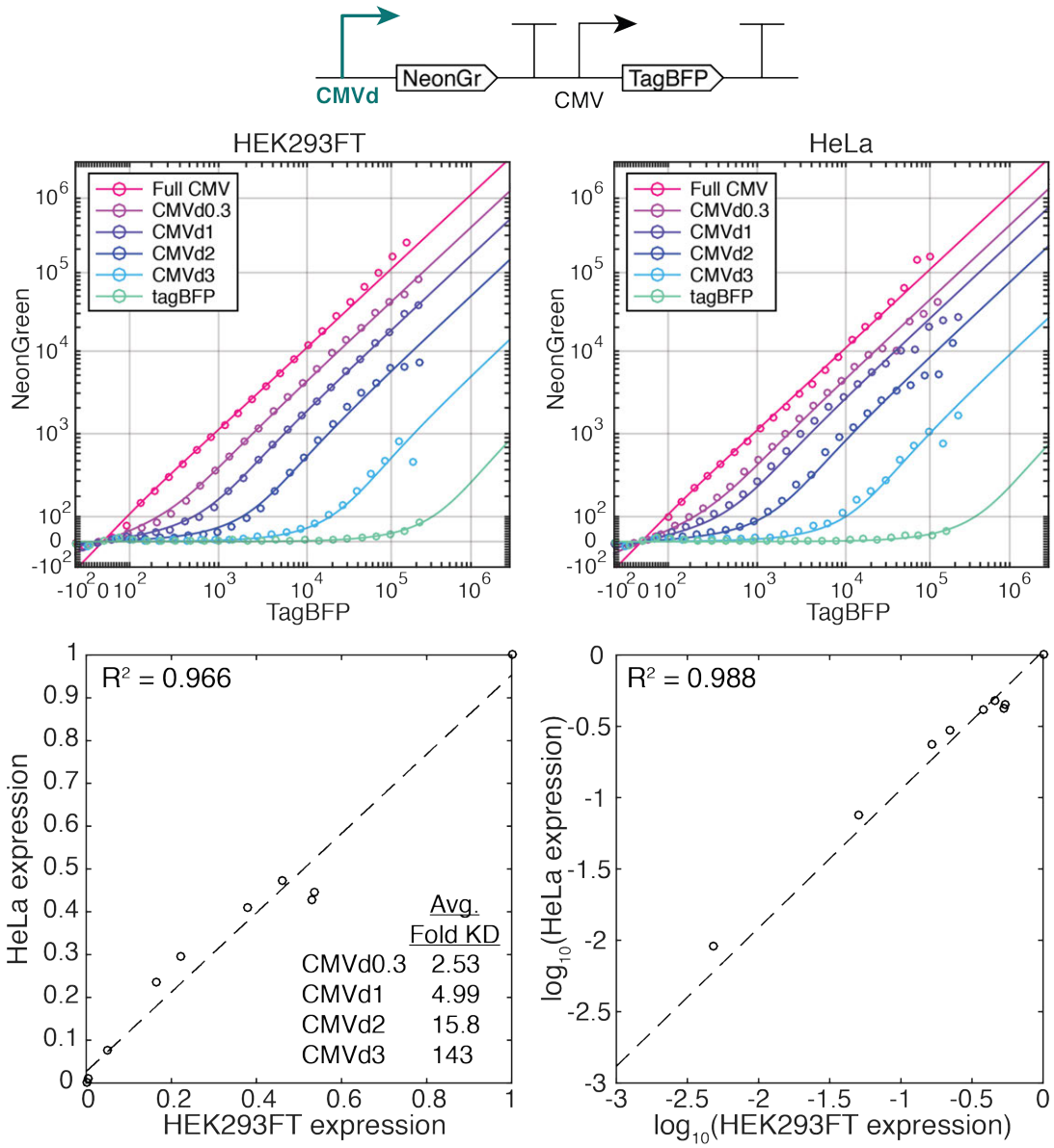

CMVd
NeonGr
CMV
TagBFP
HEK293FT
HeLa
Full CMV
CMVd0.3
CMVd1
CMVd2
CMVd3
tagBFP
NeonGreen
TagBFP
R² = 0.966
R² = 0.988
HeLa expression
HEK293FT expression
log10(HeLa expression)
log10(HEK293FT expression)
Avg.
Fold KD
CMVd0.3 2.53
CMVd1 4.99
CMVd2 15.8
CMVd3 143

Figure 4-17: Tuning gene expression using promoter truncations.
We tested different truncations of the CMV promoter to determine the ability to tune down gene expression. We tested several previously characterized truncations along with new ones. In the test plasmids, mNeonGreen is driven by the CMV truncation of interest while a TagBFP driven by full CMV serves as a transfection marker. In both HEK293FT and HeLa cells, as CMV is truncated further there is a marked decrease in mNeonGreen compared to the TagBFP reference. Binned data (circles) were fitted (lines) using a model where mNeonGreen is simply a fraction of TagBFP level. The fit mNeonGreen/TagBFP expression ratios (and log ratios) were then plotted for HEK293FT vs. HeLa and we observed comparable knockdown of mNeonGreen across cell lines. Values for knockdown (averaged over both cell lines) of our chosen set spanning a range of values is shown in the inset, with up to 143-fold knockdown with CMVd3. R-squared values for knockdowns when compared across cell lines were high at 0.966 and 0.988 for linear and log ratios respectively. These results indicate that CMV truncations can be used to tune gene expression to a large and controllable degree in the two cell lines tested.

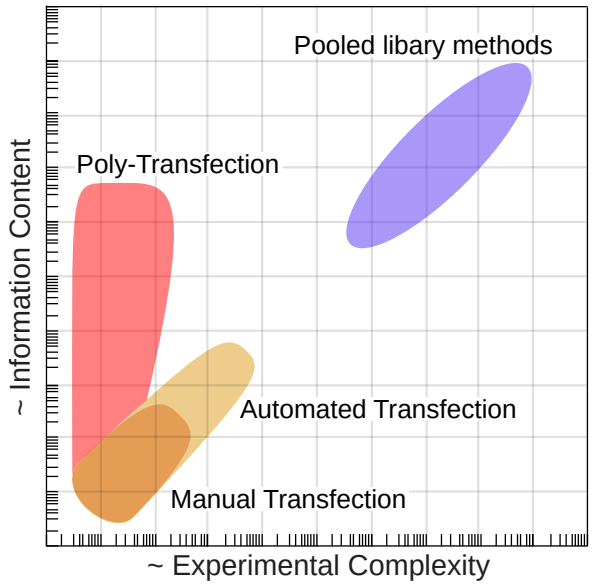


Figure 4-18: Qualitative comparison of methods to characterize and optimize genetic circuits and systems.
In general, the higher information content experiments require more experimental complexity (which is roughly correlated with cost, time or chance for error). Co-transfections are relatively simple to perform but manual pipetting steps limit the number of samples (and thus information) per assay. Automated transfection and measurements may increase the number of samples that can be acquired by 1-2 orders of magnitude, resulting in a small increase in experimental complexity and information content. Pooled library approaches require many complex steps but yield much more information (e.g. around 5×10^4 combinations for CombiGEM[118]). Typical steps include generating a uniform library of variants - either through combinatorial DNA assembly or mutagenesis strategies, delivery and integration of DNA at a single copy per cell (e.g. with low MOI lentivirus or landing pads), flow sorting of cells with different output levels into groups, then conducting high throughput sequencing on each sorted group to determine which constructs expressed which output levels. Poly-transfections can yield information with similar information content to smaller pooled library experiments. For instance, with a 4-component system using 15 expression level bins per dimensions, the parameter space is approximately 5×10^4 potential combinations. Additionally, since the poly-transfection requires only a single well of cells, an experiment is only as complex to perform as a manual transfection.

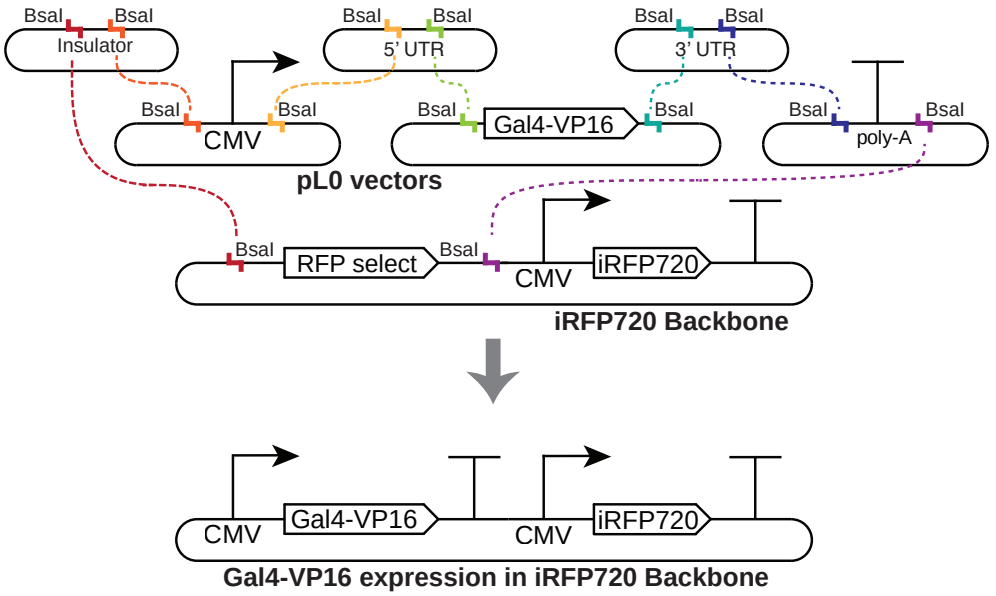


Figure 4-19: Simplified assembly strategy for one-pot-ready plasmids. We have constructed a set of backbone plasmids containing constitutive expression of one of five fluorescent proteins (Sirius, TagBFP, mNeonGreen, mKO2, iRFP720) along with restriction sites and colorimetric selection for BsaI-mediated Golden Gate assembly. In a single Golden Gate step, plasmid level 0 (pL0) plasmids encoding insulator, promoter, 5' UTR, gene, 3' UTR, and poly-A sequences can be assembled into the backbone and white colonies can be selected for against red-expressing background. Plasmids assembled in this way are immediately ready for use in poly-transfections since they encode simultaneous expression of a circuit part (Gal4-VP16 in the illustrated example) and a fluorescent transfection marker (iRFP720 in the example). The backbone plasmids will be deposited into Addgene for to aid researchers in their one-pot characterization or optimization experiments.

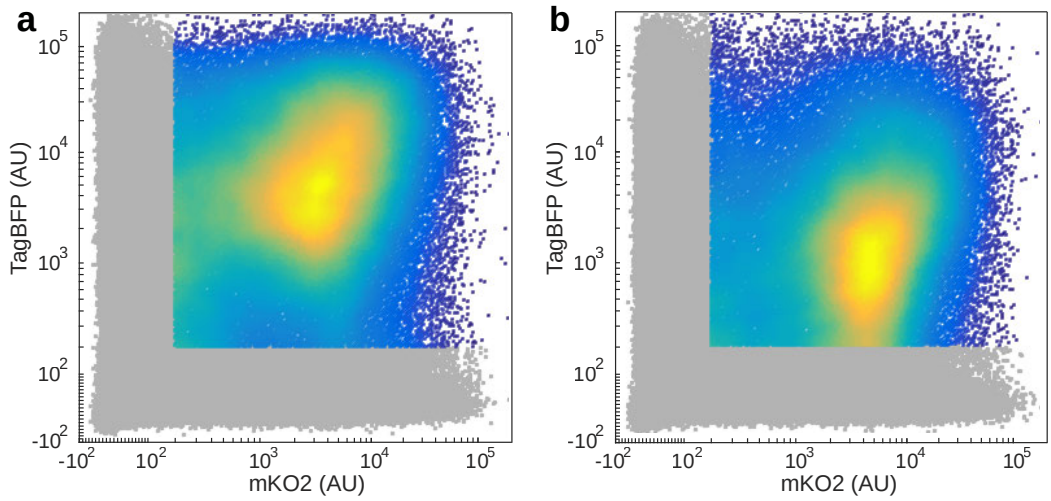


Figure 4-20: Poly-transfection distributions can be changed by titrating DNA.

(a) Poly-transfection with equal amounts of TagBFP- and mKO2-expressing plasmids. This poly-transfection used 250 ng of each plasmid, resulting in a high proportion of double-positive cells (points colored by density, yellow = greater density) expressing medium-high levels of both TagBFP and mKO2. Single-positive and untransfected cells are colored in gray. **(b)** Poly-transfections with 5-fold decreased TagBFP plasmid. For the TagBFP plasmid-lipid complexes, 50 ng of TagBFP plasmid and 200 ng of non-expressing dummy vector were premixed and transfection reagent added as before. In this case, double-positive cells express low-medium levels of TagBFP but still medium-high levels of mKO2. Thus the cells in the lower TagBFP polytransfection (and in an additional lower mKO2 polytransfection) could be used to more efficiently sample the parameter space between doubly positive and singly-positive populations, compared to the poly-transfection in (a) alone.

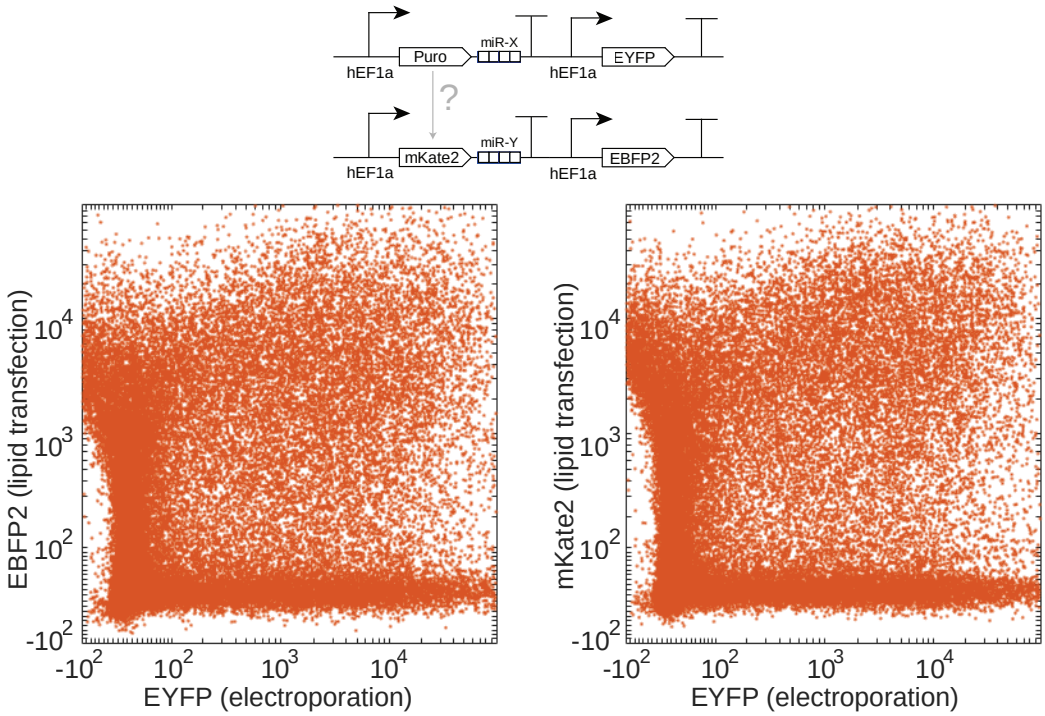


Figure 4-21: Expression distribution from sequential electroporation and lipid transfection.
In the following experiment, an EYFP containing plasmid was transfected to determine the effect of miRNA target sites on a separate miRNA sensor with EBFP2 and mKate2 fluorescent proteins. The EYFP plasmid was first electroporated into HEK293FT cells, incubated for 24 hours, then the EBFP2 and mKate2-containing plasmid was transfected with lipofectamine 3000. Similar broad expression distributions were obtained with the sequential electroporation/lipid transfection compared with the lipid-only method. It is possible that other methods of transfection can be combined to yield poly-transfection data, though we focused on lipid-only methods for simplicity. Note that some cells are off-scale due to compensation to correct for mKate2 bleedthrough into the other channels.

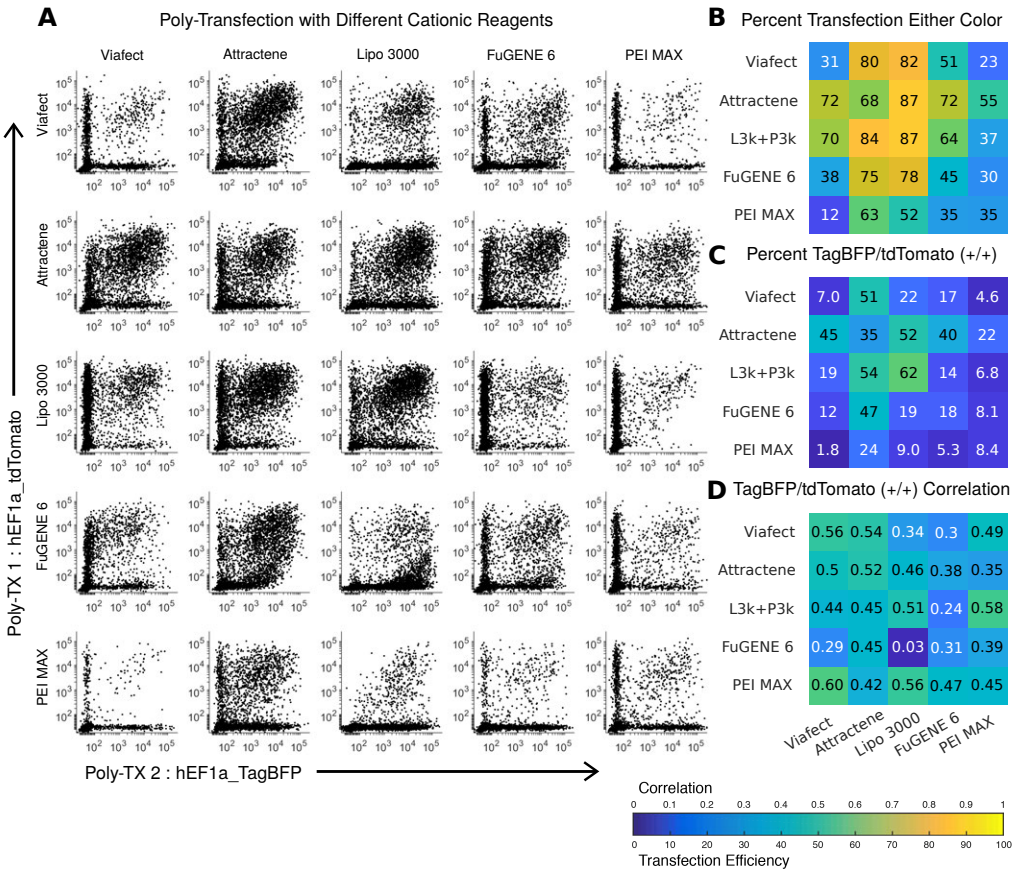
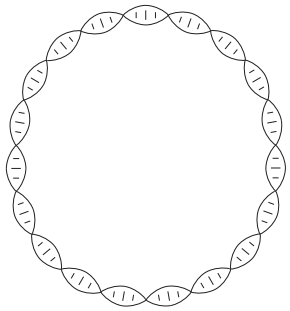


Figure 4-22: Poly-TX with different cationic transfection reagents. **(a)** HEK-293FT cells poly-transfected with hEF1a_TagBFP (x-axis/columns) and hEF1a_tdTomato (y-axis/rows) each complexed with one of five cationic transfection reagents (Viafect, Attractene, Lipofectamine 3000, FuGENE6, PEI MAX). 4000 cells shown in each plot. **(b)** % cells positive for TagBFP or tdTomato in each sample (single- or double-positive). **(c)** % cells positive for both TagBFP and tdTomato in each sample (double-positive). **(d)** Correlation between $\log_{10}$-transformed TagBFP and tdTomato values in the double-positive population of each sample.



This page intentionally left blank to keep captions near figures.

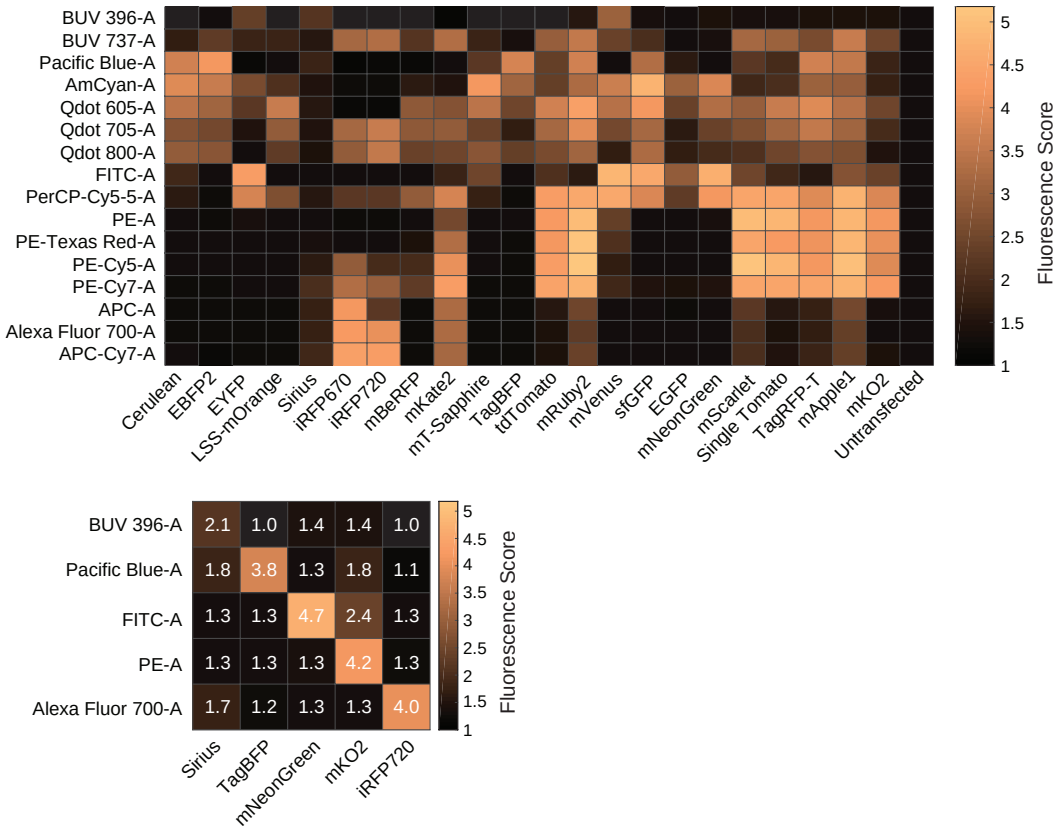


Figure 4-23: Screening of fluorescent proteins to determine compatible sets.

We tested a panel of 22 different fluorescent proteins to determine how many could be measured without significant bleedthrough in signal between channels. Fluorescent scores were calculated for each FP and measured with each channel using the following equation: score = log10(99th percentile channel fluorescence from sample)-log10(channel background from untransfected). Fluorescent proteins were chosen for emission across a range of spectral space and for excitation with the 5-laser cytometer we used for the study. We also tested several high stokes shift proteins (LSS-mOrange, mBeRFP, mT-Sapphire) to determine whether two fluorescent proteins could be measured from the same laser, though those still resulted in significant bleedthrough to other channels. We observed several incompatibilities of note: mKate2 and mRuby2 have high bleedthrough into many channels, yellow fluorescent proteins (EYFP and mVenus) show high signal in the BUV 396 channel used to measure Sirius so green FPs are more desirable when both are to be measured, and

all red fluorescent proteins we measured (notably tdTomato, mScarlet, TagRFP-T, mApple1) showed stochastic signal in the Pacific Blue channel which meant compensation could not be used to correct for bleedthrough - instead orange fluorescent proteins (mKO2 measured in the PE channel) appear to be an alternative. We determined that the set comprised of Sirius, TagBFP, mNeonGreen, mKO2, and iRFP720 (measured in BUV396, Pacific Blue, FITC, PE, and Alexa Fluor 700 channels respectively) appeared to show the most orthogonal signals and bleedthrough that could be compensated for. When designing experiments with fewer than five fluorescent proteins, it may be useful to note the lower brightness of Sirius and moderate bleedthrough of mKO2 into mNeonGreen.

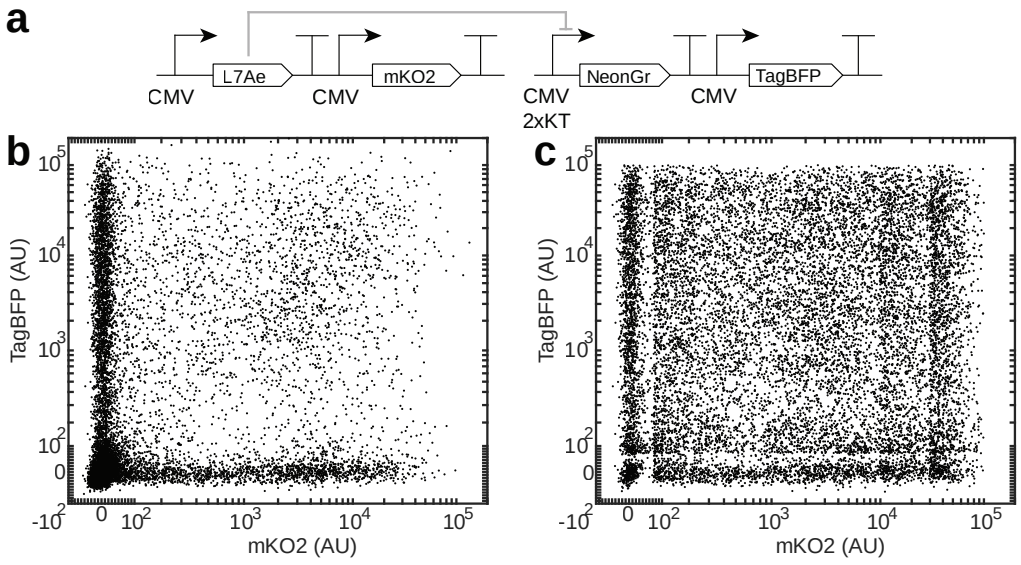
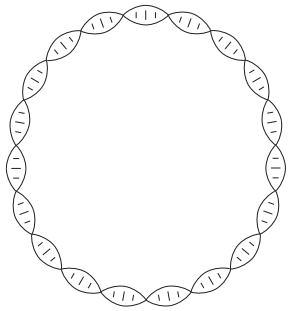


Figure 4-24: Normalization of poly-transfection data to reduce effects of transfection distribution.
(a) L7Ae repression circuit. Note that mKO2 and TagBFP are the input components (i.e. independent variables). Ideally data would sample across these two dimensions uniformly **(b)** Raw poly-transfection distribution. As previously noted, poly-transfections show a bias towards null-, singly-, and doubly-transfected cells, with fewer cells falling in between these regions. The first 14,700 data points from a total of >600,000 events are plotted here. **(c)** Normalized poly-transfection data. After binning with 7 bin edges across each dimension (49 total bins), 300 cells in each bin were randomly selected to be included in the normalized data. This resulted in a total of 14,700 data points, which is the same as in (b). The normalized scatter plot shows significantly less bias, though a slight bias remains in the singly-transfected bin. Bias could be further reduced by including more bin edges, or using other approaches to subsample (e.g. tessellation-based methods). More uniform sampling will likely help analysis of more sensitive circuits or those that cannot be analyzed by taking the median/geometric mean of binned output fluorescence (which already corrects for transfection biases to a degree).



This page intentionally left blank to keep captions near figures.

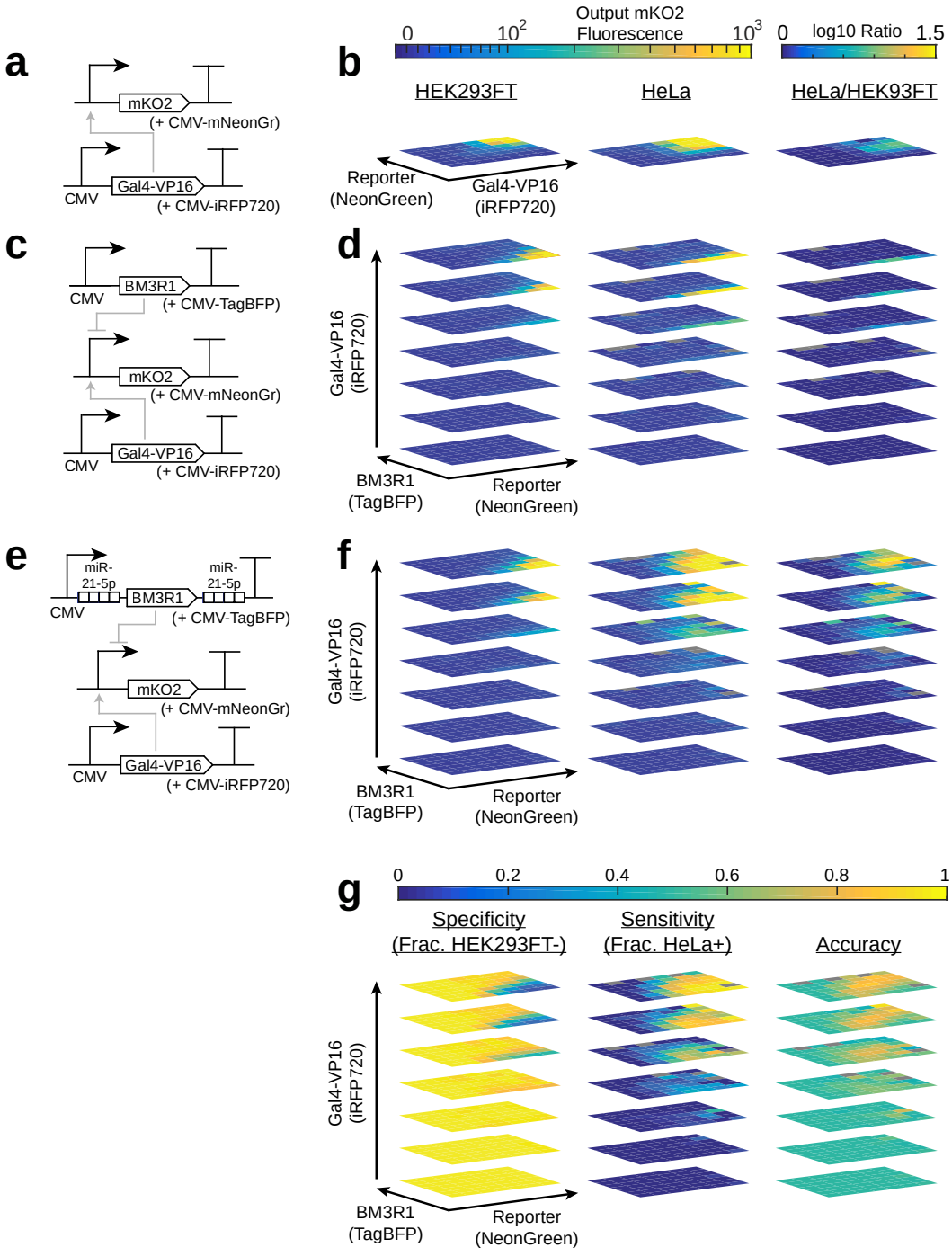

a
mKO2
(+ CMV-mNeonGr)
CMV
Gal4-VP16
(+ CMV-iRFP720)
b
Output mKO2
Fluorescence
0
10^2
10^3
0
log10 Ratio
1.5
HEK293FT
HeLa
HeLa/HEK93FT
Reporter
(NeonGreen)
Gal4-VP16
(iRFP720)
c
CMV
BM3R1
(+ CMV-TagBFP)
mKO2
(+ CMV-mNeonGr)
CMV
Gal4-VP16
(+ CMV-iRFP720)
d
Gal4-VP16
(iRFP720)
BM3R1
(TagBFP)
Reporter
(NeonGreen)
e
miR-
21-5p
miR-
21-5p
CMV
BM3R1
(+ CMV-TagBFP)
mKO2
(+ CMV-mNeonGr)
CMV
Gal4-VP16
(+ CMV-iRFP720)
f
Gal4-VP16
(iRFP720)
BM3R1
(TagBFP)
Reporter
(NeonGreen)
g
0
0.2
0.4
0.6
0.8
1
Specificity
(Frac. HEK293FT-)
Sensitivity
(Frac. HeLa+)
Accuracy
Gal4-VP16
(iRFP720)
BM3R1
(TagBFP)
Reporter
(NeonGreen)

Figure 4-25: Rapid characterization of a miRNA classifier and its sub-components.
(a) Circuit for Gal4-VP16 activation. The Gal4-VP16 activation circuit previously shown in Fig. 4-27 comprises a portion of the miR-21-5p classifier. For all circuit diagrams, the indicated CMV-fluorescent protein transfection markers were encoded on the same plasmid as the circuit parts but displayed separately to simplify the diagrams. **(b)** Heatmaps of activation by Gal4-VP16 in HEK293FT and HeLa cells. Heatmaps show that output is restricted to the corner where Gal4-VP16 and reporter are high. Taking the ratio of output in HeLa to HEK293FT shows minimal difference in output between the two cell lines. **(c)** Addition of BM3R1 repression to the circuit. BM3R1 is added to the circuit with a TagBFP transfection marker, such that mKO2 output is repressed when BM3R1 is expressed. **(d)** Heatmaps of output for BM3R1 and Gal4-VP16 circuit. In both HEK293FT and HeLa cells, output is restricted to the corner in parameter space where Gal4-VP16 and reporter are high but BM3R1 remains low. The minimal difference in HeLa/HEK293FT ratio shows that behavior of the circuit is similar in both cell types. Gray indicates that no cells were collected within a given bin. **(e)** A miRNA classifier circuit to distinguish HEK293FT and HeLa cells. With the addition of miR-21-5p target sites in both UTRs of BM3R1, a miRNA classifier is constructed. In HeLa cells, miR-21-5p activity is high, repressing BM3R1 and allowing mKO2 output to be expressed. Whereas in HEK293FT cells, miR-21-5p activity is low, such that BM3R1 remains high and represses the output. **(f)** Heatmaps for the complete miRNA classifier. In HEK293FT cells, behavior is similar to the case without target sites (d) since miR-21-5p activity is low,resulting in minimal effects from addition of miRNA target sites to BM3R1. In contrast, HeLa cells show output in an expanded region of parameter space, since endogenous miR-21-5p is able to knock down BM3R1 to some degree. Taking the ratio of HeLa to HEK293FT shows a band of parameter space where output is high in HeLa but low in HEK293FT. **(g)** Specificity, sensitivity and accuracy for the miRNA classifier. Sensitivity is calculated as the fraction of HEK293FT cells correctly determined to be low in output (mKO2 $> 10^2$ AU), sensitivity is calculated as the fraction of HeLa cells correctly determined to be high in output (mKO2 $< 10^2$ AU), while accuracy is calculated as the fraction of all correctly classified cells. Again a band in parameter space exhibits the optimal classification accuracy.

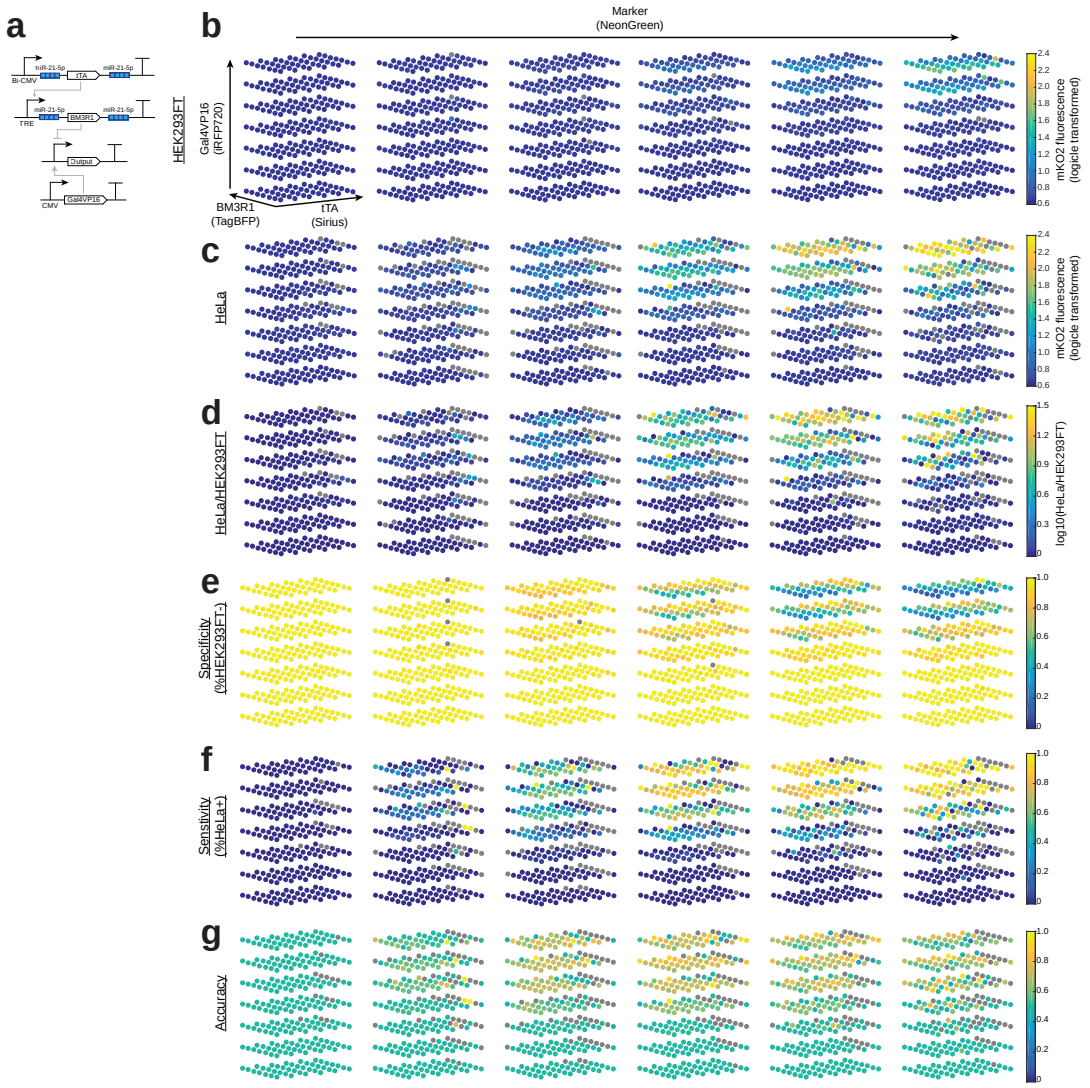


Figure 4-26: Data for 4-component miRNA classifier.
(a) In the 4-component miRNA classifier, BM3R1 is activated by tTA rather than transcribed from a constitutive promoter. Both tTA and BM3R1 contain miRNA target sites to give the miRNA the potential to regulate multiple circuit components and possibly enhance classification. **(b)** In these heatmaps, each dot represents one bin or region of parameter space. In HEK293FT cells, output remained generally low across parameter space, though there was some expression (as also seen for the 3-component classifier in Fig. 4-25) at high Gal4-VP16 and marker when BM3R1 and tTA were low **(c)** For HeLa cells output was high with Gal4-VP16 and marker

and low-medium levels of tTA and BM3R1. **(d)**One area of parameter space shows output in HeLa compared to HEK293FT as indicated by high HeLa/HEK293FT fluorescence ratio. **(e)** Specificity, or the fraction of HEK293FT cells correctly classified as negative in output (threshold = 100 AU), indicated most parameter space regions showed minimal HEK293FT output, with the exception of again high Gal4-VP16 and marker with low BM3R1 and tTA. **(g)** Sensitivity, or the fraction of HeLa cells correctly classified as positive in output, showed the inverse relationship compared to specificity. **(f)** Similarly to the 3-component classifier, a heatmap of accuracy shows a region in parameter space where high specificity and sensitivity overlap.

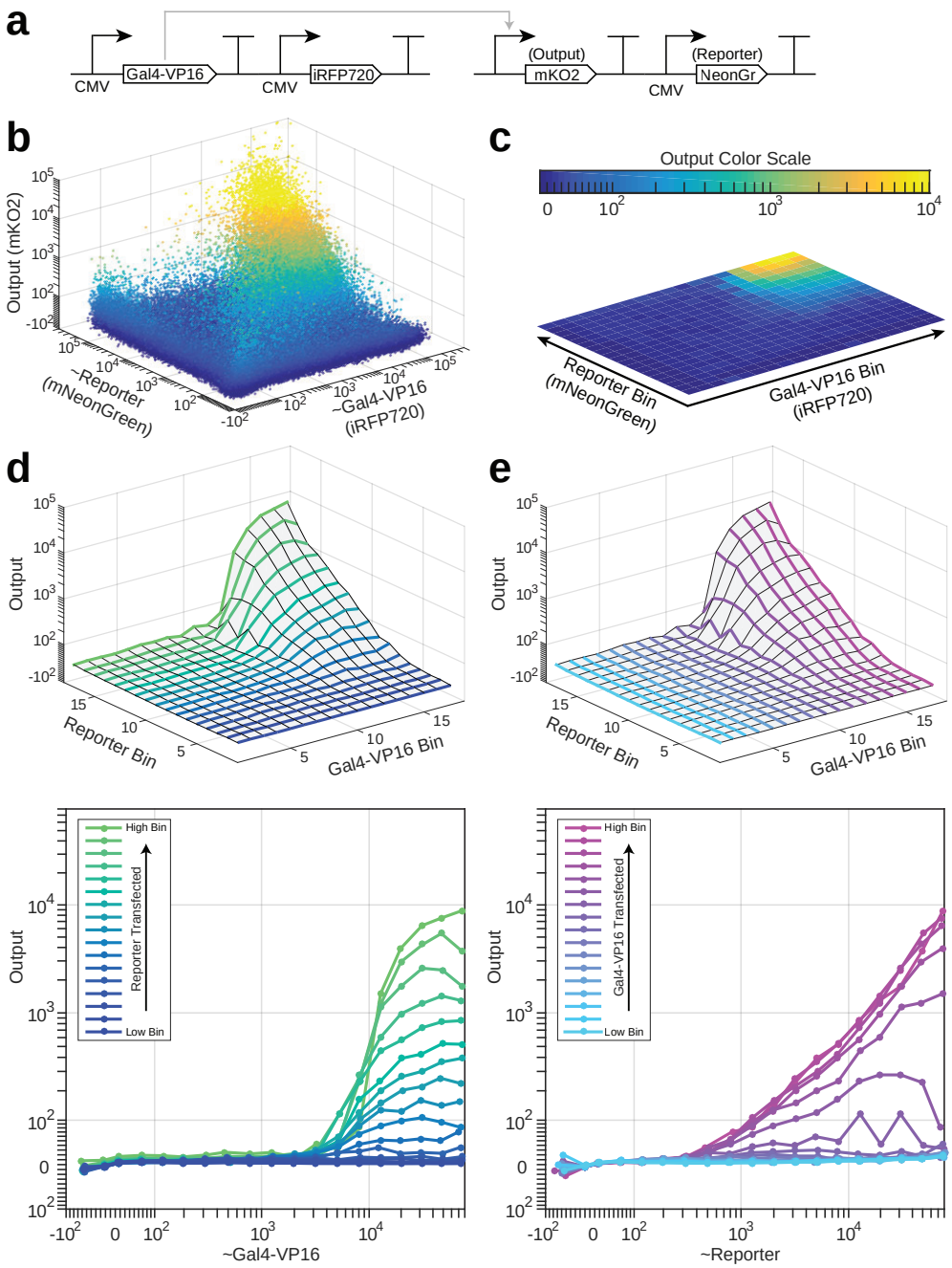

a
CMV
Gal4-VP16
CMV
iRFP720
(Output)
mKO2
CMV
(Reporter)
NeonGr
b
Output (mKO2)
~Reporter (mNeonGreen)
~Gal4-VP16 (iRFP720)
c
Output Color Scale
Reporter Bin (mNeonGreen)
Gal4-VP16 Bin (iRFP720)
d
Output
Reporter Bin
Gal4-VP16 Bin
e
Output
Reporter Bin
Gal4-VP16 Bin
High Bin
Reporter Transfected
Low Bin
Output
~Gal4-VP16
High Bin
Gal4-VP16 Transfected
Low Bin
Output
~Reporter

Figure 4-27: One-pot characterization of Gal4-VP16 activation.
(a) Circuit diagram for Gal4-VP16 activation. The first plasmid contains both constitutive expression of Gal4-VP16 and iRFP720. iRFP720 fluorescence serves as an indicator for the amount of DNA encoding Gal4-VP16 transfected into any given cell. The second plasmid contains an mKO2 fluorescent output that is activated by Gal4-VP16 and an mNeonGreen transfection marker that serves as a reporter for how much output plasmid was delivered to each cell. **(b)** 3D scatter plot for raw data obtained by poly-transfection. The two input parameters - Gal4-VP16 and reporter concentrations - are plotted on the x and y axes, while the output is plotted on the z-axis. Units for each axis are given by fluorescence units (AU) for the associated fluorescent proteins indicated in parentheses. Each point represents one cell colored according to output fluorescence using the same colorbar as in (c). Poly-transfected cells span the entire input parameter space, as demonstrated by coverage across Gal4-VP16 and reporter dimensions. **(c)** Heatmap of Gal4-VP16 activation. Data were binned by subdividing the data according to the two input dimensions and median output fluorescence for each bin was calculated. Medians are shown as different colors according to the indicated colorbar. **(d)** Surface plot and transfer curves where Gal4-VP16 is varied while reporter is held constant. Surface plots for the binned data provide a different way to illustrate the combined effects of either Gal4-VP16 and reporter on the output fluorescence. Colored lines represent transfer curves where reporter concentration is held constant (blue = low, green = high concentration) while Gal4-VP16 is increased. As Gal4-VP16 increases, output rises in a threshold-like manner, a behavior maintained across all reporter transfection levels. **(e)** Surface plot and transfer curves where Gal4-VP16 is held constant while reporter is varied. Colors represent transfer curves where Gal4-VP16 is constant (teal = low, pink = high concentration) while reporter is increased. As the reporter increases, output rises in a linear manner, showing that maximal output is limited by how much output plasmid is delivered via the transfection.

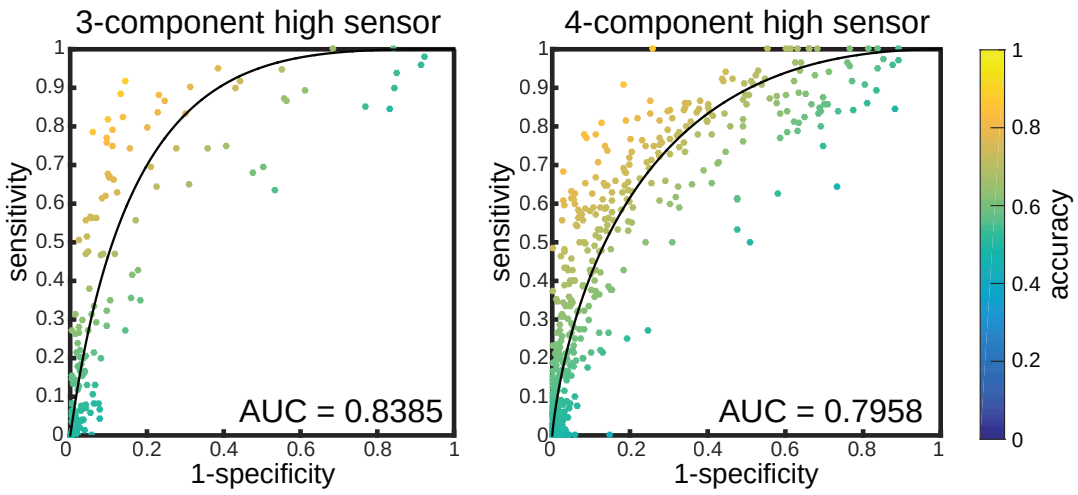
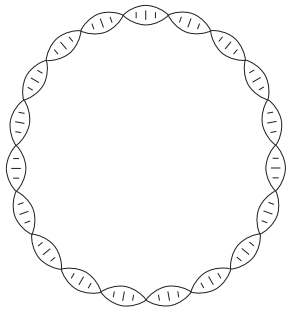


Figure 4-28: ROC-like curves to compare 3- and 4-component miRNA classifiers.
We derived ROC-like curves from the data for both 3-component (Fig. 4-25) and 4-component (Fig. 4-26) miRNA classifiers by taking the sensitivity and specificity within each bin and plotting the resulting scatter plot. Lines were fit according to a bi-normal model and area under the curve (AUC) was calculated. Points are colored according to accuracy in each bin. The AUC can give a general measure of how well each circuit design performs relative to the other, since all areas of parameter space are used to calculate the ROC curve. In this case the AUC values for 3- and 4-component classifiers were similar, indicating minimal performance gain by addition of tTA activation. Thus we chose to focus on the 3-component classifier since it would be easier to tune due to the reduced numbers of components.



This page intentionally left blank to keep captions near figures.

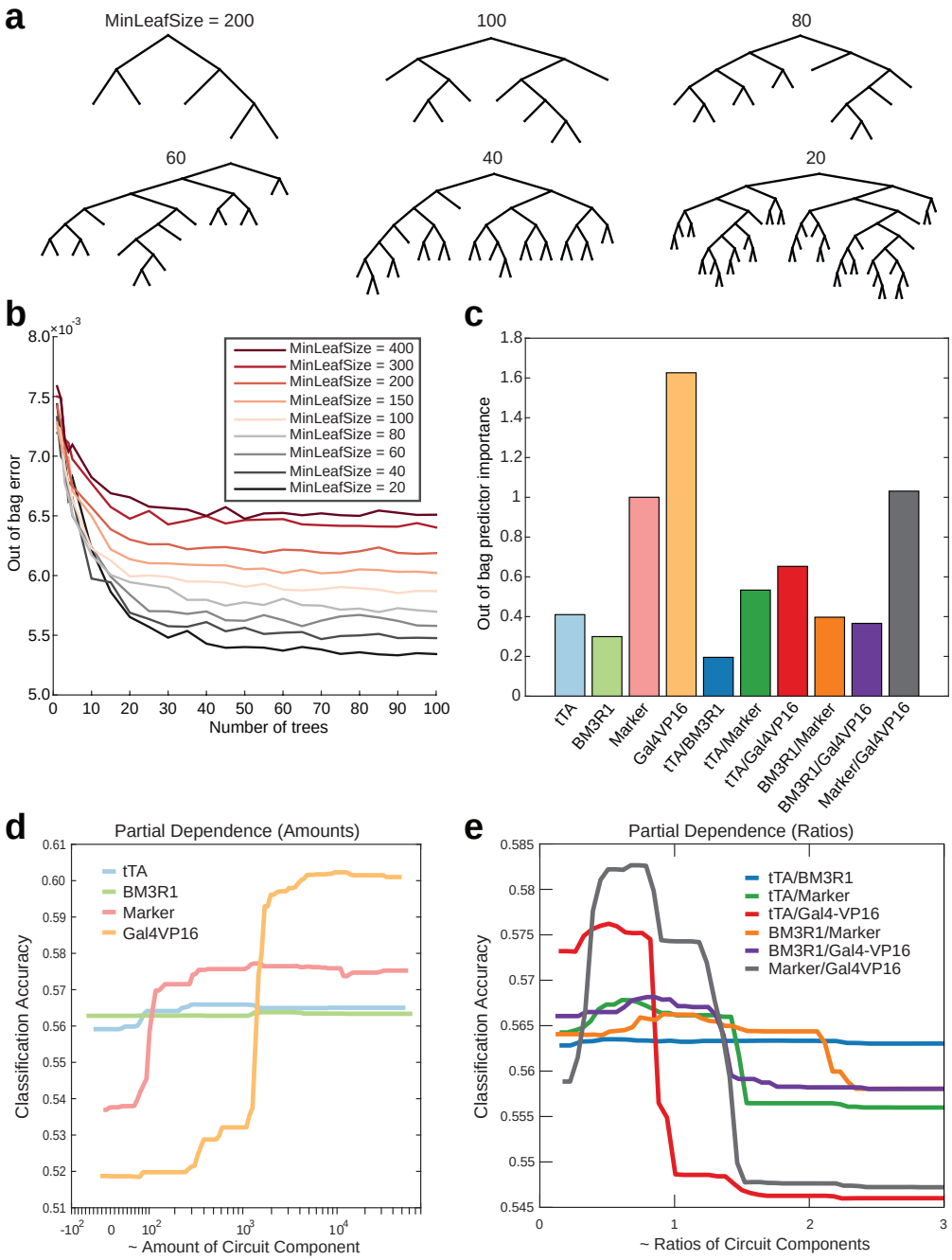

a
MinLeafSize = 200
100
80
60
40
20
b
Out of bag error
Number of trees
MinLeafSize = 400
MinLeafSize = 300
MinLeafSize = 200
MinLeafSize = 150
MinLeafSize = 100
MinLeafSize = 80
MinLeafSize = 60
MinLeafSize = 40
MinLeafSize = 20
c
Out of bag predictor importance
tTA
BM3R1
Marker
Gal4VP16
tTA/BM3R1
tTA/Marker
tTA/Gal4VP16
BM3R1/Marker
BM3R1/Gal4VP16
Marker/Gal4VP16
d
Partial Dependence (Amounts)
Classification Accuracy
~ Amount of Circuit Component
tTA
BM3R1
Marker
Gal4VP16
e
Partial Dependence (Ratios)
Classification Accuracy
~ Ratios of Circuit Components
tTA/BM3R1
tTA/Marker
tTA/Gal4-VP16
BM3R1/Marker
BM3R1/Gal4-VP16
Marker/Gal4VP16

Figure 4-29: Analysis of 4-component classifier with machine learning. We utilized random forest regression to analyzed the binned accuracy data for the 4-component miR-21-5p classifier in order to derive design rules insight into the circuit. The following example serves as a proof of concept for analyzing even higher dimensional data that could be obtained via poly-transfections in the future. We chose random forest regression due to the ease of implementation while still retaining the ability to regress more complicated datasets. **(a)** The effect of minimum leaf size (MinLeafSize) on the size of decision trees comprising the random forest. As MinLeafSize was decreased, trees contained more branches and likely higher risk of overfitting. **(b)** Optimization of MinLeafSize and number of trees in the random forest. Plotting out of bag error as a function of number of trees and MinLeafSize shows that around 80 trees are sufficient to obtain minimal error. We chose to use MinLeafSize of 100 to reduce the likelihood of overfitting while still maintaining relatively low error. **(c)** Parameter importance for amounts of each circuit component and also the ratios of each. We chose to include ratios (i.e. diagonal splits in parameter space, rather than only vertical or horizontal) since often behavior of a circuit can be accounted for by ratiometric measures. Marker and Gal4-VP16 amounts appear most important for circuit function, followed by the ratios of tTA or BM3R1 (the repressing side of the circuit) to Gal4-VP16 or marker (the activating side of the circuit). This suggests that contributions for repressing and activating components should be balanced for optimal miRNA sensing. **(d)** Partial dependence plots for amounts of circuit components show that a prerequisite amount of marker ($\sim 10^2$ AU) and Gal4-VP16 ($\sim 10^3$ AU) are necessary for classification accuracy. This is apparent when looking at the binned accuracy data, as accuracy is only high when both of these conditions are met (Fig. 4-26). **(e)** Partial dependence plots for the ratios of circuit components shows that generally better classification is obtained when components are present at similar ratios. In general, repressing components (tTA, BM3R1) should be equal or less than activating components (marker, Gal4-VP16). Also important was the marker/Gal4-VP16 ratio which should be similar. As both one-pot methods and machine learning are further developed in the future we expect even greater insight into circuit design rules to be possible, in addition to facilitation of larger, more complex and sophisticated circuits.

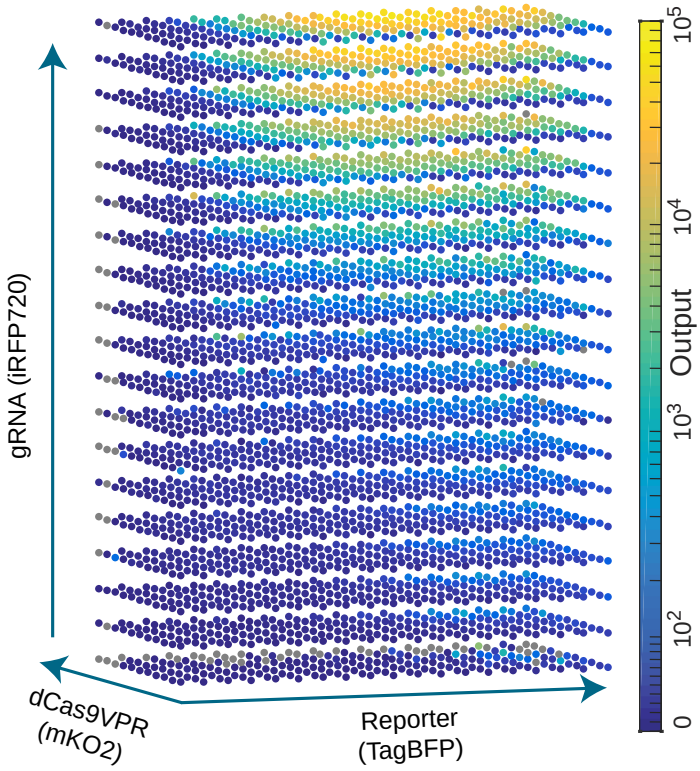


Figure 4-30: Heatmap for binned VPR activation data across the full input parameter space.
Plotted is the fluorescent output as a function of the three different circuit components (dCas9-VPR, gRNA, and output plasmid). Output fluorescent is colored according to the indicated scale bar (AU). Note as previously indicated (Fig. 4-6h) that there is a graded response of output with increasing gRNA, but more threshold-like behavior with dCas9-VPR. Addition of reporter plasmid results in graded increase in output expression, though interestingly there appears to be a slight decrease at very high concentrations. As a result, the optimal region for activation occurs at high dCas9-VPR and gRNA concentration and medium-high reporter plasmid concentration.

			Recommended Values			
	Step	Parameter	2-complex	3-complex	4-complex	n-complex
Day 0	Grow appropriate number of cells	Cells per sample =	> 4e4	> 2e5	>2e6	2,000 x 10^{n-1}
Day 1	Create lipid-DNA complexes	DNA =	30 ng	200 ng	1500 ng	30 x (2/n) x 10^{n-1}
		P3000 =	0.1 uL	0.33 uL	2.5 uL	0.05 x (2/n) x 10^{n-1}
		Lipo3000 =	0.1 uL	0.33 uL	2.5 uL	0.05 x (2/n) x 10^{n-1}
		Opti-MEM = (each per complex)	10 uL	16.5 uL	125 uL	2.5 x (2/n) x 10^{n-1}
	Trypsinize and plate cells	Cells per sample =	> 4e4	> 2e5	>2e6	2,000 x 10^{n-1}
		Media per sample =	100 uL	0.5 mL	2x 2mL = 4mL	varies
		Plate format	96-well	24-well	2 wells in 6-well	varies
	Add lipid-DNA complexes to cells	DNA total =	2x30 = 60 ng	3x200 = 600 ng	4x1500 = 6 ug	30 x 2 x 10^{n-1}
		Volume total =	2x10 = 20 uL	3x17.2 = 52 uL	4x130 = 520 uL	2.6 x 2 x 10^{n-1}
	Incubate cells	Time =	48 hours	48 hours	48 hours	48 hours
Day 3	Analyze cells	Cells per sample =	As many as sampler allows (ideally 8e4)	>4e5	>2e6	~2,000 x 10^{n-1}

Table 4.1: Quick start guide for poly-transfections.
Overview of poly-transfection method and recommended volumes and amounts for preparing and transfecting cells. Different values are given depending on the number of complexes needed for the genetic circuit. Values shown here are for HEK293FT cells; note that optimal values may change depending on transfection efficiency, cell size etc. Further details may be found on Protocol Exchange #6667.

Sample	**BM3R1** (ng)	**Output** (ng)	**Gal4-VP16** (ng)
1	5	60	435
2	10	60	435
3	20	60	435
4	40	60	435
5	60	60	435
6	80	60	435
7	100	60	435
8	150	60	435
9	200	60	435
10	5	5	435
11	5	10	435
12	5	20	435
13	5	40	435
14	5	60	435
15	5	80	435
16	5	100	435
17	5	150	435
18	5	200	435
19	5	60	10
20	5	60	25
21	5	60	50
22	5	60	100
23	5	60	200
24	5	60	300
25	5	60	400
26	5	60	435
27	5	60	480

Table 4.2: DNA amounts for co-transfection testing of miR-21-5p classifier

4.11.1 Complete workflow for one-pot optimization experiments

Step 1. Determine the components of the circuit to be tested

One should plan in advance what the circuit(s) to be tested will contain. For example throughout this protocol we will plan an experiment to characterize the ability of Gal4VP16 to activate a new design for a Gal4VP16 activatable promoter. Our circuit components include:

1. constitutive expression of Gal4VP16

- Insulator = CTCF binding motif
- Promoter = CMV
- 5' UTR = inert sequence
- Gene = Gal4VP16
- 3' UTR = inert sequence
- Terminator = SV40 poly A

2. fluorescent protein output from a Gal4VP16 activatable promoter.

- Insulator = CTCF binding motif
- Promoter = Gal4VP16 activatable promoter
- 5' UTR = inert sequence
- Gene = Fluorescent protein (later determined to be mKO2)
- 3' UTR = inert sequence
- Terminator = SV40 poly A

Step 2. Determine which fluorescent proteins to use in the experiment

One will need to determine the set of fluorescent proteins to be used, which may depend on what equipment is available to make fluorescent measurements. For the flow cytometer we used (BD Fortessa with 355nm laser + 379/28nm filter, 405 + 450/50, 488 + 530/30, 561 + 582/15, 640 + 710/50) the following sets of fluorescent proteins were demonstrated to work well:

- Two-color experiments = TagBFP + mNeonGreen or mKO2
- Three-color experiments = TagBFP + mNeonGreen + (mKO2 or iRFP670 or iRFP720)
- Four-color experiments = TagBFP + mNeonGreen + mKO2 + iRFP720
- Five-color experiments = Sirius + TagBFP + mNeonGreen + mKO2 + iRFP720

Use of additional colors is possible, especially with spectral analyzers. Though one will need to make sure that bleedthrough of signal from one fluorescent protein to another is minimal (after compensation if necessary).

For our example, we will use TagBFP as a transfection marker for Gal4VP16 expression, mKO2 as the output signal, and mNeonGreen as the transfection marker for the output plasmid. This design results in the following transfection marker components to be assembled:

3. constitutive expression of TagBFP

- Insulator = CTCF binding motif
- Promoter = CMV
- 5' UTR = inert sequence
- Gene = TagBFP
- 3' UTR = inert sequence
- Terminator = SV40 poly A

4. constitutive expression of mNeonGreen

- Insulator = CTCF binding motif
- Promoter = CMV
- 5' UTR = inert sequence
- Gene = mNeonGreen
- 3' UTR = inert sequence
- Terminator = SV40 poly A

Note that not all fluorescent proteins may be suitable for multi-color cytometry. We found that almost all red fluorescent proteins had significant signal in the 450/50 channel with a strange stochastic behavior. We believe this is due to a blue intermediate during folding of the red fluorescent proteins. For this reason, we chose to use an orange fluorescent protein (mKO2) instead which has some spillover into the 530/30 channel but has predictable behavior which can be accounted for with compensation. We also observed that yellow fluorescent proteins spill into the 379/28 channel so we used green fluorescent proteins instead. Finally, some red fluorescent proteins have signal in the 710/50 channel, limiting the choice of red/orange FP.

Step 3. Clone needed plasmid level 0 (pL0) constructs

One should clone the needed parts (eg. promoters, coding sequences, etc) into pL0 plasmids with the appropriate Golden Gate assembly overhangs - or obtain existing pL0's from other researchers. pL0 plasmids generally have spectinomycin resistance and the backbones have LacZ which is displaced after assembly for blue/white screening. pL0 overhangs are as follows and are generated by the BbsI restriction enzyme for cloning into the pL0 backbone:

- pL0-I (insulator) = [GGAG] - [TACT]
- pL0-P (promoter) = [TACT] - [CAGA]
- pL0-5 (5' UTR) = [CAGA] - [AGGT]

- pL0-G (gene/coding sequence) = [AGGT] - [GCTT]
- pL0-3 (3' UTR) = [GCTT] - [CAAC]
- pL0-T (terminator/pA) = [CAAC] - [CGCT]

For example, a pL0-P with CMV can be obtained by generating the double-stranded DNA fragment below by PCR or synthesis and cloned into the pL0-P backbone vector with BbsI Golden Gate (bolded sequences correspond to BbsI recognition site and underlined sequences correspond to BbsI cut site):

NNN**GAAGAC**NN<u>TACT</u> - CMV sequence - <u>CAGA</u>NN**GTCTTC**NNN

In general, Golden Gate reactions consist of the following:

- 40 fmol of each DNA part
- Water to a total reaction volume of 10 uL
- 1 uL of T4 ligase buffer (Promega)
- 0.4 uL of Type IIs restriction enzyme (NEB)
- 0.2 uL of T4 ligase HC (Promega)

Golden Gate thermocycling consists of the following:

1. 37C for 5min
2. 50 cycles of 37C for 2min, 16C for 5min
3. 50C for 5min
4. 80C for 10min
5. Hold at 4C

Golden Gate reactions can be transformed into chemically competent cells at relatively high efficiency.

Step 4. Clone plasmid level 1 (pL1) constructs containing transcription units

Using the pL0 constructs, one should assemble transcription units (TUs) into pL1 backbones. These should contain one each of pL0's designated as I, P, 5, G, 3, and T. In most cases, circuit components should be assembled into the pL1-S1S2 backbone while fluorescent protein transfection markers should be assembled into the pL1-S2S3 backbone. Assembly of pL1 plasmids is conducted using BsaImediated Golden Gate assembly.

For our example, we will assemble:

1. pL1-S1S2 with TU for constitutive Gal4VP16
2. pL1-S1S2 with Gal4VP16 activatable promoter and mKO2 output
3. pL1-S2S3 with constitutive expression of TagBFP
4. pL1-S2S3 with constitutive expression of mNeonGreen

Step 5. Clone plasmid level 2 (pL2 constructs) that contain one circuit TU and one transfection marker TU

Using Gibson assembly, one should assemble together pairs of constructs with one circuit TU and one transfection marker TU. In most assemblies the pL2-S3S1 backbone may be used, but other Gibson backbones with Weiss lab overlapping sequences can also be used. Each plasmid should be digested with I-SceI restriction enzyme, PCR purified or gel extracted, and the linear fragments used for a Gibson reaction. We obtained reaction mix from SGI as the Gibson Ultra Kit and followed the manufacturer's protocol for assembly. We electroporated assembly products electrocompetent cells (NEB 10-beta cells) as we generally found larger constructs harder to transform with chemically competent cells.

We have also deposited pL2 backbones already containing constitutive expression of different fluorescent proteins, such that only a single Golden Gate step is required to assemble functional pL2 vectors (Addgene vectors #109150-109154).

In our example, pL1-S1S2 with constitutive Gal4VP16 (#1) should be assembled with pL1-S2S3 with TagBFP (#3) and pL2-S3S1 to generate construct A. And also pL1-S1S2 with activatable promoter and mKO2 (#2) should be assembled with pL1-S2S3 with mNeonGreen (#4) and pL2-S3S1 to generate construct B. This should generate two final plasmids for poly-transfection. Both constructs should be midiprepped to obtain transfection quality DNA with reduced endotoxin concentrations.

Alternatively, it is possible to conduct poly-transfections using pL1's instead of pL2's. In this case, equimolar concentrations of pL1's should be mixed together in each lipid complexes, such that mixtures contain the same constructs as those assembled into pL2's listed here.

Step 6. Grow and maintain cells of interest

Cells of interested should be obtained from a reliable source, thawed into the appropriate growth medium, and verified to be free of mycoplasma or other contaminants. Cells should be regularly split into new flasks/dishes when appropriate confluence is reached (use the appropriate maintence conditions provided for the cells) using trypsin or other dissociation medium. For our study, we grew HEK293FT cells (Thermo Fisher) and HeLa cells (ATCC) in DMEM DMEM with 4.5 g/L glucose, L-glutamine, and sodium pyruvate (Cellgro) and supplemented with 10% FBS (Cellgro) and at 37C / 5% CO2 . Before transfection, one should calculate approximately how many cells will be required for the experiment and determine if the necessary

number of cells have grown. We use 200,000 HEK293FT cells per well or 150,000 HeLa cells per well in a 24-well plate and then scale that number according to plate surface area.

Step 7. Prepare transfection mixtures

Transfection mixtures can be separately prepared according to the manufacturer's protocol, then simultaneously added to cells. We have mostly used lipofectamine 3000 for transfection, but have shown that poly-transfections work with a variety of transfection agents.

In our protocol, we diluted 900 ng of each input plasmid with 75 uL of Opti-MEM (Thermo Fisher), mixed in 1.5 uL of P3000 reagent, then mixed in 1.5 uL of Lipofectamine 3000 reagent. Each complex was mixed separately and incubated for 30 min before being added to cells plated in a 6-well format. Note that in a 24-well format, these volumes correspond to 150 ng of DNA in 12.5 uL of OptiMEM, 0.25 uL of P3000, and 0.25 uL of lipofectamine per complex. If using transfection markers on a separate plasmid from the circuit component, one can use equimolar concentrations of each to a total of 150 ng of DNA (24-well format) or 900 ng (6-well) then add the corresponding amount of Opti-MEM and lipofectamine reagents.

Step 8. Add transfection mixtures to cells

Add the desired complexes to cells. In our example, separate complexes prepared for constructs A and B would be applied to cells at the same time and cells would be allowed to grow in order to express the circuit components as well as fluorescent proteins. In our experiments we waited 48 hours between transfection and flow cytometry.

Step 9. Prepare cells for flow cytometry

Cells should be suspended into PBS or FACS buffer using methods appropriate for the cell lines tested. For our adherent cells, we trypsinized cells (0.2 mL in 24-well), resuspended them in growth media (0.5 mL in 24-well), spun then down at 100 x g, aspirated the media and resuspended in PBS (0.5 mL for 24-well).

Step 10. Prepare flow cytometer and run calibrations

The flow cytometer should be operated according to the manufactuer's protocols. Single color controls should be run and PMT voltages set to obtain best dynamic range without saturation of highly expressing cells. We highly recommend running multi-peak fluorescent calibration beads (we have used Spherotech URCP-100-2H or RCP-30-5A) to make sure that the cytometer is running correctly and to enable more quantitative and reliable measurements that are reproducible across different cytometers.

Step 11. Run samples through the flow cytometer

After running controls through the cytometer, samples should be run, making sure that a wide range of fluorescent color space is covered when fluorescent channels are plotted against each other. As many cells as necessary should be collected such that an adequate number of events are recorded across the fluorescent color space. Generally we used all cells from a well in a 24-well plate for 2- complex experiments, a 6-well plate for 3-complex experiments, and 2 pooled wells from a 6-well plate for 4-complex experiments.

Step 12. Export data and finish cytometer session

Data should be exported as fcs files and the cytometer cleaned and shut down.

Step 13. Set up existing analysis pipelines

We have developed code to quickly bin data along multiple dimensions using MATLAB. The code can be found at GitHub under `https://github.com/jonesr18/MATLAB_Flow_Analysis` Additonal code may also be needed for downstream analysis in order to read in fcs files (Laszlo Balkay: `https://www.mathworks.com/matlabcentral/fileexchange/9608-fca-readfcs`), and logicle transform (Rachel Finck: `https://www.mathworks.com/matlabcentral/fileexchange/45022-logicle-transformation`).

Step 14. Process data using the FlowData method

Use a MATLAB script to call the FlowData method for binning. See the attached zip file for an example m-file and fcs files.

Step 15. Further analyze binned data to generate characterization or conclusions

If custom analysis needs to be conducted, binned or raw data can be analyzed and visualized in Matlab.

References

1. Brown, B. D. *et al.* Endogenous microRNA Can Be Broadly Exploited To Regulate Transgene Expression According To Tissue, Lineage and Differentiation State. *Nature Biotechnology* **25,** 1457–1467 (2007).

35. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333,** 1307–1311 (2011).

44. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature genetics* **43,** 854–9 (2011).

88. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).

102. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12,** 326–8 (2015).

104. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for Metabolic Engineering. *Frontiers in Bioengineering and Biotechnology* **3,** 1–11 (2015).

105. Shih, S. C. C. & Moraes, C. Next generation tools to accelerate the synthetic biology process. *Integrative Biology* **8,** 585–588 (2016).

106. Gorochowski, T. E. *et al.* Genetic circuit characterization and debugging using RNA-seq. *Molecular Systems Biology* **13,** 952 (2017).

107. Lu, Y. Cell-free synthetic biology: Engineering in an open world. *Synthetic and Systems Biotechnology* **2,** 23–27 (2017).

108. Ranji, A., Wu, J. C., Bundy, B. C. & Jewett, M. C. in *Synthetic Biology Tools and Applications* 277–301 (Elsevier Inc., 2013).

109. Slusarczyk, A. L., Lin, A. & Weiss, R. Foundations for the design and implementation of synthetic genetic circuits. *Nature Reviews Genetics* **13,** 406–420 (2012).

110. Zhou, H., Vonk, B., Roubos, J. A., Bovenberg, R. A. & Voigt, C. A. Algorithmic co-optimization of genetic constructs and growth conditions: Application

to 6-ACA, a potential nylon-6 precursor. *Nucleic Acids Research* **43,** 10560–10570 (2015).

111. Ghodasara, A. & Voigt, C. A. Balancing gene expression without library construction via a reusable sRNA pool. *Nucleic Acids Research* **45,** 8116–8127 (2017).

112. Woodruff, L. B. A. *et al.* Registry in a tube : multiplexed pools of retrievable parts for genetic design space exploration. *Nucleic Acids Research* **45,** 1553–1565 (2017).

113. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Research* **41,** 10668–10678 (2013).

114. Gach, P. C., Iwai, K., Kim, P., Hillson, N. & Singh, A. K. Droplet Microfluidics for Synthetic Biology. *Lab Chip* **17,** 3388–3400 (2017).

115. Woodruff, K. & Maerkl, S. J. A High-Throughput Microfluidic Platform for Mammalian Cell Transfection and Culturing. *Scientific Reports* **6,** 1–12 (2016).

116. Klein, J. C. *et al.* Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Research* **44,** e43 (2015).

117. Kim, H. *et al.* 'Shotgun DNA synthesis' for the high-throughput construction of large DNA molecules. *Nucleic Acids Research* **40** (2012).

118. Wong, A. S., Choi, G. C., Cheng, A. A., Purcell, O. & Lu, T. K. Massively parallel high-order combinatorial genetics in human cells. *Nature Biotechnology* **33,** 952–961 (2015).

119. Beal, J. Bridging the Gap: A Roadmap to Breaking the Biological Design Barrier. *Frontiers in Bioengineering and Biotechnology* **2,** 1–16 (2015).

120. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352,** aac7341 (2016).

121. Yu, J. *et al.* Human Induced Pluripotent Stemm Cells Free of Vector and Trangene Sequence. *Science* **324,** 797–801 (2009).

122. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

123. Papapetrou, E. P. *et al.* Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proceedings of the National Academy of Sciences* **106,** 12759–12764 (2009).

124. Nagamatsu, G. *et al.* Optimal ratio of transcription factors for somatic cell reprogramming. *Journal of Biological Chemistry* **287,** 36273–36282 (2012).

125. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–824 (2013).

126. Fujii, W., Kawasaki, K., Sugiura, K. & Naito, K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Research* **41,** e187 (2013).

127. Hori, Y., Kantak, C., Murray, R. M. & Abate, A. R. Cell-free extract based optimization of biomolecular circuits with droplet microfluidics. *Lab on a Chip* **17,** 3037–3042 (2017).

128. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 14024–14029 (2013).

129. Li, Y. *et al.* Modular construction of mammalian gene circuits using TALE transcriptional repressors. *Nature Chemical Biology* **11,** 201–213 (2015).

130. Wong, A. S. L. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proceedings of the National Academy of Sciences* **113,** 2544–2549 (2016).

131. Schwake, G. *et al.* Predictive modeling of non-viral gene transfer. *Biotechnology and Bioengineering* **105,** 805–813 (2010).

132. Bhise, N. S., Shmueli, R. B., Gonzalez, J. & Green, J. J. A novel assay for quantifying the number of plasmids encapsulated by polymer nanoparticles. *Small* **8** (2012).

133. Davidsohn, N. *et al.* Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition. *ACS Synthetic Biology* **4,** 673–681 (2015).

134. Bullock, M. D. *et al.* Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression. *Cell Death and Disease* **4,** 1–10 (2013).

135. Ferreira, J. P., Overton, K. W. & Wang, C. L. Tuning gene expression with synthetic upstream open reading frames. *Proceedings of the National Academy of Sciences* **110,** 11284–11289 (2013).

136. Ferreira, J. P., Peacock, R. W. S., Lawhorn, I. E. B. & Wang, C. L. Modulating ectopic gene expression levels by using retroviral vectors equipped with synthetic promoters. *Systems and Synthetic Biology* **5,** 131–138 (2011).

137. Wroblewska, L. *et al.* Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nature Biotechnology* **33,** 839–841 (2015).

138. Hawley, T. S., Hawley, R. G. & Telford, W. G. Fluorescent proteins for flow cytometry. *Current Protocols in Cytometry* **2017,** 9.12.1–9.12.20 (2017).

139. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. *GAL4-VP16 is an unusually potent transcriptional activator* 1988.

140. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268,** 1766–1769 (1995).

141. Kelleher, R. J., Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61,** 1209–1215 (1990).

142. Wang, T., Xie, Y., Tan, A., Li, S. & Xie, Z. Construction and characterization of synthetic microRNA cluster for multiplex RNA interference in mammalian cells. *ACS Synthetic Biology* **565,** acssynbio.5b00180 (2015).

143. Saito, H. *et al.* Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nature Chemical Biology* **6,** 71–78 (2010).

Chapter 5

Conclusions and future directions

In this work we have described several advances in applying engineering principles towards the design and implementation of new genetic circuits, as well as key technologies for facilitating the engineering cycle itself. Our work towards generating fundamental design rules and predictive models for miRNA repression should enhance the construction of more complex miRNA sensors and classifiers in the future. Our advances to the design-build-test-learn (DBTL) cycle will also translate more generally to other projects in synthetic biology. The main contributions to the DBTL cycle include: the formulation of the Ant/Syn model (design), faster and easier DNA assembly techniques - namely microfluidics and DNA assembly plasmids like the multi-assembly vector and low sensor backbones (build), poly-transfection (test), and quantitative analysis along with machine learning to parse diverse datasets (learn). While we have made several advances in these areas, key limitations and future work remain for each.

Design: The Ant/Syn model we describe explains some of the combinatorial effects that occur when multiple miRNAs regulate a synthetic reporter. However, the

Ant/Syn model is only one step towards a more ideal model which would be able to predict miRNA repression effects of any sequence of DNA - including endogenous or even random sequences. As such, future models will need to be improved in several key ways. First, the mechanisms for Ant/Syn and other interactions must be ascertained so that models may be updated accordingly. Such studies will also enhance the general understanding of cell signaling and could have impacts in areas where miRNAs are disregulated (e.g. cancer). Second, researchers should investigate possible explanations for the poor correlation between miRNA expression and miRNA activity. If such effects could be adequately explained such that miRNA activity could be accurately inferred from miRNA expression data, miRNA repression models could take advantage of the vast quantity of small RNA abundance data that is available. As it is now, abundance data is only modestly useful and primarily used to make educated guesses as to which miRNAs might be useful to sense, followed by experimental determination of miRNA activity.[11] We expect that significant and collaborative work will be required to build a 'miRNA activity dictionary' that can relate these two quantities in a wide range of useful cell types. Finally, prediction of miRNA target sites and their relative strengths will need to be more reliable. Current miRNA target site predictions generate many false positives,[144] meaning that the are of little use for predicting the strength of miRNA repression for endogenous transcripts for instance.

Build: We have made several advances in our DNA assembly platform to speed up construction of genetic circuits. Our lab has previously adapted and expanded the previous modular cloning (MoClo[145]) system to include mammalian parts.[54] More recently we have used a Golden Gate-Gibson system* to assemble even larger cir-

*work pioneered by Dr. Jin Huh in the Weiss lab

cuits, where transcription units (TU) are assembled via Golden Gate as in MoClo, but multi-TU plasmids are assembled using a Gibson step[146] similar to the another previous assembly strategy.[53] For miRNA sensing circuits we built a common backbone, termed the miRNA low sensor backbone (LSB), that includes four repeats of sequence validated target sites. LSB plasmids can be used directly for transfection experiments, providing a red fluorescent signal from mKate2 that is regulated by the miRNA target sites, and a constitutive blue signal from EBFP2 for reference. Sensors can also be integrated into the genome using BxB1 recombinase.[54] Moreover these backbones can serve as PCR templates for amplification of miRNA targets for assembly into other miRNA sensors - including multi-input ones - since the overhangs can be easily changed during the PCR step. The low sensor backbone is also one of several designs based on a multi-assembly vector, where TUs can be assembled by a Gateway step to insert the promoter and gene, the 3' UTR is added with a Golden Gate step, and then multiple TUs can be assembled together using Gibson. We anticipate that the flexibility of the Golden Gate-Gibson assembly will make it an invaluable tool for assembling larger and more complex circuits in mammalian synthetic biology, with up to 7 TUs assembled into a single plasmid. These and future developments should aid the development of larger systems, for instance metabolic engineering in mammalian cells, once other barriers to their design are overcome. Future advances in DNA assembly might utilize ligase cycling reaction (LCR)[56] instead of Gibson assembly, which may offer even more efficient assembly of many different TUs in a single step. One long term challenge is how to assemble constructs together without having to worry about the presence of Type IIS recognition sites within the assembled sequences. At this time, such sites are often mutated to avoid unwanted cleavage, but these mutations in many types of sequences (e.g. promoters, poly-A signals, enhancers, insulators) may affect their function. However, leaving in

the Type IIS recognition sites can severely hamper assembly efficiency, especially if there are several such sites. Other future improvements in DNA assembly could involve miniaturization of assembly reactions to reduce cost and increase throughput. In collaboration with Lincoln Laboratory we have explored the use of microfluidics to streamline DNA assembly.[147] We anticipate these and other microfluidic methods (e.g. droplet microfluidics[148]) will greatly accelerate the build portion of the design-build-test-learn cycle.

Implementation and design of miRNA sensors/classifiers also have room for improvement. Most miRNAs have small (∼2-fold) effects on transcripts they regulate,[71,149] which can often limit the dynamic range possible with sensors. Circuits that have ultrasensitive behavior - steep input-output curves - could potentially be used to amplify small changes in input miRNA signal to larger changes in output. However, work in implementing ultrasensitivity in mammalian systems is an ongoing effort. We have also encountered technical limitations when making miRNA activity measurements. In our work, we primarily use transfections to deliver miRNA sensors to cells due largely to the simplicity of the method. However, this limits our data to those cell lines which can be easily transfected - namely cell lines. While other platforms like AAV could possibly offer better and more uniform transduction of cells,[43] several cell lines and cell types are still lowly transduced with such methods.[150] We believe cell-free methods could prove a useful alternative for measuring miRNA activity. In such experiments, miRNA sensors could be added to cell lysate and output could be measured as bulk fluorescence or luminscence. Since DNA would not need to be delivered past the cell wall, presumably any cell type could be measured, including primary cells, assuming enough cells can be extracted. Possibly the reactions could be scaled down using microfluidics in order to make measurements in rarer cell types possible.

Test: We have developed the poly-transfection method in order to obtain significantly more information per transfected sample. Since the formation of independent transfection complexes requires minimal modification of traditional transfection procedures, we imagine that most labs could easily adopt poly-transfections. There are some remaining research questions that will need to be answered for poly-transfections to be used to their full potential. Poly-transfection and co-transfections yield largely similar data, though in some systems the outputs are up to ∼2-fold different. This discrepancy has minimal effect when characterizing parts for building medium-scale systems (∼4 components), though the cumulative discrepancy could be significant if using separate characterization data to predict function of larger systems. It should be noted however that a poly-transfection of the larger system itself should not be affected. There are limitations on how large a system can be tested with poly-transfection. Namely the number of cells required to sample parameter space for a poly-transfection at a given coverage still grows exponentially with the number of components. Thus poly-transfections drastically simplify and speed up testing for small- and medium-size systems but their applicability to large systems remains to be seen. First of all, the number of distinguishable fluorescent proteins with the flow cytometers we used was around five. Spectral analyzers have the potentially to immediately increase the number fluorescent proteins that can be measured, since they could allow better deconvolution of fluorescent proteins with overlapping emission spectra. In the future, mass cytometers could further increase the number of possible signals, since using heavy metal ions as readout currently enables readout from 23 signals. Though a significant amount of optimization and testing would be required to link transcriptional output to a metal ion signal. Another technology that could further extend the number of components that can be measured with poly-transfections may be a combination of RNA barcodes expressed

from each component and single-cell sequencing. A much simpler approach would be to sample the parameter space at lower (perhaps sparse) coverage and use algorithms to impute or smooth the data. Again significant optimization would be required to determine what degree of coverage is necessary for what kinds of genetic circuits.

Learn: Throughout this thesis, we have used custom MATLAB scripts* to conduct more quantitative analysis of flow cytometry than possible with commercial software such as FlowJo. Previously, researchers would gate cells and measure mean, median, or geometric means of the output, under the assumption that delivery of DNA was equally efficient in all samples. However, this method neglects the transfection distribution present which ranges across several orders of magnitude within a single sample. Moreover, any variation in transfection efficiency across samples could have drastic effects on measured output. For instance, if comparing two transcriptonal activators and one activates 2-fold higher but transfects <0.5-fold as efficiently (due to plasmid size or preparation quality), it would be deemed a worse activator. Instead, building upon previous work,[133] we use binning for one or several transfection markers to account for transfection efficiency within the sample. We have used binned data to fit biochemical models - as in the miRNA repression model - and used binning across multiple transfection markers to characterize genetic circuits using poly-transfection. We also demonstrated proof of concept work for using machine learning to determine design rules for a genetic circuit, from binned poly-transfection data. Significant future work will be required to make this analysis easy to use. Recent work in our lab has been oriented towards building a similar analysis platform in Python with an intuitive graphical user interface†. These examples illustrate the

*work spearheaded by Breanna DiAndreth and Ross Jones within the Weiss lab

†Work by Dr. Brian Teague in the Weiss lab; CytoFlow: `http://bpteague.github.io/cytoflow/`

growing need to quantitatively understand genetic circuits and systems in order to make more reliable predictions. We anticipate that a key part of that understanding will be analysis based on the techniques we have shown here.

Our work presented here has advanced several applied and foundational areas of synthetic biology. The library of miRNA sensors and Ant/Syn model will aid the design of future miRNA classifiers and even RNAi knockdown efforts, and poly-transfections and quantitative analysis will promote the characterization of modules and systems in mammalian cells. These developments not only facilitate exciting applications in mammalian synthetic biology, but also serve as a foundation for biologists to probe and model endogenous cellular signaling in new ways.

References

11. Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods* **9,** 840–846 (2012).

43. Tian, W. *et al.* High-Throughput Functional MicroRNAs Profiling by Recombinant AAV-Based MicroRNA Sensor Arrays. *PLoS ONE* **7,** 1–10 (2012).

53. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Research* **41** (2013).

54. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Research* **42,** 13440–13451 (2014).

56. Kok, S. D. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synthetic Biology* **3,** 97–106 (2014).

71. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).

133. Davidsohn, N. *et al.* Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition. *ACS Synthetic Biology* **4,** 673–681 (2015).

144. Pinzon, N. *et al.* MicroRNA target prediction programs predict many false positives. *Genome Research* **27,** 234–245 (2017).

145. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6** (2011).

146. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–5 (2009).

147. Kong, D. S. *et al.* Open-source, community-driven microfluidics with Metafluidics. *Nature Biotechnology* **35,** 523–529 (2017).

148. Shih, S. C. C. *et al.* A Versatile Microfluidic Device for Automating Synthetic Biology. *ACS Synthetic Biology* **4,** 1151–1164 (2015).

149. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).

150. Ellis, B. L. *et al.* A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. *Virology Journal* **10,** 1 (2013).

Chapter 6

Appendix

6.1 Methods for DNA assembly with microfluidics

6.1.1 Genetic circuit assembly

For each genetic circuit assembly, pairs of assembly reactions were performed both at conventional volume scales (2 to 10 uL depending upon the biochemistry) and in the microfluidic device (300 to 650 nL scale). The reactions performed in conventional volume formats (200 uL PCR tubes) are hereafter referred to as "tube" reactions. Assembled genetic circuits were then transformed into E. coli for clonal selection, validated by restriction digestion, and finally tested in live cells to verify functionality. Ligation Assembly Protocol Ligation assembly was utilized to construct a 5,362 bp genetic circuit featuring a GFP reporter expressed from a constitutive pTet promoter. The circuit was assembled from two BioBricks30– a constitutive GFP insert (`BBa_I13522`) and high copy plasmid backbone (`I140322`). These BioBricks were previously digested with EcoRI and PstI to generate linear fragments for ligation assembly. The `I13522` GFP insert and `I14032` backbone fragments were gel extracted,

yielding a concentration of 26.8 ng / ml and 47.1 ng /ml, respectively. Tube reactions were performed at a 2 uL scale. 0.5 uL of T4 DNA Ligase, 0.5 uL of Ligase buffer, 0.5 uL of GFP insert and 0.5 uL of backbone were mixed and incubated at 25 °C for 30 min.

6.1.2 Ligation assembly protocol

Ligation assembly was utilized to construct a 5,362 bp genetic circuit featuring a GFP reporter expressed from a constitutive pTet promoter. The circuit was assembled from two BioBricks30– a constitutive GFP insert (`BBa_I13522`) and high copy plasmid backbone (`I140322`). These BioBricks were previously digested with EcoRI and PstI to generate linear fragments for ligation assembly.

The `I13522` GFP insert and `I14032` backbone fragments were gel extracted, yielding a concentration of 26.8 ng / ml and 47.1 ng /ml, respectively. Tube reactions were performed at a 2 uL scale. 0.5 uL of T4 DNA Ligase, 0.5 uL of Ligase buffer, 0.5 uL of GFP insert and 0.5 uL of backbone were mixed and incubated at 25 °C for 30 min.

Antibiotic selection was performed by transforming 1 uL of each tube reaction into 50 uL of chemically competent, high efficiency DH5alpha E. coli one-shot cells (Invitrogen), followed by incubating on ice for 30 min. The cells were then heat shocked at 42 °C in a water bath for 45 s and then returned to ice for 2 min. 300 uL of SOC was then added, followed by shaking incubation at 225 rpm and 37 °C. Finally, 1 uL of the 350 uL SOC mixture was plated onto LB with ampicillin (100 ug / mL) after 1 hour of shaking. The 1 uL aliquot was then mixed with 50 uL of SOC on the plate to allow even spreading of the mixture. Plates were inverted and incubated at 37 °C overnight and then colonies were counted. Assembled constructs

were then verified by restriction digestion with enzymes PstI and EcoRI resulting in 4kb bp and 949 bp fragments, respectively.

6.1.3 Ligation-assembled GFP expression circuit biological validation

The assembled plasmids were biologically validated for constitutive GFP fluorescence in E. coli. 1 uL of each DNA miniprep was transformed into 10G cells (Lucigen) by heat shock and plated onto plates containing 50 ug / mL kanamycin. Single colonies were picked and grown in 2 mL cultures of kanamycin-containing LB overnight. Overnight cultures were diluted to OD600=2e-6 and grown for 6 hours to an OD600 of 0.01. GFP fluorescence was analyzed by flow cytometry with an LSRFortessa cell analyzer (BD Biosciences) and FlowJo software (Tree Star).

6.1.4 Gateway assembly protocol

Gateway assembly was utilized to construct a 3,479 bp that constitutively expresses EGFP. 1 plasmid part and 1 entry vector were assembled to form the final construct in a 10 uL bench top reaction. Entry vector plasmid stock pDonr221 was obtained from Life Technologies (part 12536017). Linear insert cGFP was amplified by PCR with Phusion Master Mix from New England Biolabs (part M0531L), using primers attB1 (GGGGACCACTTTGTACAAGAAAGCTGGG) and attB2 (CCCCACAAGTTTGTACAAAAAAGCAGG) and a source plasmid (pEGFP, BD Biosciences). PCR product was analyzed via electrophoresis through a 1% agarose-TAE gel. The PCR fragment was excised from the gel and purified using Qiagen Gel Extraction Kit (part 28706) following the manufacturer's protocol. The concentration of the purified fragment was determined by NanoDrop spectroscopy.

A Gateway BP assembly master mix sufficient for ten 10 uL reactions was made by adding 10 uL of pDonr221 (150 ng / uL stock), 10 uL of cGFP PCR fragment (150 ng / uL stock), and 60 uL TE. For bench top reactions, 8 uL of the DNA master mix was added to 2 uL of BP recombinase (Life Technologies), mixed by pipette and incubated for 16 hours at 25 °C on a PCR machine followed by incubation at 4 C upon completion of the reaction. Reactions (bench top and microfluidic) were then treated with 1 uL of Proteinase K for 10 minutes at 37 °C.

Antibiotic selection was performed by transforming 1 uL of each bench top reaction into 50 uL of chemically competent, high efficiency Mach T1 E. coli one shot cells (Life Technologies, C862003). Transformation involved incubation on ice for 30 min, followed by heat shock at 42 C in a water bath for 45 s and then returning to ice for 2 min. 300 uL of SOC was then added to the mixture, followed by shaking incubation at 225 rpm and 37 °C. A 10 uL aliquot of the 350 uL SOC mixture was plated onto LB and Kanamycin (30 ug / mL) after 1 hour in the shaker. Plates were inverted and incubated at 37 °C overnight and then colonies were counted.

Assembled constructs were then verified by growing colonies in LB and Kanamycin (30 ug / mL) overnight at 37 °C, purifying plasmid DNA (Qiagen spin 250 miniprep), and performing restriction digestion with enzymes AflII and EcoRV, resulting in 2,317 bp and 1,162 bp fragments.

6.1.5 Gateway-assembled GFP expression circuit biological validation

The assembled BP plasmids were biologically validated for constitutive GFP fluorescence in E. coli. 1 uL of each DNA miniprep was transformed into 10G cells (Lucigen) by heat shock and plated onto plates containing 50 ug / mL kanamycin.

Single colonies were picked and grown in 2 mL cultures of kanamycin-containing LB overnight. Control 10G cells (Lucigen) were grown in LB overnight. Overnight cultures were diluted 1000:1 and grown for 3 hours to an OD600 of 0.01-0.04 (0.06-0.08 for control cultures, which grew faster without a plasmid). Cultures and control cultures were then diluted 100:1 in 1X PBS and GFP fluorescence was analyzed by flow cytometry with an LSRFortessa cell analyzer (BD Biosciences). For each sample and control, 100,000 events (after gating by forward and side scatter) were processed with FACSDiva software (BD Biosciences) to determine mean fluorescence. FITC-A laser power was set to 420V.

6.1.6 Gibson assembly protocol

Gibson assembly was utilized to construct a 4708 bp experimental genetic circuit for cell density dependent gene expression. Four linear fragments (1070 bp, 1130 bp, 1132 bp and 1614 bp, respectively) were utilized for each 5 uL Gibson tube assembly reaction. The fragments were generated by polymerase chain reaction (PCR), gelpurified (Qiagen) and concentrated. Plasmid stocks of DNA templates for PCR reactions were made by transforming into DH5alpha E. coli (New England Biolabs, NEB) and purifying plasmid DNA (Qiagen) after overnight growth of a single colony in LB-Miller growth medium (referred to as simply LB hereafter) in the presence of either chloramphenicol or kanamycin. PCR primers were ordered from Integrated DNA Technologies (IDT) and re-suspended in water to a stock concentration of 100 uM. PCR reactions were performed using a Phusion High-Fidelity PCR Kit (NEB) with 0.5 uM of each primer and 50 ng of template plasmid for thirty cycles of 98 °C for 45 s, 55 °C for 45 s and 72 °C for 2 min at a 300 uL scale per fragment. Fragments were then run on a 1% agarose TAE gel and gel extracted (Qiagen, 3 columns per

fragment). Column eluates were concentrated to 30 uL in a Centricon 10 (Millipore) at 10,000 rpm for 30 min. DNA concentrations were determined by absorbance at 260 nm using a NanoDrop spectrophotometer.

A DNA mixture sufficient for forty 5 uL reactions was made by combining DNA and water to yield 125 fmol of DNA per reaction for each fragment. For these reactions, 2.5 uL of DNA mixture was added to 2.5 uL of Gibson Assembly Master Mix (NEB), mixed by pipette and then placed immediately on a preheated PCR machine for 45 minutes at 50 °C, then cooled to 4 °C.

Antibiotic selection was performed as described above, except transforming into chemically competent, high efficiency DH5alpha E. coli one-shot and plating onto LB and chloramphenicol (15 ug / mL) after 2 hours of shaking. Assembled constructs were then verified by picking individual colonies for screening, purifying plasmids from each clone and performing restriction digestion with enzymes NcoI, XbaI, XhoI and HindIII, resulting in 2262 bp, 1289 bp, 629 bp, 405 bp, and 143 bp fragments, for the expected assembly product.

6.1.7 Gibson-assembled cell density dependent gene expression circuit biological validation

To test circuit function in cells, the assembled circuit was transformed into 10G competent cells and plated on LB-agar plates containing 20 ug/mL chloramphenicol. A single colony was inoculated into 5 mL of LB with 20 ug/mL of chloramphenicol and grown overnight with shaking at 37°C. The overnight cultures were inoculated at an OD600 of 0.001 into cultures containing 1 mL of LB per time point. 1 mL aliquots of the culture were taken at each time point and put on ice to halt bacterial growth and quorum sensing. OD600 was measured for each aliquot using a NanoDrop

2000c spectrophotometer and fluorescence was measured using the LSRFortessa cell analyzer and FACSDiva software.

6.1.8 Golden Gate assembly protocol

Golden Gate assembly was utilized to construct a 5,571 bp aTc-inducible circuit for GFP expression. Five plasmids were assembled to form the final construct in a 10 uL reaction. Plasmid part stocks were prepared by transforming into dcm- E. coli (NEB) and midiprepping DNA (Qiagen) after overnight growth of a single colony in LB in the presence of either ampicillin or spectinomycin.

A DNA mixture sufficient for twenty 10 uL Golden Gate assembly reactions with 40 fmol of each plasmid part per reaction, with BSA/Albumin (NEB) added to a 1.4x concentration. 7 uL of Golden Gate assembly master mix was added to 1 uL of BsaI enzyme, 1 uL of T4 Ligase, and 1 uL of T4 Ligase buffer and mixed by pipette. Reactions were then incubated at 37 °C for 10 min and then cycled at 37 °C for 2 min 30 s, 4 °C for 30 s and 16 °C for 5.5 min 45 times. After cycling, reactions were incubated for 10 min at 37 °C and then 10 min at 80 °C before storing at 4 °C.

Antibiotic selection was performed as described above by transforming into DH5alpha E. coli one shot cells and plating 1 uL of the 350 uL SOC mixture onto LB and ampicillin (100 ug / mL) after 1 hour in the shaker. The 1 uL aliquot was then mixed with 50 uL of SOC on the plate to allow even spreading of the mixture. Plates were inverted and incubated at 37 °C overnight and then colonies were counted. Finally, assembled constructs were verified by restriction digestion with enzymes HindIII and NotI, resulting in 4222 bp, 936 bp and 413 bp fragments for the expected assembly product.

6.1.9 Golden Gate-assembled aTc-inducible GFP circuit biological validation

DH5alpha-PRO competent cells expressing TetR were transformed with the Golden Gate-assembled plasmid using heat shock and plated on LB Agar (BD) plates containing 100 ug/mL ampicillin (IBI Scientific). Frozen stocks of bacteria expressing the plasmid were inoculated into 5 mL cultures of LB with 100 ug/mL of ampicillin and grown overnight with shaking at 37 °C. Cultures were then inoculated into 1 mL of LB containing ampicillin for each time point at a starting OD600 of 0.001. Anhydrotetracycline (aTc) was added to cultures at a concentration of 100 ng/mL and 1 mL aliquots of the cultures were taken at each time point and put on ice. OD600 measurements were taken with a NanoDrop 2000c spectrophotometer and GFP fluorescence was assayed acquired using the LSRFortessa cell analyzer and FACSDiva software.

6.1.10 Microfluidic device fabrication

Microfluidic devices were fabricated using multilayer soft lithography. Molds for the "flow" channel layer (for assembly biochemistry) and "control" channel layer (for microfluidic valving) were patterned by first spin coating two layers of AZ9260 photoresist (AZ Electronic Materials) on a 3" silicon wafer (2 spins for 60 s at 800 rpm), followed by a five minute bake on a hot plate at 90 °C. Photolithography of the coated wafers was performed using high-resolution transparency masks (CadArt Services, 20,000 dpi) on a contact aligner (Karl Suss MJ-6). The photoresist was exposed with the mask patterns (25 s x 4 exposures, 30 s rest between exposures) and developed in 1:2 Microchem 400K developer. After developing, the molds were baked at 120 °C to round the channel molds and silanized with tri(choromethyl)silane (Sigma Aldrich)

for ∼90 seconds in vapor phase to facilitate PDMS mold release and increase mold lifetimes. The microchannel heights on the mold (30 – 40 um) were subsequently measured using a profilometer (Dektak XTL, Bruker).

For the flow layer, 20:1 part A:B PDMS, either Sylgard 184 (Dow Corning) or RTV 615 (Momentum) was spin-coated (1000 rpm, 50 s) to a thickness of - 50 um while a 5 mm thick layer of 5:1 PDMS was poured on the control layer mold in a Petri dish, followed by an initial thermal cure (20 min, 80 °C for Sylgard, 45 min, 80 °C for RTV). After baking, both molds were removed from the oven for alignment. The negative PDMS channel replica was released from the control mold and cut to size (individual patterned devices) with a razor blade. A blunt-tipped 20G surgical steel luer stub (Becton Dickinson and Company, BD) was used to punch flow inlet and outlet ports in the PDMS, followed by an isopropyl alcohol wash to remove debris and drying under a nitrogen stream. The processed thick PDMS control layer was then aligned over the spin-coated flow layer under a dissecting scope. To bond the aligned flow and control layers, the composite PDMS substrate was cured (30 min at 80 °C for Sylgard, 45 min at 80 °C for RTV). The composite devices were then cut from the flow mold and the flow layer inlet/outlet holes were punched as previously described. The assembled PDMS layers were subsequently bonded to 0.5 mm thick glass coverslips (VWR) using oxygen plasma exposure (30 s/ 300W, Gasonics), and then baked at 80 °C overnight.

6.1.11 Microfluidic controller

The PDMS device was controlled by a custom-fabricated microfluidic controller. The controller contains four eight-channel solenoid valve arrays for pneumatic latching of the microfluidic chip valves, as well as four eight channel banks of constant pressure

outlets to provide driving pressure for fluids in the microfluidic flow channels. An external lowpressure (0-50 psig) air compressor (Panther) provides constant pneumatic pressure to the solenoid valve arrays and constant pressure outlets, with a secondary low-pressure regulator providing fine pressure control (0-10 psig) to the constant pressure outlets. Programmable operation of solenoids that actuate the microfluidic control valves, which function as latches and pumps, is achieved using digital output signals from an integrated Arduino Mega board with an Atmel ATmega2560 microprocessor. Complete assembly information including circuit board layouts and full design schematics are available in Metafluidics.

6.1.12 Microfluidic genetic circuit assembly

Automated microfluidic genetic circuit assembly reactions were conducted in the ring-mixer device. The ring-mixer features 5 inlets and outlets for loading and collecting reagents and 8 independently addressable control channels. A single assembly reaction was conducted per device; devices were not re-used.

The full assembly protocol consisted of 7 stages, comprised of the four device operations (stage number, device operation): (1, Incubate) Initially, all valves are latched at 15 psi to prime the control channels with 1x TE buffer; (2, Flow) A blocking agent, either 1x Bovine Serum Albumin (BSA) or 0.1% n-Dodecyl beta-Dmaltoside (DDM), is flushed through the ring mixer to mitigate non-specific protein adsorption. Blocking agents were loaded at 3 psi. (3, Incubate) All valves are latched and the blocking agent is incubated in the ring mixer for 10 min; (4, Flow) The blocking agent is purged from the ring mixer with compressed air at 3 psi; (5, Fill) The three compartments of the ring mixer are filled with the desired reagents. In this step compressed air at 7 psi backs the reagent loaded into each compartment inlet and

pushes out existing air in the channel through the bulk PDMS until all air bubbles are removed. This step typically takes 10-20 min. (6, Mix) Once reagents are loaded and air bubbles are removed, the three mixing valves are actuated at 25 Hz for 1000 cycles (4 min total) to achieve complete mixing. (7, Incubate) All valves are latched and the device, which is placed on top of an Eppendorf Mastercycler thermocycler with in situ adapter, is then temperature cycled according to the appropriate assembly biochemistry protocol described above. Finally, (8, Flow) the sample is collected from the ring mixer outlet in a several uL plug of 1x TE or distilled, deionized H2O for storage or transformation. Typically two devices (and thus two assembly reactions) were run in parallel; the microfluidic controller is capable of controlling up to 4 devices simultaneously.

The assembly protocol was similar for each biochemistry, differing only in thermocycling conditions, blocking agent (BSA for all reactions except DDM for Gateway) and the reagents loaded in each of the three ring-mixer compartments. For ligation reactions, "insert" DNA was loaded in compartment 1, "backbone" DNA was loaded in compartment 2, and ligase and ligase buffer in equal proportion were loaded in compartment 3. For Gateway reactions, BP Clonase was loaded in compartment 1, while DNA master mix was loaded in compartments 2 and 3. For Gibson reactions, DNA master mix (containing all 4 linear fragments) was loaded in compartments 1 and 2 (comprising half the ring mixer) while 2X Gibson enzyme mix was loaded in compartment 3 (comprising the other half). Finally, for Golden Gate reactions, BsaI, ligase, and ligase buffer were loaded in compartment 1, and DNA master mix (containing all 5 plasmids) was loaded in compartments 2 and 3. Negative control experiments were also conducted for each biochemistry where enzymes and associated buffers were replaced with distilled, deionized water.

Circuits assembled in the microfluidic device were transformed into E. coli cells

and processed as described in the protocols for larger-volume tube assembly, with the exception that the full volume eluted from the device mixing chamber was transformed. Restriction digestion and biological validation were conducted with the same protocols.

All genetic circuit assembly reactions were performed in devices made of Sylgard PDMS with the exception of Gateway reactions, which were performed in devices composed of RTV PDMS. Microfluidic devices held either 300, 400, 560, 575, 600, or 650 nL mixer volumes, depending upon which mold they were cast from.

6.2 Methods for miRNA activity measurements and multi-input predictions

6.2.1 Construction of miRNA sensor library

A common low sensor backbone (LSB) plasmid was constructed using a Gateway-Gibson strategy[53] to enable assembly of miRNA low sensor libraries in a single Golden Gate[55] step. Long DNA oligonucleotides encoding miRNA target site repeats were synthesized as ultramers by IDT, annealed and restriction-ligated via BbsI-mediated Golden Gate. White/blue screening enabled selection of mostly correct clones which were further verified by Sanger sequencing. Overhangs for Golden Gate steps are listed in Table 2.4.

6.2.2 Construction of miRNA classifiers

Several new backbones were constructed to expedite assembly of miRNA classifiers. JG107 and JG108 plasmids were assembled similarly to LSB and contain a LacZ

cassette into which miRNA target sites were cloned using BbsI-mediated Golden Gate. An important feature for these plasmids is that their miRNA target sites possess distinct overhangs when digested with BsaI. The low sensor backbone reverse (LSBr) plasmid retains the same design of LSB but with BsaI and BbsI recognition sites reversed. The LSBr-5and3UTR construct introduces a further addition of a BsaI/BbsI-flanked mCherry selection cassette in the 5' UTR of the mKate2 transcription unit. miRNA target sites present in LSB, JG108, and JG107 were then used as inputs to clone 3-input classifiers into LSBr using BsaI-mediated Golden Gate. 4-input classfiers were cloned into LSBr-5and3UTR using a similar strategy, with target sites for the 3' UTR originating from LSB and JG108 plasmids while JG107 encoded an inert sequence. Target sites were cloned into the 5' UTR from annealed oligonucleotides flanked by BsaI sites. White colonies were selected and verified by Sanger sequencing

6.2.3 Cell culture

HEK293FT were purchased from Thermo Fisher, HeLa and HepG2 cells were obtained from ATCC. HEK293FT and HeLa cells were grown in DMEM with 4.5 g/L glucose, L-glutamine, and sodium pyruvate (Cellgro) supplemented with 10% characterized FBS (HyClone). HepG2 cells were grown in DMEM with high glucose, 2 mM L-glutamine, without sodium pyruvate (HyClone) supplemented with 10% characterized FBS (HyClone). All cell lines were grown at 37° C and 5% CO_2. All cell lines tested negative for mycoplasma.

6.2.4 Transfection of miRNA sensors and classifiers into cell lines

miRNA low sensor libraries were transfected into cells using a reverse transfection method[62]. Example volumes are provided for a 100 uL mix which was sufficient for roughly 25 wells in 96-well plate format. 2% or 0.2% gelatin (Sigma-Aldrich) was diluted to 0.05% in sterile water. 1 volume of 2M sucrose was added to 9 volumes of buffer EC from an Effectene transfection kit (Qiagen) to make a master mix with total volumes depending on how much DNA was transfected. 37.5 uL Buffer EC mixture was added to 2 ug of DNA sample for a final DNA concentration of 20 ng/uL after addition of all components (DNA, buffer EC, enhancer, Effectene, gelatin). 3.75 uL of enhancer solution was added to the DNA/EC mixture, mixed gently, and incubated at room temperature for 5 min. 12.5 uL of Effectene was added and mixed followed by incubation for 10 min at room temperature. 53.75 uL of 0.05% gelatin was added and mixed gently to obtain the final transfection mix. 4 uL of the mixture was added to each well in a 96-well plate, wells were covered with sterile adhesive aluminum covers, and stored at -80C without dessication. Cells were reverse transfected by dissociating and counting cells as normal, followed by seeding of cells at usual concentration for forward transfection in 100 uL of cell media.

miRNA classifiers were transfected using forward transfection methods with lipofectamine 3000 (Thermo Fisher) according to the manufacturer's protocols. All transfections were conducted with 400 ng of DNA, 1 uL of P3000 reagent, 1 uL of lipofectamine 3000 in Opti-MEM (Thermo Fisher). HEK293FT cells were seeded in 24-well plates at various densities in 0.5 mL volume of media (HEK293FT = 2×10^5 cells/well, HeLa = 1×10^5 cells/well, HepG2 = 1.5×10^5 cells/well). For cell mixtures, 5×10^4 of each cell type was seeded per well and a 1:3 ratio of DMEM

high glucose:DMEM was used. miR-21-5p mimic was obtained from Sigma-Aldrich and cotransfected using the same method as for DNA. Target site combinations for 4-input sensors are listed in Table 2.2.

6.2.5 Flow cytometry

Flow cytometry was conducted using an LSR Fortessa cytometer with 405, 488, and 561 nm lasers (BD Biosciences). Reverse transfected cells in 96-well plate format were analyzed using the high throughput sampler option to collect >10,000 cells as quickly possible. Forward transfected cells in 24-well plate format were analyzed in tube format with >50,000 events collected per sample. mKate2 was detected using a 561 nm laser with 610/20 filter and 235 PMT voltage, EBFP2 was detected using a 405 nm laser with 450/50 nm filter and 240 PMT voltage, and EYFP was detected using a 488 nm laser with 530/30 nm filter and 200 PMT voltage. For multi-input circuits,

6.2.6 Model-based fitting and prediction of miRNA activities

Models and predictions were implemented in MATLAB with steady-state assumptions. Briefly, cytometry files for single-input sensors were read in, gated using forward and side scatter, binned by EBFP2 fluorescence, and parameters fit to the model using lsqcurvefit. Parameter fits were used to generate predictions either by multiplication of fold repression (synergistic-only model) or taking the minimization of mKate2 expression (antagonistic-only model) or a combination of the two (Ant/Syn model). Rate constants used are listed in Table 2.3.

6.2.7 Statistical analysis

Statistics are based on a log-normal distribution in fluorescence with a geometric standard deviation of approximately 2.8, an assumption supported by all samples in the flow cytometry data we collected. Greater than 50,000 events were chosen to allow at least 100 data points per fluorescent bin, and binned data are presented as medians. All error bars indicate standard deviations from three technical replicates. For ROC curves, error bars were calculated using threshold averaging and again indicate standard deviations from technical triplicates.

6.2.8 Data availability

All data files supporting this study are available on request from the corresponding author. Plasmids will be available from Addgene.

6.3 Methods for one-pot optimization

Detailed experimental methods and guidelines for designing poly-transfection experiments may be found on Protocol Exchange under protocol #6667 "One-pot Optimization of Genetic Circuits using Poly-transfections" and a quickstart guide is provided in Table 4.1.

6.3.1 DNA assembly framework

The plasmids used in this study were assembled by a hierarchical Golden Gate-Gibson assembly method similar to previous hierarchical methods used in our lab.[53] First, Golden Gate assembly was used to assemble plasmids for each transcription unit (termed plasmid level 1 or pL1 plasmids) from a number of basic input parts (termed plasmid level 0 or pL0) including insulators (pL0-I), promoters (pL0-P), 5' UTRs (pL0-5), gene coding sequences (pL0-G), 3' UTRs (pL0-3), and transcriptional terminators/poly-A sequences (pL0-T). Then plasmids containing multiple transcription units (termed pL2) were constructed from the assembled pL1 plasmids using Gibson assembly. For the Gibson step, plasmids were linearized using I-SceI endonuclease to generate overlapping regions, followed by assembly with Gibson Assembly Ultra Kit from SGI and electroporation into electro-competent cells. Typical Golden Gate and Gibson reaction protocols are given in Protocol Exchange. As a simplified alternative to hierarchical assembly, we have constructed plasmids that utilize a single Golden Gate step to obtain plasmids encoding two transcription units - one expressing the circuit part and the other expressing a fluorescent protein marker (Fig. 4-19). Since the backbone vector already encodes fluorescent protein expression, Gibson assembly would no longer be required. DNA sequences and plasmids comprising the simplified DNA assembly framework are available to researchers via

Addgene (#109150-109154).

6.3.2 Cell culture

The HEK293FT cell line was obtained from Thermo Fisher and HeLa cells were obtained from ATCC. Both cell lines were maintained in DMEM containing 4.5 g/L glucose, L-glutamine, and sodium pyruvate (Cellgro) which was supplemented with 10% FBS (Cellgro) and were grown at 37C and 5% CO_2. All lines used in this study tested negative for mycoplasma.

6.3.3 Transfections

Unless indiacted otherwise, transfections were conducted using Lipofectamine 3000 (Thermo Fisher) and were performed according to the manufacturer's dilution protocols with modifications for poly-transfections. Specifically we used 100-300 ng of DNA per circuit component with 1 uL of P3000 reagent and 1 uL of Lipofectamine 3000, all prepared in 100 uL Opti-MEM per 24-well plate. In the 24-well plate format, 200,000 HEK293FT cells or 150,000 HeLa cells were transfected per well. For poly-transfections each DNA-lipid mixture was prepared separately before being applied to cells, while for co-transfections DNA was mixed at the specified ratios and added to the transfection reagent. The amount of DNA and transfection reagent per component was scaled down according to the total number of circuit components in order to maintain similar total amounts compared to traditional transfections. For a quick reference for setting up poly-transfections, see Table 4.1. Reliability of poly-transfections across different transfection reagents and protocols was assessed by repeating poly-transfections using the same amount of DNA and the corresponding amount of transfection reagent recommended by the manufacturer (Fig. 4-20,

4-21).

6.3.4 Flow cytometry

Flow cytometery was performed using a BD LSRFortessa equipped with a 355nm laser with 379/28nm filter at 410 PMT voltage for measuring fluorescence from Sirius, 405nm laser with 450/50nm filter at 200 V for measuring TagBFP, 488 laser with 530/30 filter at 180 V for measuring mNeonGreen, 561nm laser with 582/15nm filter at 210 V for measuring mKO2, and 640 laser with 710/50nm filter at 320 V for measuring iRFP720. For co-transfections >50,000 cells were collected per sample, while for poly-transfections >200,000 cells were collected for two-component experiments. For greater numbers of circuit components than two, roughly 10 times more cells were collected per component.

To determine an orthogonal set of fluorescent proteins, we separately transfected HEK293FT cells with 22 constitutively expressed fluorescent proteins and measured signal from all available channels on the flow cytometer (Fig. 4-23). Fluorescence score was calculated as log10(99th percentile channel fluorescence from sample)-log10(channel background from untransfected). We then determined that Sirius, TagBFP, mNeonGreen, mKO2, and iRFP720 formed a relatively orthogonal set.

6.3.5 Data analysis

We have developed a MATLAB pipeline to facilitate analysis of multi-dimensional poly-transfection data. Code may be found on Github at `https://github.com/Weiss-Lab/MATLAB_Flow_Analysis`. Generally, data was binned across fluorescence corresponding to relevant input parts and statistics (e.g. median) for the output calculated for each bin. In some cases, data was 'sliced' to only analyze data with a

certain level of transfection marker prior to binning. Example analyses and associated code from this study may be found within the Github repository.

For subsampling of poly-transfection data, a series of points sampling a ratiometric trajectory was determined and the distance between each cell and the trajectory determined. To subsample, a number of cells were selected for inclusion based on the distance between the cell and the trajectory according to a log-normal distribution. For our analysis, each length between points on the trajectory corresponded to selection of approximately 310, 160, 100, 50, and 20 cells at a distances of 0.1, 0.2, 0.3, 0.4, 0.5 (logicle transformed units). Also, uniform subsampling may be used to normalize poly-transfection data to reduce the effect of higher density of singly- and doubly-transfected cells (Fig. 4-24).

To visualize the effects of each genetic component on system output, data were binned according to fluorescent of each transfection marker and plotted as medians in 2D, 3D, and 4D heatmaps to (Fig. 4-9-4-13, 4-25-4-26, 4-27). We then compared performance of 3-component and 4-component high sensors by using an ROC-like analysis, where each point within the ROC curve corresponds to the sensitivity and specificity calculated within a fluorescent bin (Fig. 4-28).

We used random forest regression to generate design rules for the miR-21-5p classifier from the poly-transfection data (Fig. 4-29). Briefly, we generated random forests with different minimum leaf sizes to maintain low out of bag error while reducing the number of branches in order to reduce the chance for overfitting. We then determined parameter importances and generated partial dependence plots using a minimum leaf size of 100.

6.3.6 Statistical analysis

Statistical analysis for this study is based on log-normal distributions for production of fluorescent proteins in cells. Data collected in this study supported the assumption of geometric standard deviation of approximately 2.8. For co-transfections, >50,000 cells were collected such that at least 100 data points could be analyzed per transfection bin. For poly-transfections, 10 times more cells were collected per circuit component to minimize sample collection time while still allowing at least 99% and 83% coverage of parameter space in 4- and 5-color experiments respectively.

References

53. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Research* **41** (2013).

55. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS ONE* **4** (2009).

62. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110 (2001).

Bibliography

1. Brown, B. D. *et al.* Endogenous microRNA Can Be Broadly Exploited To Regulate Transgene Expression According To Tissue, Lineage and Differentiation State. *Nature Biotechnology* **25,** 1457–1467 (2007).

2. Lin, S. L., Miller, J. D. & Ying, S. Y. Intronic microRNA (miRNA). *Journal of Biomedicine and Biotechnology* **2006,** 1–13 (2006).

3. Kozomara, A., Hunt, S., Ninova, M., Griffiths-Jones, S. & Ronshaugen, M. Target repression induced by endogenous microRNAs: large differences, small effects. *PloS one* **9,** e104286 (2014).

4. Slack, F. J. *MicroRNAs in Development and Cancer* (2011).

5. Talebizadeh, Z. Regulation of Gene Expression by Small RNAs. *The American Journal of Human Genetics* **86,** 328–330 (2010).

6. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508,** 66–71 (2014).

7. Grimson, A. *et al.* MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell* **27,** 91–105 (2007).

8. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Molecular Cell* **64,** 320–333 (2016).

9. Landgraf, P. *et al.* A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129,** 1401–1414 (2007).

10. Sokilde, R. *et al.* Global microRNA analysis of the NCI-60 cancer cell panel. *Molecular Cancer Therapeutics* **10,** 375–384 (2011).

11. Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods* **9,** 840–846 (2012).

12. Kozomara, A. & Griffiths-Jones, S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42,** 68–73 (2014).

13. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460,** 479–486 (2009).

14. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153,** 654–665 (2013).

15. Olsen, P. H. & Ambros, V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology* **216,** 671–680 (1999).

16. Bagga, S. *et al.* Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122,** 553–563 (2005).

17. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769–773 (2005).

18. Cloonan, N. Re-thinking miRNA-mRNA interactions: Intertwining issues confound target discovery. *BioEssays* **37,** 379–388 (2015).

19. Hendrickson, D. G. *et al.* Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biology* **7,** 25–29 (2009).

20. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).

21. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science* **336,** 233–237 (2012).

22. Eichhorn, S. W. *et al.* MRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular Cell* **56,** 104–115 (2014).

23. Clancy, J. L. *et al.* mRNA isoform diversity can obscure detection of miRNA-mediated control of translation. *RNA* **17,** 1025–1031 (2011).

24. Chen, C. Y. A., Zheng, D., Xia, Z. & Shyu, A. B. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nature Structural and Molecular Biology* **16,** 1160–1166 (2009).

25. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics* **16,** 421–433 (2015).

26. Guo, Z. W. *et al.* MtiBase: A database for decoding microRNA target sites located within CDS and 5'UTR regions from CLIP-Seq and expression profile datasets. *Database* **2015,** 1–9 (2015).

27. Lee, I. *et al.* New class of microRNA targets containing simultaneous 5' -UTR and 3'-UTR interaction sites. *Genome Research* **19,** 1175–1183 (2009).

28. Fang, Z. & Rajewsky, N. The impact of miRNA target sites in coding sequences and in 3' UTRs. *PLoS ONE* **6,** 1–6 (2011).

29. Lytle, J. R., Yario, T. a. & Steitz, J. a. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 9667–9672 (2007).

30. Pillai, R. S. *et al.* Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells. *Science* **309,** 1573–1576 (2005).

31. Humphreys, D. T., Westman, B. J., Martin, D. I. K. & Preiss, T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 16961–6 (2005).

32. Petersen, C. P., Bordeleau, M. E., Pelletier, J. & Sharp, P. A. Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell* **21,** 533–542 (2006).

33. Aeschimann, F. *et al.* LIN41 Post-transcriptionally Silences mRNAs by Two Distinct and Position-Dependent Mechanisms. *Molecular Cell* **65,** 476–489.e4 (2017).

34. Gurtan, A. M. & Sharp, P. A. The role of miRNAs in regulating gene expression networks. *Journal of Molecular Biology* **425,** 3582–3600 (2013).

35. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333,** 1307–1311 (2011).

36. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).

37. Leidinger, P. *et al.* A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biology* **14,** R78 (2013).

38. Ikeda, S. *et al.* Altered microRNA expression in human heart disease. *Physiological Genomics* **31,** 367–373 (2007).

39. Miki, K. *et al.* Efficient Detection and Purification of Cell Populations Using Synthetic MicroRNA Switches. *Cell Stem Cell* **16,** 699–711 (2015).

40. Sayeg, M. K. *et al.* Rationally Designed MicroRNA-Based Genetic Classifiers Target Specific Neurons in the Brain. *ACS Synthetic Biology* **4,** 788–795 (2015).

41. Bloom, R. J., Winkler, S. M. & Smolke, C. D. A quantitative framework for the forward design of synthetic miRNA circuits. *Nature Methods* **11,** 1147–53 (2014).

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

43. Tian, W. *et al.* High-Throughput Functional MicroRNAs Profiling by Recombinant AAV-Based MicroRNA Sensor Arrays. *PLoS ONE* **7,** 1–10 (2012).

44. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature genetics* **43,** 854–9 (2011).

45. Mukherji, S. & van Oudenaarden, A. Synthetic biology: Understanding biological design from synthetic circuits. *Nature Reviews Genetics* **10,** 859–871 (2009).

46. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Molecular Systems Biology* **3** (2007).

47. Kinkhabwala, A. & Guet, C. C. Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS ONE* **3** (2008).

48. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & Van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **31,** 69–73 (2002).

49. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351,** 720–724 (2016).

50. Antebi, Y. E. *et al.* Combinatorial Signal Perception in the BMP Pathway. *Cell* **170,** 1184–1196.e24 (2017).

51. Li, P. *et al.* Morphogen Gradient Reconstitution Reveals Hedgehog Pathway Design Principles. *Science* **360,** 543–548 (2018).

52. Yuan, Y. *et al.* Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 3158–63 (2015).

53. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Research* **41** (2013).

54. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Research* **42,** 13440–13451 (2014).

55. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS ONE* **4** (2009).

56. Kok, S. D. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synthetic Biology* **3,** 97–106 (2014).

57. Liang, J., Liu, Z., Low, X. Z., Ang, E. L. & Zhao, H. Twin-primer non-enzymatic DNA assembly: An efficient and accurate multi-part DNA assembly method. *Nucleic Acids Research* **45,** e94 (2017).

58. Halleran, A. D., Swaminathan, A. & Murray, R. M. Single day construction of multi-gene circuits with 3G assembly, 1–9 (2018).

59. Felgner, J. H. *et al.* Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *The Journal of Biological Chemistry* **269,** 2550–2561 (1994).

60. James, J. *et al.* A Novel Mechanism Is Involved in Cationic Lipid-Mediated Functional siRNA Delivery. *Molecular Pharmaceutics* **6,** 763–771 (2009).

61. Awais, M., Voronina, S. G. & Sutton, R. An efficient method is required to transfect non-dividing cells with genetically encoded optical probes for molecular imaging. *Analytical Sciences* **31,** 293–8 (2015).

62. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110 (2001).

63. McCoy, A. M., Collins, M. L. & Ugozzoli, L. A. Using the Gene Pulser MX-cell Electroporation System to Transfect Primary Cells with High Efficiency. *Journal of Visualized Experiments* **35,** 1–7 (2010).

64. Golden, R. J. *et al.* An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* **542,** 197–202 (2017).

65. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4,** 1–38 (2015).

66. Grey, F. *et al.* A viral microRNA down-regulates multiple cell cycle genes through mRNA 5'UTRs. *PLoS Pathogens* **6** (2010).

67. Rouya, C. *et al.* Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. *RNA* **20,** 1398–1409 (2014).

68. Hon, L. S. & Zhang, Z. The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. *Genome Biology* **8,** 1–18 (2007).

69. Wu, S. *et al.* Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* **29,** 2302–2308 (2010).

70. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).

71. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).

72. Mintz, P. J. *et al.* MicroRNA-181a* Targets Nanog in a Subpopulation of CD34(+) Cells Isolated From Peripheral Blood. *Mol Ther Nucleic Acids* **1,** e34 (2012).

73. Wang, W.-C. *et al.* miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics* **10,** 328 (2009).

74. Chou, T.-C. & Talalay, P. Generalized Equations for the Analysis of Inhibitions of Michaelis-Menten and Higher-Order Kinetic Systems with Two or More Mutually Exclusive and Nonexclusive Inhibitors. *European Journal of Biochemistry* **115,** 207–216 (1981).

75. Broderick, J., Salomon, W., Ryder, S., Aronin, N. & Zamore, P. Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *Rna* **17,** 1858–1869 (2011).

76. Elkayam, E. *et al.* Multivalent Recruitment of Human Argonaute by GW182. *Molecular Cell* **67,** 646–658.e3 (2017).

77. Denzler, R. *et al.* Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Molecular Cell* **64,** 565–579 (2016).

78. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature reviews. Genetics* **12,** 99–110 (2011).

79. Fabian, M. R. & Sonenberg, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature Structural and Molecular Biology* **19,** 586–593 (2012).

80. Lam, J. K., Chow, M. Y., Zhang, Y. & Leung, S. W. siRNA versus miRNA as therapeutics for gene silencing. *Molecular Therapy - Nucleic Acids* **4** (2015).

81. Xuan, M., Xiaofeng, C., Beixin, M. & Xuemei, C. Trip to ER: MicroRNA-mediated translational repression in plants. *RNA Biology* **10,** 1586–1592 (2013).

82. Gu, S. & Kay, M. A. How do miRNAs mediate translational repression? *Silence* **1,** 11 (2010).

83. Valencia-Sanchez, M. A., Liu, J., Hannon, G. J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes and Development* **20,** 515–524 (2006).

84. Zhou, H. & Rigoutsos, I. MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *RNA (New York, N.Y.)* **20,** 1431–9 (2014).

85. Siciliano, V. *et al.* MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nature Communications* **4,** 1–7 (2013).

86. Moore, W. A. & Parks, D. R. Update for the logicle data scale including operational code implementations. *Cytometry Part A* **81 A,** 273–277 (2012).

87. Zhu, X. P. *et al.* Microarray analysis of the aberrant microRNA expression pattern in gliomas of different grades. *Oncology Reports* **34,** 318–324 (2015).

88. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).

89. Darzacq, X. *et al.* In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14,** 796–806 (2007).

90. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).

91. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology* **4,** 1707–1719 (2006).

92. Dar, R. D. *et al.* Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 17454–9 (2012).

93. Halter, M., Tona, A., Bhadriraju, K., Plant, A. L. & Elliott, J. T. Automated live cell imaging of green fluorescent protein degradation in individual fibroblasts. *Cytometry Part A* **71,** 827–834 (2007).

94. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature structural and molecular biology* **11,** 599–606 (2004).

95. Bosson, A. D., Zamudio, J. R. & Sharp, P. A. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Molecular Cell* **56,** 347–359 (2014).

96. Weil, T. T., Parton, R. M. & Davis, I. Making the message clear: Visualizing mRNA localization. *Trends in Cell Biology* **20,** 380–390 (2010).

97. Sulc, P., Ouldridge, T. E., Romano, F., Doye, J. P. & Louis, A. A. Modelling toehold-mediated RNA strand displacement. *Biophysical Journal* **108,** 1238–1247 (2015).

98. Siciliano, V. *et al.* Engineering modular intracellular protein sensor-actuator devices. *Nature Communications* **9,** 1881 (2018).

99. Seluanov, A., Mao, Z. & Gorbunova, V. Analysis of DNA Double-strand Break (DSB) Repair in Mammalian Cells. *Journal of Visualized Experiments,* 1–7 (2010).

100. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology* **33,** 543–548 (2015).

101. Gaidukov, L. *et al.* A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Research* **46,** 1–15 (2018).

102. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12,** 326–8 (2015).

103. Song, J. *et al.* RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nature Communications* **7,** 1–7 (2016).

104. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for Metabolic Engineering. *Frontiers in Bioengineering and Biotechnology* **3,** 1–11 (2015).

105. Shih, S. C. C. & Moraes, C. Next generation tools to accelerate the synthetic biology process. *Integrative Biology* **8,** 585–588 (2016).

106. Gorochowski, T. E. *et al.* Genetic circuit characterization and debugging using RNA-seq. *Molecular Systems Biology* **13,** 952 (2017).

107. Lu, Y. Cell-free synthetic biology: Engineering in an open world. *Synthetic and Systems Biotechnology* **2,** 23–27 (2017).

108. Ranji, A., Wu, J. C., Bundy, B. C. & Jewett, M. C. in *Synthetic Biology Tools and Applications* 277–301 (Elsevier Inc., 2013).

109. Slusarczyk, A. L., Lin, A. & Weiss, R. Foundations for the design and implementation of synthetic genetic circuits. *Nature Reviews Genetics* **13,** 406–420 (2012).

110. Zhou, H., Vonk, B., Roubos, J. A., Bovenberg, R. A. & Voigt, C. A. Algorithmic co-optimization of genetic constructs and growth conditions: Application to 6-ACA, a potential nylon-6 precursor. *Nucleic Acids Research* **43,** 10560–10570 (2015).

111. Ghodasara, A. & Voigt, C. A. Balancing gene expression without library construction via a reusable sRNA pool. *Nucleic Acids Research* **45,** 8116–8127 (2017).

112. Woodruff, L. B. A. *et al.* Registry in a tube : multiplexed pools of retrievable parts for genetic design space exploration. *Nucleic Acids Research* **45,** 1553–1565 (2017).

113. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Research* **41,** 10668–10678 (2013).

114. Gach, P. C., Iwai, K., Kim, P., Hillson, N. & Singh, A. K. Droplet Microfluidics for Synthetic Biology. *Lab Chip* **17,** 3388–3400 (2017).

115. Woodruff, K. & Maerkl, S. J. A High-Throughput Microfluidic Platform for Mammalian Cell Transfection and Culturing. *Scientific Reports* **6,** 1–12 (2016).

116. Klein, J. C. *et al.* Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Research* **44,** e43 (2015).

117. Kim, H. *et al.* 'Shotgun DNA synthesis' for the high-throughput construction of large DNA molecules. *Nucleic Acids Research* **40** (2012).

118. Wong, A. S., Choi, G. C., Cheng, A. A., Purcell, O. & Lu, T. K. Massively parallel high-order combinatorial genetics in human cells. *Nature Biotechnology* **33,** 952–961 (2015).

119. Beal, J. Bridging the Gap: A Roadmap to Breaking the Biological Design Barrier. *Frontiers in Bioengineering and Biotechnology* **2,** 1–16 (2015).

120. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352,** aac7341 (2016).

121. Yu, J. *et al.* Human Induced Pluripotent Stemm Cells Free of Vector and Trangene Sequence. *Science* **324,** 797–801 (2009).

122. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

123. Papapetrou, E. P. *et al.* Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proceedings of the National Academy of Sciences* **106,** 12759–12764 (2009).

124. Nagamatsu, G. *et al.* Optimal ratio of transcription factors for somatic cell reprogramming. *Journal of Biological Chemistry* **287,** 36273–36282 (2012).

125. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–824 (2013).

126. Fujii, W., Kawasaki, K., Sugiura, K. & Naito, K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Research* **41,** e187 (2013).

127. Hori, Y., Kantak, C., Murray, R. M. & Abate, A. R. Cell-free extract based optimization of biomolecular circuits with droplet microfluidics. *Lab on a Chip* **17,** 3037–3042 (2017).

128. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 14024–14029 (2013).

129. Li, Y. *et al.* Modular construction of mammalian gene circuits using TALE transcriptional repressors. *Nature Chemical Biology* **11,** 201–213 (2015).

130. Wong, A. S. L. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proceedings of the National Academy of Sciences* **113,** 2544–2549 (2016).

131. Schwake, G. *et al.* Predictive modeling of non-viral gene transfer. *Biotechnology and Bioengineering* **105,** 805–813 (2010).

132. Bhise, N. S., Shmueli, R. B., Gonzalez, J. & Green, J. J. A novel assay for quantifying the number of plasmids encapsulated by polymer nanoparticles. *Small* **8** (2012).

133. Davidsohn, N. *et al.* Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition. *ACS Synthetic Biology* **4,** 673–681 (2015).

134. Bullock, M. D. *et al.* Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression. *Cell Death and Disease* **4,** 1–10 (2013).

135. Ferreira, J. P., Overton, K. W. & Wang, C. L. Tuning gene expression with synthetic upstream open reading frames. *Proceedings of the National Academy of Sciences* **110,** 11284–11289 (2013).

136. Ferreira, J. P., Peacock, R. W. S., Lawhorn, I. E. B. & Wang, C. L. Modulating ectopic gene expression levels by using retroviral vectors equipped with synthetic promoters. *Systems and Synthetic Biology* **5,** 131–138 (2011).

137. Wroblewska, L. *et al.* Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nature Biotechnology* **33,** 839–841 (2015).

138. Hawley, T. S., Hawley, R. G. & Telford, W. G. Fluorescent proteins for flow cytometry. *Current Protocols in Cytometry* **2017,** 9.12.1–9.12.20 (2017).

139. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. *GAL4-VP16 is an unusually potent transcriptional activator* 1988.

140. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268,** 1766–1769 (1995).

141. Kelleher, R. J., Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61,** 1209–1215 (1990).

142. Wang, T., Xie, Y., Tan, A., Li, S. & Xie, Z. Construction and characterization of synthetic microRNA cluster for multiplex RNA interference in mammalian cells. *ACS Synthetic Biology* **565,** acssynbio.5b00180 (2015).

143. Saito, H. *et al.* Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nature Chemical Biology* **6,** 71–78 (2010).

144. Pinzon, N. *et al.* MicroRNA target prediction programs predict many false positives. *Genome Research* **27,** 234–245 (2017).

145. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6** (2011).

146. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–5 (2009).

147. Kong, D. S. *et al.* Open-source, community-driven microfluidics with Metafluidics. *Nature Biotechnology* **35,** 523–529 (2017).

148. Shih, S. C. C. *et al.* A Versatile Microfluidic Device for Automating Synthetic Biology. *ACS Synthetic Biology* **4,** 1151–1164 (2015).

149. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).

150. Ellis, B. L. *et al.* A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus

(AAV1-9) and one engineered adeno-associated virus serotype. *Virology Journal* **10,** 1 (2013).