

NORTHWESTERN UNIVERSITY

Enabling cell-based therapies through environmental sensing and signal processing

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences

By

Patrick Sean Donahue

EVANSTON, ILLINOIS

December 2020

© Copyright by Patrick Donahue 2020

All Rights Reserved

Abstract

Cell-based therapies are an exciting frontier in medicine. This field is built on a simple premise—cells can be engineered to recognize and treat various human diseases. The paradigm of cell-based therapy uses biosensors to interrogate a cell's environment and distinguish disease from health, intracellular signaling pathways and genetic circuitry to process, integrate, and interpret this information, and effector functions to enact a therapeutic response against the disease. Best exemplified at present by chimeric antigen receptor modified T cells, which are programmed to patrol the body and to seek out and destroy tumor cells, cellular therapeutics hold promise for treating cancer and many other pathologies. Though several cell-based therapies have gained FDA approval in recent years for clinical use against hematologic malignancies, the reach of cell-based therapies is limited by many factors, including the availability of fundamental technologies that could enable us to target the cells against a broad range of diseases. This thesis aims to address this problem through two overarching efforts: (I) developing fundamental technologies for cell-based biosensors and therapeutics and (II) translating these cell-based devices for clinical applications.

Towards the first aim, I first refined technologies for sensing hallmarks of the tumor microenvironment and discovered that employing different transmembrane domains in a synthetic receptor system could mitigate ligand-independent signaling. This advance will enhance the specificity, and thereby safety, of cell-based therapies that rely on synthetic, transmembrane receptors to sense their environments. I also investigated biosensors for detecting hypoxia, a feature common to many cancers and other pathologies, elucidated several principles for engineering these sensors, and designed genetic circuits to modulate their signaling. These circuits may ultimately make these biosensors more robust and resultingly expand the range of applications for this technology to many disease indications. Finally, I developed a toolkit for the engineering of genetic programs in mammalian cells. Through a thorough investigation and characterization of synthetic promoters and transcription factors, I established principles for tuning this system and enabled the design of a mathematical model that predicts how these components and circuits function. This toolkit, termed the Composable Mammalian Elements of Transcription (COMET), has broad applications for composing genetic circuits that form the signaling pathways in cell-based therapies,

including those that convert the signaling from biosensors into an effector function.

Towards the second aim, I developed a natural killer cell-based strategy that can be deployed against a broad range of solid tumors. Microenvironment induced natural killer cells (MINK) recognize hallmarks of the tumor microenvironment (TME), such as hypoxia, rather than tumor cell-surface antigens, and respond with a therapeutic effector function, such as producing a cytokine to stimulate an immune response against the tumor. By relying on features that arise from tumor physiology and are thus common across tumors rather than on tumor-specific antigens, MINK and other tumor TME-recognizing therapies may find wide utility. As evidenced by MINK, the technologies developed in this thesis enable the engineering of cell-based therapies against cancer; these synthetic biology technologies will ultimately enable the development of cell-based therapies against a broad range of diseases.

Acknowledgements

First and foremost, I would like to thank my amazing mentor Josh Leonard for shaping me into the scientist I am today, giving me so many opportunities, and providing constant encouragement. The past several years have been nothing short of incredible. I have learned so much from working with you, from every meeting in your office, discussion at lab meeting, and round of revision on our grants and papers. I look forward to our future relationship and collaborations.

This work would not be what it is without the support and guidance of my colleagues in the Leonard Lab. Michelle Hung, Kelly Schwarz, and Andrew Younger were always available to answer my never-ending stream of questions. Taylor Dolberg, Joe Muldoon, and Devin Stranford were with me nearly every single day of this journey. So many of my ideas and experiments were bounced off of, and improved by, them. Thank you for always being there to talk, science or otherwise. Alex Prybutok was a fantastic resource for all things CAR related, critical analysis of experiments and conclusions, and positive encouragement. Hailey Edelstein worked side-by-side with me on the COMET and MESA manuscripts and so many other things, and together we planned and carried out many fantastic projects and experiments. Kate Dray and I collaborated to develop next generation COMET-TFs, models, software, and applications. It has also been great to work alongside and learn from the newer laboratory members (Jon Boucher, Will Corcoran, Simrita Deol, Taylor Gunnels, Katie Dreyer, and Roxi Mitrut) who will take many of the projects described here in exciting new directions. I learned so much and was constantly inspired by my labmates, who made my graduate school experience complete and every day enjoyable. During this time, I was fortunate to mentor a number of talented undergraduates, through the academic years and summer sessions. Pear Dhantravan, Imran Khan, Joe Draut, Everett Allchin, Benji Leibowitz, Lauren Battaglia, Brandon Lim, Katie Zhu, Marya Ornléas, Liz Schauer, and Kate Chambers all put in a tremendous amount of work. Their efforts and reagents underly many of the experiments presented here.

I owe a huge debt of gratitude to my family and friends who supported me through my education at Northwestern and through the many years leading up to it. I am also appreciative of the many mentors and teachers I have had over the years who inspired my love for science, encouraged me to pursue the MD/PhD pathway, believed I could, and taught me the foundational knowledge and skills I needed to be successful.

List of Abbreviations*Medicine-related*

ALL	Acute lymphoblastic leukemia
BCG	Bacille Calmette-Guérin
CMV	Cytomegalovirus
DLBCL	Diffuse large B-cell lymphoma
FDA	Food and Drug Administration
GVHD	Graft versus host disease
MCL	Mantle cell lymphoma
r/r	Relapsed/refractory
SV40	Simian vacuolating virus 40
TME	Tumor microenvironment

Cell types

CAF	Cancer-associate fibroblast
hMSC	Human mesenchymal stem cell
MDSC	Marrow-derived suppressor cell
NK	Natural killer
RBCs	Red blood cell
TAM	Tumor-associate macrophage
TAN	Tumor associated neutrophil
Treg	Regulatory T cell

Cellular components and processes

ADCC	Antibody-dependent cell-mediated cytotoxicity
AKT	Protein kinase B
CAR	Chimeric antigen receptor

CD80/CD86	Cluster of differentiation 80/cluster of differentiation 86
CTLA-4	Cytotoxic T-lymphocyte-associated protein 4
EGF	Epidermal growth factors
ERK	Extracellular signal-regulated kinases
FGF	Fibroblast growth factors
GM-CSF	Granulocyte macrophage colony-stimulating factor
GPCR	G protein-coupled receptor
HIF	Hypoxia inducible factor
HRE	Hypoxia response element
IFN	Interferon
IGF	Insulin like growth factors
IL	Interleukin
IRES	Internal ribosome entry site IRES
JAK	Janus kinase
MAPK	Mitogen-activated protein kinases/
MHCs	Major histocompatibility complex
NFAT	Nuclear factor of activated T-cells
NFAT	Nuclear factor of activated T-cells
NHEJ	Non-homologous end joining
PD-1	Programmed death-1
PD-L1	Programmed death ligand 1
PI3K	Phosphoinositide 3-kinase
PLCG	Phospholipase C
RANTES	Regulated upon activation, normal T Cell expressed and presumably secreted
RNAPII	RNA polymerase II
RTK	Receptor tyrosine kinase
STAT	Signal transducer and activator of transcription

TCR	T cell receptor
TGF- β	Transforming growth factor- β ,

Synthetic biology technologies

CARs	Chimeric antigen receptor
COMET	COMposable mammalian elements of transcription
GEARs	Generalized Engineered Activation Regulators
GEMS	Generalized extracellular molecule sensor
haNK	High affinity NK cell
HBS	Hypoxia biosensor
LP	Landing pad
MESA	Modular extracellular sensor architecture
MINK	Microenvironment induced natural killer cells
synNotch	Synthetic Notch
TALEN	Transcription activator-like effector nuclease
taNK	Target-activated NK cell

COMET-related terminology

AD	Activation domain
NES	Nuclear export signal
NLS	Nuclear localization signal
RaZFa	Rapamycin-activated ZFa
TF	Transcription factor
ZF	Zinc finger
ZFa	ZF activator
ZFi	ZF inhibitor

MESA features and domains

TC	Target chain
PC	Protease chain
ECD	Extracellular domain (comprises ligand binding domain and extracellular linker)
ICD	Intracellular domain (comprises protease or transcription factor and inner linker)
TMD	Transmembrane domain
TEVp	Tobacco etch virus protease
NTEVp	N-terminal protein component of split TEVp
CTEVp	C-terminal protein component split TEVp
PRS	Protease recognition sequence (for TEVp)
AIP	Autoinhibitory peptide (for TEVp)
IL	Inner linker
PCIL	Protease chain inner linker

Transmembrane domains

CD28	Cluster of differentiation 28
EphA4	Ephrin type-A receptor 4
FGFR	Fibroblast growth factor receptor (1–4)
FGFR-Syn	Fibroblast growth factor receptor synthetic (a consensus sequence of 1–4)
GpA	Glycophorin A
VEGFR1	Vascular endothelial growth factor 1 receptor

Ligand binding domains

FKBP	FK506-Binding protein	binder of rapamycin/rapalog
FRB	FKBP rapamycin binding	binder of rapamycin/rapalog
GID1	Gibberellin insensitive dwarf 1	binder of gibberellin
GAI	Gibberellin insensitive	binder of gibberellin

ABI1	Abscisic acid-insensitive1	binder of abscisic acid
PYL1	Pyrabactin like protein 1	binder of abscisic acid
GBP	GFP-binding protein	binder of GFP, nanobody
scFv	Single chain variable fragment	a type of binding protein
G6-311	scFv against VEGF	binder of VEGF, previously termed "V1"
B20-4.1	scFv against VEGF	binder of VEGF, previously termed "V2"

Ligands

EtOH	Ethanol
DMSO	Dimethyl sulfoxide
Rapalog, Rapa	Rapamycin analog, here specifically Takara AP21967
GA3	Gibberellin
GA3-AM	Gibberellin analog, cell-permeable
ABA	Abscisic acid
GFP	Green fluorescent protein
sGFP	Secreted green fluorescent protein
VEGF	Vascular endothelial growth factor
sVEGF	secreted VEGF

Fluorescent proteins and other reporter proteins

EBFP2	Enhanced blue fluorescent protein 2
EYFP	Enhanced yellow fluorescent protein
mC	mCerulean
miRFP670	monomeric infrared fluorescent protein (with 670 nm emission)
miRFP720	monomeric infrared fluorescent protein (with 720 nm emission)
mV	mVenus
NanoLuc	NanoLuciferase

Reagents

BSA	Bovine serum albumin
CTFR	Cell Trace Far Red
CTV	Cell Trace Violet
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DTZ	Diphenylterizine
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IgG	Immunoglobulin G
IMDM	Iscove Modified Dulbecco Media
IV	Integration vectors
LB	Luria-Bertani
MEM α	Minimal Essential Media with alpha modification
PBS	Phosphate-buffered saline
PVDF	Polyvinylidene fluoride
RCP	Rainbow calibration particles
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RNase	Ribonuclease that degrades RNA
SDS	Sodium dodecyl sulfate
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween
TE	Tris EDTA buffer
TUPV	Transcription unit positioning vector
URCP	Ultra-rainbow calibration particles

Other terms

ANOVA	Analysis of variance
APC	Allophycocyanin
AU	Arbitrary units
BCA	Bicinchoninic acid assay
BRET	Bioluminescence resonance energy transfer
COVID-19	Coronavirus disease 2019
ECL	Enhanced chemiluminescence
FACS	Fluorescence activated cell sorting
FRET	Förster resonance energy transfer
FSC-A	Area in the Forward Scatter channel
FSC-H	Height in the Forward Scatter channel
HSD	Tukey's honest significance test
HyD	Hybrid detector
IBIS	Interdisciplinary Biological Sciences Program
MEFLs	Molecules of Equivalent Fluorescein
MEPB	Molecules of Equivalent Pacific Blue
MEPTRs	Molecules of Equivalent PE-Texas Red
MFI	Mean fluorescence intensity (in arbitrary units)
MOI	Multiplicity of infection
NA	Numerical aperture
NEB	New England Biolabs
NFRET	Normalized FRET signal
PDB	Protein Data Bank
PE	Phycoerythrin
RLU	Relative luciferase units
S.E.M.	Standard error of the mean

SNR	Signal-to-noise ratio
SSC-A	Area in the Side Scatter channel
WT	Wild type

Table of Contents

Abstract	3
Acknowledgements	5
List of Abbreviations	6
Table of Contents	14
List of Tables and Figures	17
Chapter 1. Introduction and background	22
1.1 The hallmarks of cancer underly our therapeutic successes	23
1.2 Moving beyond chemotherapy: targeting other hallmarks of cancer with immunotherapy	24
1.3 Checkpoint inhibitors demonstrate the wide impact of immunotherapy	25
1.4 CAR T cells drive immunotherapy success in the 21st century	26
1.5 Challenges in CAR T cell treatments highlight challenges facing immunotherapy	27
1.6 CAR T cells are a preview of the future of cancer therapeutics	29
1.7 Targeting signatures of the tumor microenvironment with synthetic biology: an opportunity for broad impact	29
1.8 Synthetic biology tools for engineering next-generation anticancer therapies: what do we have and what do we need?	33
1.9 Conclusions	42
Chapter 2. COMET: a toolkit for engineering custom genetic programs in mammalian cells	43
2.1 Preface	44
2.2 Abstract	44
2.3 Introduction	45
2.4 Materials and methods	46
2.4 Results	61
2.4 Discussion	85
2.6 Data and code availability	89
2.7 Acknowledgements	89

2.8 Competing Interests	90
Chapter 3. Elucidation and refinement of synthetic receptor mechanisms	91
3.1 Preface	92
3.2 Abstract	93
3.3 Introduction	93
3.4 Materials and methods	95
3.5 Results	107
3.6 Discussion	126
3.7 Data availability	129
3.8 Acknowledgements	129
Chapter 4. Developing robust cell-based biosensors for hypoxia	131
4.1 Preface	132
4.2 Abstract	132
4.3 Introduction	133
4.4 Materials and methods	135
4.5 Results	143
4.6 Discussion	162
4.7 Acknowledgements	165
Chapter 5. Tumor microenvironment induced natural killer (MINK) cells for cancer immunotherapy	166
5.1 Preface	167
5.2 Abstract	167
5.3 Introduction	168
5.4 Materials and methods	169
5.5 Results	175
5.6 Discussion	192
5.7 Acknowledgements	194

Chapter 6. Conclusions and future directions	195
6.1 Project impact and future directions	196
6.2 Recent progress and challenges in mammalian synthetic biology	203
6.3 Next generation therapeutics for diseases beyond cancer	206
References	207
Appendix 1. Disposition of reagents	225
Appendix 2. Supplementary information for Chapter 2	227
Appendix 3. Supplementary information for Chapter 3	256
Appendix 4. Supplementary information for Chapter 4	277
Appendix 5. Supplementary information for Chapter 5	279

List of Tables and Figures

Chapter 1.

Figure 1.1	CAR T cells illustrate the cell-based therapy paradigm	30
Figure 1.2	Native versus synthetic components for cell-based therapies	35

Chapter 2.

Figure 2.1	Investigation of COMET promoter design rules	62
Figure 2.2	A model for COMET-mediated gene regulation	65
Figure 2.3	Characterizing an expanded panel of ZFa	67
Figure 2.4	Tuning transcription through ZF mutants and AD variants	70
Figure 2.5	Transcriptional inhibition	73
Figure 2.6	Characterization of promoter design rules in the genome	75
Figure 2.7	Engineering small molecule-responsive TFs	78
Figure 2.8	Strategies for design of CID systems	80
Figure 2.9	Evaluation of new CID systems for COMET	82
Figure 2.10	Composing Boolean logic	84

Chapter 3.

Table 3.1	Flow cytometry setup for FRET analysis	102
Figure 3.1	Protease chain tuning to improve MESA receptor performance	109
Figure 3.2	TMD contributions to MESA receptor signaling	113
Figure 3.3	Development of a flow cytometric FRET approach to probe receptor chain association	115
Figure 3.4	Effect of TMD choice on receptor chain association	118
Figure 3.5	Effects of TMD dimerization geometry on receptor signaling	121
Figure 3.6	Tuning an expanded panel of MESA receptor systems	123
Figure 3.7	Generalizing principles for receptor engineering	125

Chapter 4.

Figure 4.1	Oxygen diffusion through various volumes of media	144
Figure 4.2	Oxidation rates of fluorescent proteins	146
Figure 4.3	Evaluating oxidation conditions for fluorescent protein maturation	147
Figure 4.4	Evaluating the performance of a genomically integrated HBS	149
Figure 4.5	Evaluating the choice of minimal promoter for genomically integrated HBS lines	151
Figure 4.6	Evaluation of genomically integrated HBS under physoxic conditions	153
Figure 4.7	HBS activation with stable HIF1 α	155
Figure 4.8	Evaluation of HBS feedback with HIFs in a B16F10-LP line	157
Figure 4.9	Possible issues with initial feedback circuits and potential solutions	158
Figure 4.10	Genetic circuits for HBS modification with COMET	160
Figure 4.11	Conferring oxygen sensitivity to COMET TFs	161
Figure 4.12	Hypoxia-sensing genetic circuits with COMET instead of HIFs	163

Chapter 5.

Figure 5.1	Development and validation of methodologies for NK-92 cell laboratory work	177
Figure 5.2	Evaluating infrared and BRET reporter systems	179
Figure 5.3	Imaging BRET with the IVIS spectrum and under biologic tissue	181
Figure 5.4	Comparing strategies for conferring exogenous IL-2 independence to the NK-92 cell line	183
Figure 5.5	Characterization of an IL-2 independent NK-92 line with luciferase expression	186
Figure 5.6	Investigating HBS transduced NK-92 cell lines	188
Figure 5.7	Engineering NK-92 lines with HBS with other minimal promoters	189
Figure 5.8	Comparison of performance of NK-92 HBS lines with various minimal promoters by flow cytometry	191

Appendix 2.

Figure A2.1	Effects of the choice of ZF and the number and spacing of ZF binding sites	228
Figure A2.2	Differences between spaced and compact promoters	229
Figure A2.3	A model for ZFa-regulated gene expression	231
Figure A2.4	Characterization of the panel of ZFa	233
Figure A2.5	Characterization of the panel of ZFa	234
Figure A2.6	Properties of ZFa with ZF mutants and AD variants	235
Figure A2.7	Investigation of ZFi-mediated and ZFi-DsRed-mediated inhibition	237
Figure A2.8	Further investigation of ZFi-mediated and ZFi-DsRed-mediated inhibition and the COMET mechanism	238
Figure A2.9	Workflow for landing pad (LP) integration and stable cell line generation	240
Figure A2.10	Characterization of genomically integrated COMET TFs	242
Figure A2.11	Characterization and tuning of RaZFa activity	243
Figure A2.12	Characterization and tuning of RaZFa activity	244
Figure A2.13	Results of RaZFa tuning	246
Figure A2.14	Implementing Boolean logic with COMET	247
Figure A2.15	Flow cytometry gating	249
Figure A2.16	Profile of fluorescent calibration beads	250
Table A2.1	ZFa and fitted parameters for x6-C promoters	251
Table A2.2	Fitted parameters for modifications to promoter architecture and ZFa domains	252
Table A2.3	Instrument specifications for analytical flow cytometry.	253
Table A2.4	Instrument specifications for FACS	253
Table A2.5	Parameters used for standard models of activation	253
Table A2.6	Parameters for the removal of cooperativity by COMET inhibitors	254
Table A2.7	Parameters used for standard models of AND gates	254

Table A2.8	Promoter states for the COMET four-input gate	254
Note A2.1	Nomenclature used in this manuscript	255
Appendix 3.		
Figure A3.1	Comparisons across MESA assays	257
Figure A3.2.	Investigating the role of PCIL on PC stability	258
Figure A3.3	Normalization of protein expression for MESA containing various TMDs	259
Figure A3.4	Surface stain and background signal of MESA containing various TMDs	261
Figure A3.5	Evaluating TCs that showed low ligand-induced signaling with TMD-matched PC	262
Figure A3.6	Investigating ligand-induced accumulation of MESA receptors containing synthetic TMDs	263
Table A3.1	Ligand concentrations for small molecule-sensing MESA systems	264
Table A3.2	Rapamycin-sensing MESA expression normalization doses	265
Table A3.3	Instrument specifications for analytical flow cytometry to quantify reporter expression	266
Table A3.4.	Instrument specifications for analytical flow cytometry to quantify surface expression	266
Table A3.5	Instrument specifications for confocal microscopy to quantify FRET	266
Note A3.1	Outcomes from one-way ANOVAs and Tukey's HSD tests	267
Note A3.2	Outcomes from two-way ANOVAs and Tukey's HSD tests	268
Note A3.3	Outcomes from three-way ANOVAs and Tukey's HSD tests	273
Appendix 4.		
Table A4.1	Filters for microscopy	278
Table A4.2	Filters for flow cytometry analysis	278
Table A4.3	Filters for flow cytometry sorting	278

Appendix 5.

Table A5.1	Filters for flow cytometry sorting	280
Table A5.2	Filters for flow cytometry analysis	280
Figure A5.1	Comparison of performance of NK-92 HBS lines with various minimal promoters by Imaging on IVIS Spectrum	281
Figure A5.2	Imaging of HBS activation through biologic tissue	282

Chapter 1. Introduction and background

1.1 The hallmarks of cancer underly our therapeutic successes

In the United States, there will be approximately 1.8 million new cancer cases and 600,000 deaths from cancer in 2020⁵. While staggering, these numbers are lower than they would have been without the two standards of modern cancer care—early detection and innovative treatments. The cancer death rate began falling in 1991, and since then, these efforts have saved an estimated 2.9 million lives⁵. Surgery, chemotherapy, and radiation comprise the standard of care, and innovation in all three modalities has led to these continually increasing survival rates. These innovations are largely driven by our increased understanding of the fundamental biology of tumors and how these treatments target various aspects of the physiology^{6, 7}.

As Hanahan and Weinberg described in their landmark 2000 paper⁸, there are several hallmarks of cancer cell physiology: self-sufficiency in growth signals, insensitivity to growth-inhibitor signals, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. Together, these lead to rapid, malignant growth; it is this property that we have exploited for the successful treatment of many cancers with chemotherapy and radiation. The physiology which drives the disease is therefore its underlying weakness—the faster the cells grow, the more susceptible they are to these agents that inhibit cell growth by damaging DNA, inhibiting mitosis, and depriving the cells of nutrients, among other mechanisms. The first decades of the 21st century saw the years of fundamental research into cell and tumor cell biology that revealed the driving mutations in many cancers bear great fruit in the form of precision medicine⁹. For each patient, we could sequence the tumor and choose small molecule drugs targeted to the specific dysregulated pathways or the specific driving mutations of the tumor. As more patients were treated with targeted therapeutics, we learned more about cancer physiology from the patients who did and did not respond—many patients have mutations for which we don't yet have drugs, tumor heterogeneity means some cells may not have the mutations required for the drugs to be efficacious, tumors then select for these cells, and some of the mutations are simply associated with the tumor rather than driving its physiology and are therefore not actually therapeutic targets¹⁰.

As techniques for studying biology progressed further, so did our understanding of the cancer cell and its interaction with its host. This led to a revision of these hallmarks of cancer. Hanahan and Weinberg

described in 2011 the next-generation hallmarks of cancer, including two new enabling characteristics, genome instability and tumor-promoting inflammation, and two emerging hallmarks, deregulating cellular energetics and avoiding immune destruction¹¹. Interestingly two of these areas concern the interaction between the tumor and the immune system, relating to how cancers both avoid and exploit our bodies' mechanisms for dealing with invasive threats. Recently, therapeutics focused on addressing this interaction and restoring the proper balance of anti-tumor immunity have seen breakthrough clinical successes.

1.2 Moving beyond chemotherapy: targeting other hallmarks of cancer with immunotherapy

After several decades of intense laboratory research, immunotherapies have recently shown successes in clinical trials. However, early evidence of immunotherapy dates back to 1891, when tumors were injected with bacteria to drive a response against them^{12, 13}. Currently there are a broad range of immunotherapeutic strategies under investigation in the laboratory as well as in preclinical and clinical trials. The use of Bacille Calmette-Guérin (BCG) vaccine, a mycobacteria similar to tuberculosis, as a modern intravesical therapy for bladder cancer is proposed to work through an immunostimulatory mechanism and was first demonstrated effective in 1976^{14, 15}. Cytokine injection and infusion directly into tumors simulates the immune system¹⁶, with major players in this space including interferon (IFN)- α , interleukin (IL)-2¹⁷, IL-12¹⁸, IL-15^{19, 20}, and IL-21²¹. Oncolytic viruses can express immune potentiating factors such as granulocyte macrophage colony-stimulating factor (GM-CSF) to potentiate anti-tumor immune response²². Antibodies can block checkpoints that cancer cells exploit to inhibit the immune system²³. Dendritic cells²⁴ and T cells²⁵ can be cultured *ex vivo* to stimulate their response against tumor neoantigens and reinfused. Allogenic and autologous natural killer cells, with or without genetic enhancement, can be infused to harness the innate immune response²⁶. Further, all of these cells can be genetically engineered *ex vivo* to increase their anti-tumor efficacy through a number of gene therapies²⁷. Of all of these novel therapeutics, two classes in particular show exceptional clinical efficacy and both immediate and future promise: checkpoint inhibiting antibodies and chimeric antigen receptor (CAR) modified T cells.

1.3 Checkpoint inhibitors demonstrate the wide impact of immunotherapy

Checkpoint inhibitors are a class of antibody-based drugs that block immune checkpoint signaling²³. Normally, immune checkpoints exist to prevent autoimmunity and hypersensitivity reactions. However, cancer cells can take advantage of these pathways to prevent their destruction by the immune system. Checkpoint inhibitors block this immune avoidance mechanism, taking the brakes off the immune system and allowing it to attack many tumors. As expected, however, these potent therapies usually come along with a characteristic set of immune-related adverse events from the increased activity of the immune system that, when recognized early, can be appropriately managed²⁸. Despite these toxicities, checkpoint inhibitors have seen widespread clinical success.

Two checkpoints in particular currently have demonstrated efficacy when used as therapeutic targets: CTLA-4/CD80/CD86 (cytotoxic T-lymphocyte-associated protein 4/cluster of differentiation 80/cluster of differentiation 86) and PD-1/PD-L1 (programmed death-1/programmed death ligand 1). In its landmark Phase III clinical trial (MDX010-020), the anti-CTLA-4 antibody ipilimumab increased median survival among patients with stage III or IV melanoma and progressive, metastatic disease from 6.4 to 10.1 months, when added on top of gp100 peptide vaccine²⁹. In the clinical trial leading to its first approval (Checkmate-037), the anti-PD-1 antibody nivolumab led to an objective response rate of 31.7% of patients, versus the 10.6% objective response rate in the chemotherapy group, among patients with unresectable metastatic melanoma who had progressed after ipilimumab therapy³⁰. The anti-PD-L1 antibody (PD-L1 is the ligand for PD-1) avelumab showed a 31.8% objective response rate in a trial against stage IV Merkel cell carcinoma (JAVELIN)³¹, becoming the first FDA-approved treatment for this cancer.

While ipilimumab remains the only FDA-approved anti-CTLA-4 inhibitor, pembrolizumab³² and cemiplimab³³ have joined the anti-PD-1 class and durvalumab³⁴ and atezolizumab³⁵ have become approved to block PD-L1. Approved indications for checkpoint inhibitor therapy have since greatly expanded to multiple indications in metastatic melanoma, small-cell lung cancer, non-small-cell lung cancer, renal cell cancer, squamous cell cancer of the head and neck, urothelial cancer, colorectal cancer, hepatocellular carcinoma, Hodgkin's lymphoma, patients with certain biomarkers regardless of primary tumor, gastric adenocarcinoma, esophageal adenocarcinoma, mediastinal large B-cell lymphoma, Merkel

cell carcinoma, cervical cancer, squamous cell cancer, and triple-negative breast cancer³⁶. While initial indications were only for advanced disease, several of these immune checkpoint inhibitor antibodies have now earned first-line status for their safety and efficacy. Excitingly, combination immunotherapy with two of these agents, such as nivolumab or ipilimumab as a first line therapy in advanced melanoma, shows substantial benefit over the use of either agent alone, especially in patients with PD-L1 negative tumors³⁷. Further, more targets for cancer immunotherapy with antibodies have been identified and are active areas of investigation³⁶. This novel class of drugs highlights how targeting a hallmark that is common among many malignancies can rapidly have broad, very tangible clinical impacts.

1.4 CAR T cells drive immunotherapy success in the 21st century

The second new class of drugs that showed breakout success is engineered T cells^{38, 39}. These cells are typically harvested from the patient, manipulated in some fashion, and then reinfused. The first demonstration of *ex vivo* expanded and stimulated tumor infiltrating lymphocytes to treat malignancies was in 1988⁴⁰, and innovations in this field have led to complete response rates of up to 32%, durable for at least 3 years for some cancers⁴¹. Other strategies using T cells have been pursued in parallel to this work, both aiming to develop T cells that are reactive against a specific tumor antigen. In the first approach, an antigen-specific T cell receptor (TCR) is identified and then the gene for it is delivered to a population of T cells, which then become specific for that antigen^{42, 43}. There are nearly 100 clinical trials of TCR T cell therapies underway, though none are currently FDA-approved⁴⁴.

The second approach uses chimeric antigen receptors (CARs), which are synthetic mimics of the TCR^{45, 46}. A CAR comprises an antigen binding domain (typically a single chain variable fragment (scFv)) linked through a transmembrane domain to various intracellular signaling domains from proteins involved in the T cells' native immune synapse⁴⁷. CARs differ from TCRs in several ways. First, TCRs are restricted to recognition of antigens within major histocompatibility complexes (MHCs), which vary between patients and only display some antigens, while the CARs employ scFvs to bind any antigen on the surface of the target cell. Second, the engineered intracellular signaling domains provide both the first and second signal required for T cell activation. The end result is the same—when a CAR binds an antigen on a tumor cell, it,

like the TCR, induces the cytotoxic effector function of the CD8+ T cells, resulting in the death of the cancer cell.

CAR T cells first showed efficacy in a patient in 2010⁴⁸. The clinical trials of CAR T cells thus far have shown incredible successes⁴⁹, leading to three FDA approved drugs: tisagenlecleucel (KYMRIA[®]; Novartis), axicabtagene ciloleucel (YESCARTA; Kite), and brexucabtagene autoleucel (TECARTUS; Kite). The range of indications are currently fairly narrow and include B-cell acute lymphoblastic leukemia (ALL), relapsed or refractory or relapsed/refractory and diffuse large B-cell lymphoma (r/r DLBCL), and relapsed/refractory mantle cell lymphoma (r/r MCL). The numbers out of the clinical trials were impressive. Axicabtagene ciloleucel showed a 83% overall response rate and 58% of patients achieving complete response after 27.1 months (ZUMA-1)⁵⁰. Tisagenlecleucel induced complete responses in 57% of patients with DLBCL and follicular lymphoma (FL), with nearly 90% of the patients remaining in remission at 28.6 months⁵¹; it also achieved a 40% complete response rate in patients with DLBCL with 90% of these patients surviving several years later (JULIET)⁵². Tisagenlecleucel has also demonstrated efficacy in pediatric ALL, with a 1-year survival rate in relapsed/refractory ALL of 76%, CAR T cells still detectable for at least 20 months, and complete remissions lasting nearly a decade in some patients (ELIANA)^{53, 54}. Though not yet approved, lisocabtagene maraleucel (JCAR017; Juno), showed a 73% objective response and a 53% complete response (TRANSCEND)⁵⁵. Of note, all of these results were achieved with CARs directed against the same target, CD19, a B cell antigen. New CARs against other antigens in leukemias, lymphomas, and a multitude of other cancers are an active area of investigation. There are currently at least 300 CAR T cell clinical trials under way, with many against non-CD19 antigens and in non-B cell malignancies⁵⁶.

1.5 Challenges in CAR T cell treatments highlight challenges facing immunotherapy

While CAR T-cells will no doubt save many lives, this technology currently has some limitations that are being actively investigated^{38, 57}. The manufacturing and expansion process is complex and time-consuming, making these therapeutics cost hundreds of thousands of dollars per infusion⁵⁸. Additionally, patients who do not have enough viable T cells or cannot wait for the manufacturing time are not candidates

for therapy. To address this, several groups are investigating how to remove various antigens and functions from the T cells in order to generate allogenic “universal” T cells that could be given to any patient without triggering severe graft versus host disease or immune rejections of the therapy⁵⁹.

Another challenge is that the efficacy of the CAR molecule exists in a tight window—too weak binding and signaling and it is ineffective; too much signaling, especially in the absence of ligand, and the cell becomes anergic or exhausted⁶⁰. Solutions to this have included re-engineering components of the CAR and inserting the CAR gene into the locus that usually expresses a component of the TCR rather than random integration in the genome⁶¹.

One issue that has come to the forefront as CARs are increasingly targeted against solid tumors is that the immunosuppressive tumor microenvironment results in suppression of the T cell response. This is a major part of the reason that CAR T cells have seen great successes in hematologic malignancies but are much more limited in their treatment of solid tumors⁶². Solutions to this include either inducible or constitutive secretion of immune stimulating cytokines, such as IL-12, by the CAR T cell⁶³ and combining CAR T cells with checkpoint inhibitor antibodies or removing the checkpoint molecules altogether^{64, 65}. Alternatively cells, such as NKs, that might penetrate better into some tumors can be engineered with CARs to induce their own cytotoxic effector functions⁶⁶. I elaborate more on this solution in detail later in this chapter (**Section 1.8, Cell type**).

Finally, for many tumors, we do not have antigens unique to the tumor and, more importantly, tumor heterogeneity among patients means that many patients will not have the requisite antigen and thus not be a candidate for therapy. On a related note, tumors can downregulate the antigen and escape therapy. Solutions to these issues include searching for more antigens⁶⁷, deploying novel tumor cell to T cell adapters⁶⁸, designing CAR T cells to recognize combinations of antigens (this addresses both the lack of unique antigens resulting in off target effects and the escape issue)⁶⁹, and oncolytic viral vectors that also cause the surviving tumor cells to express an antigen for which an existing CAR T cell therapy can target⁷⁰.

1.6 CAR T cells are a preview of the future of cancer therapeutics

CAR T cells, while a potent, revolutionary therapy that will save many lives, are not a silver bullet and the be-all end-all cancer therapy, but an inspiration for a therapeutic paradigm—cells can be engineered to recognize and thus target various disease states in the body and respond by executing an effector function that is therapeutic (**Fig. 1.1**). This paradigm could be applied to many diseases beyond cancer, as well as many different cell types and effector functions beyond the cytotoxic T cell. The way to accomplish this is with synthetic biology—an interdisciplinary, engineering approach to biology, where we (re)-design, model, construct, and characterize, new biological components, such as proteins, genetic circuits, and cells.

Synthetic biology provides many opportunities for engineering a large variety of therapeutic effector functions across a wide array of cell types. These therapeutic engineered cells would generally follow the basic paradigm illustrated by CAR T cells: 1) sense a marker or a variety of markers that define a pathologic state, 2) process this signal into instructions for the cell, 3) execution of an effector function that ameliorates the disease state. These approaches intrinsically address some of the challenges that CAR T cells are currently facing and ultimately provide more flexibility in terms of steps that can be taken to overcome them.

1.7 Targeting signatures of the tumor microenvironment with synthetic biology: an opportunity for broad impact

There are a large variety of non-antigenic features of the tumor environment that could be sensed by cells, many of these either being or resulting from the quintessential hallmarks of cancer, as described above. One would therefore expect therapies that target these features to find applications across a wide variety of tumor types. In this sense, we can liken such novel cell-based therapeutics to checkpoint inhibitors, chemotherapy, and radiation, methodologies which all target hallmarks of cancer and therefore have many indications. In contrast, precision medicines, such as inhibitors of specific driving mutations (*i.e.*, imatinib and vemurafenib), and CAR T cells rely on specific derangements or antigens. These require multiple drugs for multiple cancers and, while ultimately highly successful, require years of development and trials for each indication. Rather, an approach in which cells can detect features of the tumor

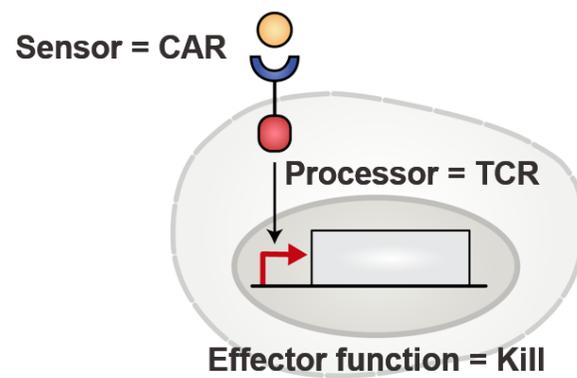


Fig 1.1 CAR T cells illustrate the cell-based therapy paradigm.

microenvironment, rather than antigens on a tumor cell, would ultimately have a broad impact after initial development, with clinical trials still required for each indication.

One target is associated with the sustained angiogenesis hallmark—vascular endothelial growth factor (VEGF). The tumor's rapid pace of growth causes a high demand for nutrients, leading to sustained angiogenesis, which, though it occurs through many different pathways, leads to increase in VEGF in the tumor microenvironment. VEGF plays a pivotal role in tumor angiogenesis and is a hallmark of malignancy⁷¹. VEGF overexpression is associated with tumor progression and poor prognosis in many cancers^{71, 72}. For instance, VEGF, from melanocytes and tumor infiltrating lymphocytes, is present in 91% of malignant melanomas but not in benign nevi or normal dermis^{73, 74, 75}. Plasma VEGF increases from 19-47 pg/mL in healthy humans to 46-104 pg/mL in patients with cancer⁷⁶. Tumor tissue is a significant VEGF reservoir, with concentrations 7-70 times greater than that in the plasma⁷⁶. While most of the VEGF in the human body is located intracellularly in skeletal muscle, the concentration of free VEGF in muscles is 1.6-4.8 pg/mL compared with 13.3-39.4 pg/mL in breast carcinomas, making it a good marker⁷⁶. In addition to VEGF, tumors also produce and rely on a variety of growth factors, such as transforming growth factor (TGF)- β , insulin like growth factors (IGFs), epidermal growth factors (EGFs), and fibroblast growth factors (FGFs), for their continued rapid proliferation⁷⁷. While all of these growth factors, including VEGF, are technically antigens, they are not the classic cell-associated antigens that CAR or TCR modified T cell therapies sense. Display of a growth factor in an MHC (as could be recognized by TCRs) does not alone differentiate healthy from cancerous cells, and these growth factors are paracrine hormones and thus generally not expressed on the surface of cells (as could be recognized by CARs). Therefore, new sensor strategies, as discussed below (**Section 1.8, Sensors**), will need to be used to sense this feature.

Despite the hyperangiogenesis resulting from VEGF overexpression, the vasculature in tumors is markedly abnormal, leading to leading to poor perfusion and resultant hypoxia⁷⁸. Unchecked, the growth of the tumor outpaces that of its vasculature and blood supply^{79, 80}. Tumor vasculature is markedly abnormal in appearance and function, classically described as elongated and tortuous, in contrast to well-organized physiologic vascular networks⁸¹. The tight barrier formed by endothelial and smooth muscle cells is frequently insufficient, resulting in unusually permeable vasculature, which affects the ability to maintain

adequate perfusion⁸². These effects lead to regions of hypoxia throughout the tumor. While healthy tissues have multiple mechanisms to counteract hypoxia, including altering cellular metabolism and rapidly increased blood flow through arterial dilatation, tumors lack the ability to respond likewise⁸³. This makes hypoxia a good marker for tumors in otherwise healthy humans, particularly those lacking ischemic disease. To sense tumor hypoxia, several groups have developed DNA-based sensors and tested them *in vivo*^{84, 85, 86}. Further it has been shown that one can use a similar hypoxia biosensor to restrict the expression of CARs in T cells to hypoxic conditions, potentially increasing the specificity of these therapies for a tumor environment⁸⁷. However, these sensors often show leaky gene expression or heterogeneous responses to hypoxia, which may limit their utility in a cell-based therapy.

Another marker of the tumor environment is elevated potassium in the interstitial fluid. Potassium homeostasis is tightly controlled by mammalian cells, as it is a key determinant of the membrane potential and thus control is essential to maintaining, among others, critical cardiac and neurological functions. This is done by keeping intracellular potassium levels high at 145 mM compared to extracellular potassium concentrations around 5 mM⁸⁸. As tumor necrosis and cell death occurs naturally during tumor progression, this sequestered potassium leaks out into the extracellular space, elevating it to approximately 40 mM in some tumors⁸⁹. It follows that if this previously intracellular potassium is residing in the extracellular space, there should also be many proteins and molecules normally present intracellularly within the tumor cells that are now present and concentrated in the extracellular tumor microenvironment, though this has not been comprehensively studied.

There are many other signatures of the tumor microenvironment that arise from dysregulated metabolism of tumor cells. For instance, the uptake of amino acids is deregulated, the demand for nitrogen is increased, glucose uptake occurs at a greatly elevated rate, and lactic acid is secreted⁹⁰. This latter property is associated with a marked acidosis in the tumor microenvironment⁹¹; even if the molecules responsible for the acidosis are not specific markers, the acidosis itself might be with the appropriate sensor. Many other metabolites and metabolic pathways are dysregulated within tumors⁹²; whether these elevations occur as detectable extracellular markers is not yet known.

The immunosuppressive tumor microenvironment is also another hallmark and has signature

features that could be detected by cells with targeted biosensors. Typically, immunosuppressive molecules such as IL-10 are upregulated. An analysis of breast cancer interstitial fluid revealed increased levels of IL-7, IL-10, IL-12 IL-13, RANTES, and other factors⁹³. Another analysis showed that patients could have one of several different characteristic patterns of cytokine expression⁹⁴. Though patients with cancer often have marked elevations of serum IL-6 levels⁹⁵, some studies have found IL-6 to be lower in tumors than the serum but noted that IL-8, VEGF, and GM-CSF are upregulated within the tumor⁹⁶. Further, tumors contain a variety of non-malignant cells that carry out immunosuppressive functions, including tumor-associated macrophages (TAMs), marrow-derived suppressor cells (MDSCs), tumor associated neutrophils (TANs), cancer-associated fibroblasts (CAFs), and regulatory T cells (Tregs)⁹⁷. A cell could sense the presence of these cells, mediated by their cell surface antigens, as a marker of an immunosuppressive tumor.

Altogether, these general features of tumors present exciting opportunities for targeting cell-based therapies against a broad variety of tumor types. However, achieving this goal will likely require the sophisticated engineering of mammalian cells for these advanced functions and of many technologies for doing so. As shown in **Fig 1.1**, each therapy will require at least three components: a sensor, a processor, and an effector function.

1.8 Synthetic biology tools for engineering next-generation anticancer therapies: what do we have and what do we need?

A choice at the outset

For a cell to recognize any of the molecules present within the tumor or any of the features of the tumor microenvironment, it must have a sensor. The sensor is generally either a co-opted native receptor or a synthetic biosensor developed against that specific marker. These two categories represent a major design choice in synthetic biology and cell engineering—to what degree will the parts that enable the engineered function come from nature (referred to as endogenous or native parts) or be synthetic (referred to as exogenous or orthogonal parts)? Though native versus synthetic is a binary description, the choice actually lies along a spectrum, with each component of each part being potentially native or synthetic (**Fig 1.2**).

However, even the most synthetic, orthogonal systems are only nominally orthogonal to the native function of the cell. They still rely on cellular components, such as transcription and translation machinery and the basic nucleotides and amino acids that go into the construction of these parts⁹⁸. This resource competition presents a major challenge for cell engineering⁹⁹. Addressing this challenge for synthetic parts is actively under investigation. One method is to use computational models that take this competition into account when designing circuits^{100, 101}. Other potential solutions include developing orthogonal translation machinery for mammalian cells and have the circuits rely on these^{102, 103}. While most of this work has been done in prokaryotes, control systems for managing resource burden in mammalian cells have been recently reported¹⁰⁴.

Both ends of the spectrum in **Fig 1.2** have advantages and disadvantages, thus the choice is a classic trade-off, one of many made when engineering mammalian cells with synthetic biology¹⁰⁵. By sticking to native components, the development process will be expedited, as these components have likely been tuned over years of evolution to robustly sense their target ligands and environmental states. However, this could ultimately make tuning the sensors to trigger ON and OFF at different levels of the input, as one would do to alter the sensitivity and specificity of the cell-based therapy, more difficult in the future. Native components are also more readily available, with many well-characterized receptors and transcription factors from a vast array of species available in the literature. However, using native components risks interference from other uses of those components by the cell. Native systems are subject to endogenous regulation, which may be counter to therapeutic goals, such as a downregulation of a cytotoxic effector program in a cell that resides in an immunosuppressive microenvironment.

A final challenge with native systems is that they may be difficult, if not impossible in some cases, to multiplex—to use several of them simultaneously in the same cell so that multiple environmental signals may be integrated into the cell's decision making process about whether it is in a pathological environment or not. This ability to engineer cells that can recognize and integrate multiple signals will be critical for developing highly specific, and therefore safe, cell-based therapies. This challenge arises from the fact that many of the co-opted native receptors share usage of readily engineerable pathways, such as NFAT. The NFAT promoter, and the upstream signaling pathways have been thoroughly characterized and as such

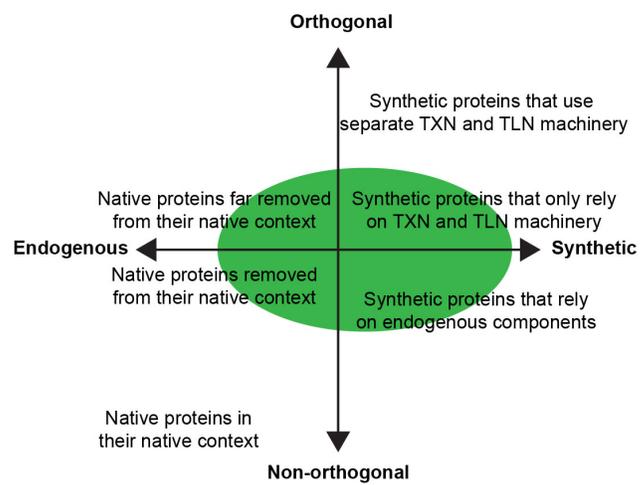


Fig 1.2 Native versus synthetic components for cell-based therapies. Green circle represents the region in which many mammalian synthetic biology technologies exist.

are well-understood. This allows the components of this pathway to be utilized to process the signals from synthetic receptors and convert them into effector functions, and many synthetic systems rely on NFAT to process their signals (see *Sensors* and *Processors*, below). Resultingly, two systems that both carry out their action through elevating intracellular calcium and/or ultimately activating an NFAT promoter, could only be used to perform OR logic in a cell, as either input would turn the system fully on.

On the other hand, synthetic systems are engineered from scratch or by utilizing components from a variety of systems and possibly assembling them in a manner that renders them orthogonal in their new context. While a significantly more time-consuming process than co-opting native systems, this has multiple advantages in the long term. Synthetic systems are often engineered to be modular, that is so that the inputs and outputs can be readily swapped—a property that results in long-term gains as each technology has a wide variety of applications open to it but represents a sophisticated engineering feat. The development of such systems also involves a substantial amount of characterization and thus accumulation of knowledge on how to tune the performance of the systems. Therefore, any adjustments required to get a certain component to work for a certain application can be done relying on this pre-existing knowledge base, rather than requiring a large investigation of how to make the desired changes to the system. Even though the development process is much longer than the time required to co-opt native components, such as native receptors as biosensors, it is likely that in the near future, computational protein design will increase the speed with which we build synthetic biosensors^{106, 107}. On the other hand, some disadvantages of synthetic components include that they may induce an immune response when used as part of a therapy, rendering the therapy ineffective; this will be a greater challenge for some therapies (those that patrol the body) than others (those that would be encapsulated in an immunoprivileged capsule that shields the cells from the immune system of the host¹⁰⁸). One strategy to address this issue is to engineer the synthetic protein to look similar to proteins native to the species through a process known as deimmunization¹⁰⁹, though this may not be an option for many synthetic components. Thus, while the repurposing of native components offers a rapid solution, maximizing the orthogonality of the components we use for sensing, signal processing, and carrying out effector functions circuits by developing synthetic solutions for each will have long term payoffs that may outweigh the expediency of utilizing endogenous components.

Sensors

The first step in programming a cell-based therapy is enabling it to recognize a feature of its environment and convert this to an intracellular signal—it does so with a sensor. Many such synthetic biology systems exist for sensing various environmental signals, each at different points along the orthogonality spectrum. As discussed above, CARs are synthetic analogues of TCRs, though they do not require presentation of the antigen in an MHC, and, like the TCR, they signal through the cell's endogenous NFAT pathway. SynNotch is an engineered Notch receptor that comprises a modular scFv antigen binding ectodomain and an intracellular synthetic transcription factor (TF). Upon antigen binding, the signal is mechanotransduced across the membrane, by rendering the intracellular domain of the receptor susceptible to cleavage by an endogenous protease, which releases the TF^{110, 111}. The generalized extracellular molecule sensor platform (GEMs) can sense a variety of protein and small molecule antigens, signaling through a menu of pathways, including JAK/STAT, MAPK, NFAT, and NF- κ B¹¹². The Tango¹¹³ and Cha-Cha¹¹⁴ sensors rewire signal from a g-protein-coupled receptor (GPCR) through a dCas9 transcription factor to an endogenous or exogenous gene. The modular extracellular sensor architecture (MESA) platform was specifically developed, by my lab, to sense extracellular antigens through dimerization of two synthetic chains and subsequent release of a synthetic transcription factor upon ligand binding¹¹⁵. Unlike the examples above, MESA does not rely on endogenous components to signal and its functional domains are derived from proteins that are not normally present in or do not interact with components normally present in mammalian cells. Further, while synNotch and CARs signal through mechanotransduction and are thus meant for detecting cellular-associated antigens, MESA can sense soluble antigens. Though it has been reported that CARs can be engineered to sense the soluble TGF- β ¹¹⁶, this finding has not yet been extended to other soluble ligands. MESA is therefore the only orthogonal, synthetic receptor system for the detection of soluble proteins.

First developed by our lab in 2014, MESA has since been used to engineer cells to sense the immunosuppressive, angiogenic VEGF and respond by producing IL-2, mediated through a released dCas9-TF targeted to the endogenous IL-2 locus². Recently, MESA have been multiplexed to sense multiple antigens in the same cell¹¹⁷. This latter study identified several challenges with engineering cell-

based therapies and synthetic receptors. Notably, the high ligand-independent signaling of MESA was one such limiting factor and is a challenge that I address in **Chapter 3** of this work.

Sensors, however, are not necessarily transmembrane receptors. For instance, some sensors detect the presence of intracellular proteases in order to sense HIV infection¹¹⁸ or cancerous changes¹¹⁹. Others can detect non-protease intracellular proteins, thereby sensing various viral infections and other pathologies¹²⁰. Recent advances in the computational design of transmembrane pores signal the advent of biosensors for extracellular ions and some small molecules¹²¹. Many RNA-based circuits exist that can discriminate between healthy and cancer cells or the inflammatory state of the cell's environment^{122, 123}. Additionally, DNA-based sensors can sense certain cell states by determining the presence or absence of transcription factors, such as biosensors for inflammation that sense NK-kB¹²⁴ or hypoxia biosensors (HBS) that detect hypoxia inducible factor (HIF) proteins^{84, 85, 86}. I focus on this latter class in **Chapter 4**.

Processors

The second step a cell performs as a therapeutic is to convert the signal from a sensor into an effector function—it does so through a processor. Signal processing systems comprise transcription factors (TFs) and other intracellular proteins and small molecules involved in signaling pathways. The available processors also span the spectrum from native to synthetic. As noted above, CAR T cell signal processing relies solely on the native NFAT signaling pathway, while synNotch, Cha Cha, and MESA use synthetic TFs. For these systems, as well as many other applications, the default synthetic TFs are tTA^{125, 126} and Gal4¹²⁷. Derived from prokaryotic, yeast, and viral proteins, these two synthetic TFs bind to and activate transcription from an exogenous piece of DNA that may or may not be integrated into the genome. Several synthetic processing systems have recently been developed for mammalian cells and are generally built on other imported prokaryotic TFs¹²⁸, Gal4¹²⁹, transcription activator-like effector nucleases (TALENs)^{130, 131, 132}, zinc fingers (ZFs)^{133, 134}, dCas9¹³⁵, and transposases¹³⁶.

Several notable strategies are evident in the middle of the spectrum, between relying entirely on native signaling pathways and bringing in a synthetic signaling system. In the first, a biosensor activates an endogenous signaling pathway, which activates an exogenous copy of its target promoter, driving

transcription of an exogenous transgene. The use of this strategy is illustrated by GEMs and a cell-based therapy for Type I diabetes—the sensing of extracellular glucose is wired through the activation of a native ATP-sensitive potassium channel, a native voltage-gated calcium channel, and subsequent calcineurin-dependent activation of an exogenous NFAT promoter that expresses insulin¹³⁷. In the second strategy, a synthetic TF (typically dCas9-based) released from a synthetic receptor activates gene expression from an endogenous locus. This strategy has been used by MESA, Cha Cha, and dCas9-synR, a receptor system similar to MESA¹³⁸. The Generalized Engineered Activation Regulators (GEARs) system employs another strategy—it taps into the native signaling pathway of a native receptor and reroutes the signal to activate an endogenous gene that is not normally the target of this pathway¹³⁹.

In addition to being less subject to interference from endogenous regulation, signal processing systems comprising mostly synthetic transcription components have many advantageous properties. Ideally these systems are well-characterized, have predictable tuning of various knobs, and can process multiple signals orthogonally to each other. These systems, once assembled and characterized, are powerful tools. One such system assembled in bacteria enabled the construction of highly sophisticated genetic circuits; when combined with a mathematical model and software package, named CELLO, this synthetic TF system enabled the predictive design of high sophisticated genetic circuits in bacteria¹⁴⁰. This allows circuits to be designed and evaluated computationally, in a massively high-throughput fashion, saving substantial amounts of time performing experimental evaluations in the wet lab. Developing a TF system that would enable predictive design of signal processing genetic circuits for mammalian cells is a feat I work towards in **Chapter 2** of this work.

Effector functions

The possible range of effector functions for a cell-based therapy is large, encompassing broad categories such as killing, metabolizing, and signaling, but here I will focus on one application in particular—driving immune response against a solid tumor. Many patients (20-40%)¹⁴¹ do not respond to immunotherapies, and we are only beginning to figure out why. Key reasons elucidated thus far include the immune status of the tumor at the initiation of therapy, including factors such as the presence or absence

of infiltrative T cells¹⁴². The tumor microenvironment is highly immunosuppressive, and this hinders the response to immune therapies such as CAR T cells and checkpoint inhibitors^{143, 144}. Turning these “cold” immunosuppressive tumors into “hot” immunostimulatory tumors could open the door for these treatments to a larger number of patients with a wider array of disease indications.

One method to do this is to administer cytokine therapies. However, doing so systemically can result in severe inflammation and has led to patient deaths^{145, 146, 147}. Intratumoral injection of these cytokines provides clinical benefit, but not all locations are accessible to injection or practical to inject, especially in metastatic disease, and the injection itself is not without risk^{148, 149, 150}. Rather, a strategy in which a cell therapy could patrol the body, sense when it is in a tumor, and only then produce these incredibly potent therapies directly within the tumor would have high levels of safety and efficacy. Any cell within the tumor, namely one from the patient’s own immune system, would experience very high levels of the cytokine produced by the therapeutic cell, while levels of this cytokine in the rest of the body would not be as high. Contrast this with a strategy in which cytokines are systemically infused or constitutively produced by infused cells and the benefit of this added layer of complexity becomes clear—by producing cytokines locally, much higher effective doses can be given without the immune related adverse events that would occur if the patient systemically experienced these high doses.

Producing cytokines locally in a tumor with a cell-based therapy has several other advantages. First, it allows for continuous, repeated dosing from these “cell factories,” reducing the number of necessary injections and complexity of the therapy. Second, it allows for mixing and matching of custom therapeutic programs to meet the specific needs of each patient or to address each presentation of a disease—as long as the processor in the therapeutic is modular, the effector molecule can be swapped at will. Third, this could also speed the adaptation of newer, more potent engineered cytokine mimetics. These new molecules, which were designed through various experimental and computational approaches to have increased potency and/or safety profiles when compared to their native analogues, include eIL-12¹⁵¹, eIL-15(RLI)^{19, 152}, Neo-IL-2/15¹⁵³, or DR-IL18¹⁵⁴.

Cell types

While the majority of FDA approved cell-based therapies are T cell-based, due to the success of the CAR T cell field, several other cell types are under active investigation for their therapeutic potential. For instance, hMSCs are an exciting therapeutic frontier as they traffic to tumors and can be engineered with anti-tumor therapeutic functions¹⁵⁵. Many cells of the innate immune system hold promise for treating cancer as well. Many tumors have a Th2 phenotype, best thought of as a low grade, smoldering inflammation that inhibits the Th1 phenotype in which cytotoxic T cells thrive^{143, 144}; this phenotype recruits cells of the innate immune system, including macrophages, which compose up to 50% of the tumor mass in some cancers¹⁵⁶. Natural killer (NK) cells are also being investigated in clinical trials for several cancers^{157, 158}, as these cells also traffic into tumors after systemic infusion^{159, 160}. Though safe^{161, 162, 163}, adoptively transferred NKs show minimal clinical benefit without further engineered effector functions, such as CAR-induced cytotoxicity. The availability of off-the-shelf, allogenic natural killer cell lines that can be infused into many patients without triggering GVHD or immune rejection, such as the NK-92 cell line, has also led to the use of NKs for off-the-shelf therapies, addressing some of the concerns with the cost and manufacturing time and complexity inherent with the current generation of CAR T cells^{164, 165}. In **Chapter 5** of this work, I investigate strategies for translating our synthetic biology work to therapeutic programs in NK-92 cells.

Considering the combination of cell type and effector function

Returning to the spectrum of synthetic, orthogonal components vs. native, endogenous ones, we can apply this framework to the choice of cell types and effector functions. For stimulating the immune response, it may be wise to rely on engineered, synthetic responses in cells without major roles in the immune process (*i.e.*, hMSCs, fibroblasts, platelets, or RBCs, among others) as these cell types and components provide insulation against immunosuppressive signals coming from the tumor environment. However, in other cases, the choice of cell type might be most influenced by the effector functions available natively within that cell type. For instance, if the goal is intratumoral production of a therapeutic antibody, one may find it most expedient to place the antibody gene into the corresponding endogenous loci in a B

cell and then have the biosensor and processor activate the endogenous antibody production effector function. In this case, rather than needing to engineer the antibody folding, glycosylation, and secretion process in another cell type, all of these components are readily available and already tuned. One could also envision a macrophage-based therapy which senses when immunosuppressive microenvironmental signals are pushing the macrophage towards M2 suppressive phenotype; the effector function would be then to activate the macrophage's M1 inflammatory phenotype, with the processor acting on the endogenous loci of the major drivers of this response. However, even for these systems that intend to activate an endogenous response, relying on synthetic parts to accomplish these goals has advantages. The more components of any particular system that rely on endogenous components, the less robust that system will be to choices of cell type and interference from the environment of the cell, namely any signals it is receiving from that environment that are contrary to its therapeutic goals.

1.9 Conclusions

Synthetic biology is an exciting frontier for achieving many of the goals of cell-based therapy engineering and holds promise for overcoming the obstacles I outlined above. Though many therapeutic opportunities are on the horizon with cell-based therapies, delivering highly potent therapeutics directly into tumors is promising strategy for enhancing the safety and efficacy of cancer therapies with engineered cells and the goal I have chosen to focus on for my thesis work. However, many of the technologies needed in order to realize the full potential of cell-based therapies require further development, which I make important strides toward addressing in this thesis. The basic paradigm outlined above is that a given cell-based therapy will need a sensor, processor, and effector function, and my efforts during my thesis work focus on each part of this. In **Chapter 2**, I develop a novel toolkit for processing signals in mammalian cells through custom genetic programs. In **Chapter 3**, I refine a system for sensing extracellular antigens. In **Chapter 4**, I investigate a DNA-based biosensor for sensing intratumoral hypoxia and strategies for its refinement with genetic circuits. In **Chapter 5**, I apply these technologies to build a cell-based therapy to deliver potent therapies into hypoxic tumors. The work in this thesis lays the foundation for a broad range of future cell-based therapies that will employ these foundational technologies to execute their therapeutic functions.

Chapter 2. COMET: a toolkit for engineering custom genetic programs in mammalian cells

2.1 Preface

A version of this chapter was previously published as¹:

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

This chapter describes the development of COMET, a toolkit of synthetic TFs and promoters for engineering gene expression in mammalian cells. When I began my time in the Leonard Lab, the performance of our biosensors was limited, in several ways, by the properties of the available TFs for processing the signal from them. COMET was designed to address these issues and be a modular, scalable system for synthetic gene circuits for many applications. This chapter includes the development of several new chemically inducible dimerization (CID) systems that were developed after the publication of the aforementioned manuscript and will be published separately in a future manuscript. I designed and performed most of the work for this study. Joseph Muldoon carried out all of the modeling work. Hailey Edelstein built and characterized the stable cell cells. Joseph Draut built and characterized the RaZF. The new CID systems were designed, built, and tested by myself, Joseph Draut, Brandon Lim, and Kate Dray. I wrote the initial draft of the manuscript, with contributions from Hailey Edelstein and Joseph Muldoon. All authors edited it.

2.2 Abstract

Engineering mammalian cells to carry out sophisticated and customizable genetic programs requires a toolkit of multiple orthogonal and well-characterized transcription factors (TFs). To address this need, we develop the COmposable Mammalian Elements of Transcription (COMET)—an ensemble of TFs and promoters that enable the design and tuning of gene expression to an extent not previously possible. COMET currently comprises 44 activating and 12 inhibitory zinc-finger TFs and 83 cognate promoters, combined in a framework that readily accommodates new parts. This system can tune gene expression over three orders of magnitude, provides chemically inducible control of TF activity, and enables single-

layer Boolean logic. We also develop a mathematical model that provides mechanistic insights into COMET performance characteristics. Altogether, COMET enables the design and construction of customizable genetic programs in mammalian cells.

2.3 Introduction

The construction of synthetic genetic programs has emerged as a powerful approach for investigating signaling and regulatory networks¹⁶⁶ and for engineering cell-based therapeutic and diagnostic devices^{105, 167}. Applications in mammalian cells often involve designing new ways for cells to sense and respond to internal states or environmental cues. Most programs utilize transcriptional regulation, and while large libraries of components such as transcription factors (TFs) and promoters have been developed for prokaryotes¹⁶⁸, a dearth of analogous parts for mammalian systems currently limits both fundamental research and applications in medicine.

Early synthetic TFs used in eukaryotic cells employ bacterial tetracycline-responsive repressor TetR^{125, 126} or yeast Gal4¹²⁷, and these proteins remain workhorses. New TFs have expanded the pool of orthogonal regulators through programmable DNA recognition, including zinc finger (ZF)-TFs^{133, 134}, transcription activator-like effectors (TALEs)^{130, 131, 132}, dCas9-TFs¹³⁵, and TetR family regulators¹²⁸. ZF-TFs are especially attractive for building a toolkit for transcriptional control, as they are the smallest of these new TFs, affording space for more complex genetic programs under constraints such as gene delivery vehicle cargo limits.

An ideal transcriptional toolkit would include well-characterized TFs and promoters; a physical understanding of how design choices impact performance characteristics; and a quantitative framework that describes how such biological parts may be combined to produce intended behaviors. Such a toolkit should include multiple orthogonal activating and inhibitory TFs; sets of TFs and promoters that enable one to experimentally scan through values of a given performance characteristic; and modularity in TF and promoter design to enable swapping and expansion of the toolkit and interfacing with other biological parts.

To address these needs, we report the COMposable Mammalian Elements of Transcription (COMET)—an ensemble of engineered promoters and modular ZF-TFs with tunable properties. We

incorporate into COMET a panel of 19 TFs that were originally developed in yeast¹⁶⁹ using designed ZF domains¹⁷⁰. We characterize new promoters and then append new functional domains onto the ZFs. In doing so, we elucidate design rules for utilizing TFs and promoters to build gene expression programs exhibiting customizable activation, inhibition, small molecule-responsiveness, and Boolean logic in mammalian cells, and we develop a mathematical model to describe the properties of these genetic parts and programs

2.4 Materials and methods

General DNA assembly

Plasmid cloning was performed primarily using standard PCR and restriction enzyme cloning with Vent DNA Polymerase (New England Biolabs (NEB)), *Taq* DNA Polymerase (NEB), Phusion DNA Polymerase (NEB), restriction enzymes (NEB; Thermo Fisher), T4 DNA Ligase (NEB), Antarctic Phosphatase (NEB), and T4 PNK (NEB). Golden gate assembly and Gibson assembly were also utilized. Most plasmids were transformed into chemically competent TOP10 *E. coli* (Thermo Fisher) and grown at 37°C, except for integration vectors, which were transformed into chemically competent Stable *E. coli* (NEB) and grown at 30°C.

Cloning strategy for COMET vectors

The COMET plasmids are in pcDNA backbones for high expression in HEK293FT cells. Restriction sites were chosen to allow for modular swapping of parts with restriction enzyme cloning. Furthermore, reporter constructs can be assembled by one-step Golden Gate reactions employing synthesized oligonucleotides. A complete list of all plasmids constructed for and utilized in this manuscript is available in **Supplementary Data A2.1**, and plasmid maps are available per **Data Availability**.

Source vectors for DNA assembly

ZF-containing and VP16-containing vectors were a generous gift from Ahmad Khalil¹⁶⁹. VP64 and VPR were sourced from SP-dCas9-VPR, which was a gift from George Church (Addgene plasmid #63798

<https://www.addgene.org/63798/>)¹⁷¹. DsRed-Express2 was obtained by site directed mutagenesis of pDsRed2-N1, which was a gift from David Schaffer (University of California, Berkeley). EBFP2 was sourced from pEBFP2-Nuc, which was a gift from Robert Campbell (Addgene plasmid #14893 <https://www.addgene.org/14893/>)¹⁷². EYFP, FKBP, and FRB were sourced from plasmids we previously described (Addgene plasmids #58855 <https://www.addgene.org/58855/>), #58877 <https://www.addgene.org/58877/>, and #58876 <https://www.addgene.org/58876/>, respectively)¹¹⁵. NanoLuciferase was synthesized as a GeneArt DNA String (Life Technologies/Thermo Fisher). The mMoClo (pLink2, pLink4, and pLink8, Destination Vector, BxB1 Recombinase Expression Vector) plasmids were a gift from Ron Weiss¹⁷³. The CHS4 insulator was sourced from PhiC31-Neo-ins-5xTetO-pEF-H2B-Citrin-ins, which was a gift from Michael Elowitz (Addgene plasmid #78099 <https://www.addgene.org/78099/>)¹⁷⁴. The CAG promoter was sourced from pR26R CAG/GFP Asc, which was a gift from Ralf Kuehn (Addgene plasmid #74285 <https://www.addgene.org/74285/>)¹⁷⁵. The SV40 minimal promoter was sourced from pYC0866 (4xHRE_minSV40-sfGFP-CMV_dsRed Exp), which was a gift from Yvonne Chen⁸⁷. EF1 α and TetON3G were sourced from pLVX-Tet3G (Clontech), and TRE3GV was sourced from pLVX-TRE3G (Clontech). Barcodes used for the TUPVs were designed by the Elledge lab¹⁷⁶. BlastR was sourced from lenti dCAS-VP64_Blast, which was a gift from Feng Zhang (Addgene plasmid #61425 <https://www.addgene.org/61425/>)¹⁷⁷.

Plasmid backbones

All plasmid backbones are modified versions of the pcDNA3.1/Hygro(+) Mammalian Expression Vector (Thermo Fisher V87020). To make pPD003, the SV40 promoter and Hygromycin resistance gene that it drove were removed, while leaving the SV40 origin of replication and SV40 poly(A) signal intact. Additionally, a sense mutation in the *AmpR* gene was introduced to remove a Bsal restriction site. To make pPD005 (referred to as “pcDNA”), the Bpil site in the bGH poly(A) signal was mutated to enable Golden Gate reactions with Bpil, and the Bsal site in the 5'-UTR was mutated to enable Golden Gate reactions with Bsal. The Bpil site was in a region of the BGH poly(A) tail that when deleted does not alter the efficiency of the polyadenylation¹⁷⁸.

Template plasmids for ZF reporter plasmids

pPD027 (the first-generation ZF reporter template) was constructed by inserting a synthesized region (containing two Bsal sites for Golden Gate-mediated ZF binding site array insertion and a YB_TATA minimal promoter⁸⁷) between the BglII and NheI sites and inserting EYFP between the NheI and NotI sites of pPD003. pPD032 and pPD033, which are the templates for ZF reporters with the binding site array moved further upstream of the minimal promoter, were constructed by inserting spacer regions into the BamHI site between the ZF binding array insertion template and the YB_TATA minimal promoter. The spacer inserts were amplified by PCR from the region of pPD003 upstream of the CMV promoter prior to insertion. These three templates (pPD027, pPD032, and pPD033) were used to construct all spaced reporters shown in **Fig. 2.1b–e** and **Fig. 2.2a**.

pPD152 (the second-generation ZF reporter template) was constructed to enable multi-round insertion of larger ZF binding arrays using alternating rounds of Golden Gate with Bsal and Bpil. To do so, the region of pPD027 between the AatII and NotI sites (the ZF binding array insertion site through the end of the EYFP coding sequence) was inserted between the corresponding sites of pPD005. pPD152 was used to make all of the ZF1 compact binding site reporters shown in **Fig. 2.1** and **Fig. 2.2a** (including pPD290 (ZF1x6-C YB_TATA EYFP), which was used as the reporter plasmid in the majority of the experiments), and the logic promoters in **Fig. 2.8**.

pPD540 (the third-generation ZF reporter template) was constructed to swap the palindromic “sticky ends” (5' or 3' overhangs) of the ZF binding array insertion site to non-palindromic sticky ends. The use of palindromic sticky ends, which were originally designed to allow construction of ZF binding arrays with either Golden Gate or EcoRI and BamHI, risks the insertion of multiple copies of the same insert in Golden Gate reactions. This redesign enabled us to inset promoters of sizes that could not be cheaply synthesized as a single insert as multiple inserts in a single round of Golden Gate. This was accomplished by synthesizing a new upstream region (containing two Bsal sites for Golden Gate-mediated ZF binding site array insertion with non-palindromic sticky ends and a YB_TATA minimal promoter) and inserting this upstream region between the BglII and NheI sites of pPD152.

Golden Gate assembly of ZF reporter plasmids

Golden Gate assembly¹⁷⁹ was used to construct most of the reporter plasmids from a reporter template. Promoter insets were synthesized as 15–100 bp oligonucleotides (some promoters were synthesized as multiple inserts) by Integrated DNA Technologies or Life Technologies (Thermo Fisher). The coding and reverse strands were synthesized separately and designed to anneal, resulting in dsDNA with a 4 nt sticky end overhang on each side. The coding and reverse oligonucleotides were mixed (6.5 μL H_2O , 1 μL T4 Ligase Buffer, 0.5 μL T4 PNK (10 U/ μL ; NEB), 1 μL of each 100 μM oligonucleotide) and phosphorylated at 37°C for 1 h. They were then denatured at 95°C for 5 min and cooled slowly to room temperature (here, approximately 22°C) to allow for annealing. The mix was then diluted 50-fold to make a 200 nM stock or 500-fold to make a 20 nM stock. While we made most of the constructs with the 200 nM stock, we later discovered that the 20 nM stock resulted in higher-efficiency reactions.

Bsal Golden Gate reaction mixtures comprise 1 μL T4 ligase buffer, 1 μL 10x BSA (1 mg/mL), 0.5 μL Bsal-HF (20 U/ μL ; NEB), 0.5 μL T4 Ligase (400 U/ μL ; NEB), 10 fmol of vector, 1 μL of each insert (diluted to 200 nM or 20 nM), and water to 10 μL total volume. The reaction was incubated at 37°C for 1 h, 55°C for 15 min, and 80°C for 20 min, and then cooled to room temperature. Up to 10 μL of reaction was immediately transformed into up to 50 μL of chemically competent Top10 *E. coli*. For reactions that did not yield many colonies on the first cloning attempt or did not produce colonies with the correct plasmids, the reaction conditions were changed to: 30 cycles of 37°C for 1 minute then 16°C for 1 minute, 55°C for 15 min, 80°C for 20 min, and cool to room temperature.

Some of the larger ZF binding site arrays were assembled through sequential rounds of alternating Bsal and Bpil Golden Gate reactions. Bpil Golden Gate reaction mixtures comprise 1 μL T4 ligase buffer, 1 μL 10x BSA (1 mg/mL), 0.4 μL Bpil-FD (Thermo Fisher), 0.4 μL T4 Ligase (400 U/ μL ; NEB), 10 fmol of vector, 1 μL of each insert (diluted to 200 nM or 20 nM), and water to 10 μL . The reaction was incubated at 37°C for 30 min, 50°C for 5 min, and 80°C for 10 min, and then cooled to room temperature prior to transformation.

Non-Golden Gate assembly of some ZF reporters

Although Golden Gate assembly was the primary strategy for cloning the promoters, the first-generation templates were not readily amenable to synthesis and insertion of ZF binding site arrays. Therefore, some spaced promoters with large numbers of binding sites used in **Fig. 2.1c** were constructed by PCR amplification of 1–8 binding sites from other reporter plasmids and insertion of these binding sites between the EcoRI and BamHI sites upstream of reporter constructs with 1–8 binding sites in the promoter. Likewise, the ZF reporters with ZF binding site arrays moved further upstream of the minimal promoter shown in **Fig. 2.1d** were constructed by PCR-amplifying the ZF binding site arrays from other constructs and inserting between the EcoRI and BamHI sites of pPD032 and pPD033.

Additionally, COMET reporter constructs were designed to include a limited set of minimal promoters; restriction enzyme cloning was employed to accomplish this as well. pPD1028 (ZF1x6-C SV40_Min EYFP) was cloned from pPD270, cut with XbaI and ApaI. Into this construct we inserted two fragments of DNA: SV40_min⁸⁷ was PCR-amplified and cut with BsaI and ApaI, and EYFP was PCR-amplified from pPD270 and cut with BsaI and ApaI. pPD1029 (ZF1x6-C CMV_min EYF) was cloned from pPD270, cut with XbaI and ApaI. CMV_min was synthesized by IDT and placed upstream of an EYFP gene in a pcDNA-based vector. A fragment comprising CMV_min and EYFP was then PCR-amplified, digested with BsaI and ApaI, and inserted into the digested pPD270.

Assembly of ZFa and ZFi

The first five ZFa tested in **Fig. 2.1b** were constructed by PCR-amplifying the ZFa sequence from¹⁶⁹ (including the N-terminal 3x-FLAG tag, SV40 NLS, VP16 AD, and ZF) and inserting between the NheI site and NotI site of pPD005. Cognate ZFi were constructed by whole-plasmid PCR-mediated deletion of the VP16 AD. During the AD deletion process, BamHI and KpnI sites were added between the SV40 NLS and the ZF, which were later used to insert a PCR-amplified DsRed Express2, thereby creating cognate ZFi-DsRed. Subsequent ZFa (i.e., any new ZFa tested in **Fig. 2.3a**) were constructed by replacing DsRed-Express2 with a PCR-amplified VP16 (BamHI/KpnI) and replacing the ZF domain with a PCR-amplified ZF domain (KpnI/NotI) from¹⁶⁹.

Assembly of ZF mutants

ZFa mutants were synthesized as multiple sets of complementary oligonucleotides, which were annealed and then inserted via Golden Gate assembly into a vector designed to encode ZFa upon insertion of all inserts. Reactions were performed with Bpil as described in **Golden Gate assembly of ZF reporter plasmids**. ZFi mutants were generated by whole-plasmid PCR-mediated deletion of the VP16 AD.

Assembly of RaZFa

RaZFa components were constructed by multi-step restriction enzyme-based cloning. The SV40 NLS was part of the original ZFa constructs¹⁶⁹, and the NES sequence was obtained from¹⁸⁰.

Gibson assembly

Gibson assembly^{181, 182} was used to specify ADs on ZF1a. Gibson reactions were performed by PCR addition of homology arms onto the target DNA. Components were mixed together: 17 fmol of backbone, 51 fmol of each insert, 7.5 μ L of Gibson Master Mix, and water to 10 μ L. 7.5 μ L of Gibson Master mix contains 2 μ L 5X isothermal reaction buffer (0.5 M Tris-HCl pH 7.5, 0.05 M MgCl₂, 1 mM dNTP, 5 mM NAD, 0.05 M DTT), 0.04 U T5 exonuclease, 0.25 U Phusion DNA Polymerase, and 40 U *Taq* DNA Ligase (NEB) in water. The reaction was incubated at 50°C for 1 h, and 5 μ L was transformed into chemically competent Top10 *E. coli* (Thermo Fisher). In subsequent cases, ADs were moved onto other ZF by restriction digest.

Construction of plasmids for mMoClo

We made several changes to the mMoClo plasmids originally described¹⁷³ in order to incorporate them into the workflow for our laboratory, in which many constructs are prototyped using pcDNA-based expression vectors. Details can be found in **Supplementary Fig. A2.9, Supplementary Data 1, 2 (Online¹)**. We modified the Destination Vector provided by the Weiss lab by adding two repeats of the CHS4 insulator into two places in the vector. The insulators upstream of the attB site are, upon genomic integration, inserted downstream of the LP, insulating the LP from the genome (and vice versa). The

insulators downstream of the RB Globin polyA terminator of the puromycin resistance gene insulate this transcription unit from TU1. This new vector is termed pPD630 (Integration Vector). We cloned pLink1, pLink3, pLink5, pLink6, pLink7, and pLink9 site directed mutagenesis via whole plasmid PCR of pLink2.

The TUPVs were cloned by making several alterations to pcDNA (pPD005), in 3 steps. In the first step, two repeats of the CHS4 insulator were placed downstream of the BGH polyA tail. Second, to enable Golden Gate cloning of the TUPV library, three pairs of BsaI sites were inserted into the vector with PCR. The first pair was upstream of the promoter, the second pair was inserted between the BGH polyA tail and the insulator, and the third pair was inserted downstream of the insulator. In the third reaction, three pairs of annealed oligonucleotides were inserted into these BsaI sites via a Golden Gate reaction. The first insert, to be placed upstream of the promoter, comprised a BpiI site, TUPV-specific sticky end, and TUPV-specific 5' barcode (barcodes unique to each TUPV enable sequencing of the TUPV contents after TUPVs are combined into an integration vector). The second insert, to be placed between the BGH/polyA tail and the insulator, comprised a TUPV-specific 3' barcode. The third insert, to be placed downstream of the insulator, comprised a BpiI site and a TUPV-specific sticky end. In this manner, 9 TUPVs each with their own unique 5' and 3' barcodes and 5' and 3' sticky ends were cloned (pPD471–479). This initial library uses a CMV promoter as the core promoter for each TU, which was placed upstream of a multiple cloning site (MCS). A second library of 9 TUPVs was then constructed by replacing the CMV promoter with the CAG promoter by restriction enzyme digest with SnaBI and NheI (pPD561–569). A third library of 9 TUPVs was constructed by replacing the promoter with EF1alpha (between MluI and NheI) and the MCS replaced with an EBFP2-P2A-BlastR gene (between NheI and NotI) (pJM450–458). Although this third library no longer contains the full pcDNA MCS, it retains the NheI and NotI genes that flank the COMET ZFa and ZFi and many of the RaZFa components.

Transferring COMET parts into mMoClo

COMET reporters and ZFa were transferred into TUPVs using restriction enzyme cloning. To construct mKate2 reporters in TUPV1, mKate2 was cloned into the MCS of pPD561 using NheI and NotI restriction sites downstream of a CAG promoter to create pHIE041. Binding site arrays were PCR-amplified

from pPD152, pPD287, pPD290, pPD296, pPD063, pPD069, and pPD095 and inserted to replace CAG in pHIE041 using BglIII and NheI, resulting in pHIE042–049. To construct constitutively expressed VP16-ZF1a in TUPV2 (pJM466), the EBFP-P2A-BlastR in pJM451 was replaced with PCR-amplified VP16-ZF1a from pD100 using NheI and NotI.

mMoClo Assembly of Integration Vectors

The mMoClo integration vectors were assembled through a Bpil-mediated Golden Gate reaction. Each 20 μ L reaction comprised 2 μ L 10x T4 ligase buffer, 2 μ L 10x BSA (1 mg/mL stock), 0.8 μ L Bpil-FD, 0.8 μ L T4 DNA Ligase (400 U/ μ L stock), 20 fmol integration vector backbone (pPD630), and 40 fmol of each transcription unit and linker plasmid to be inserted. The reaction was incubated at 37°C for 15 min, then subjected to 55 iterations of thermocycling (37°C for 5 min, 16°C for 3 min, repeat), followed by 37°C for 15 min, 50°C for 5 min, 80°C for 10 min to terminate the reactions; then the mixture was cooled to room temperature (optionally held at 4°C if the reaction ran overnight) and placed on ice prior to immediate transformation into bacteria.

Plasmid preparation

TOP10 *E. coli* were grown overnight in 100 mL of LB with the appropriate selective antibiotic. The following morning, cells were pelleted at 3000 x g for 10 min and then resuspended in 4 mL of a solution of 25 mM Tris pH 8.0, 10 mM EDTA, and 15% sucrose. Cells were lysed for 15 min by addition of 8 mL of a solution of 0.2 M NaOH and 1% SDS, followed by neutralization with 5 mL of 3 M sodium acetate (pH 5.2). Precipitate was pelleted by centrifugation at 9000 x g for 20 min. Supernatant was decanted and treated with RNase A for 1 h at 37°C. 5 mL of phenol chloroform was added, and the solution was mixed and then centrifuged at 7500 x g for 20 min. The aqueous layer was removed and subjected to another round of phenol chloroform extraction with 7 mL of phenol chloroform. The aqueous layer was then subjected to an isopropanol precipitation (41% final volume isopropanol, 10 min at room temperature, 9000 x g for 20 min), and the pellet was briefly dried and resuspended in 420 μ L of water. The DNA mixture was incubated on ice for at least 12 h in a solution of 6.5% PEG 20,000 and 0.4 M NaCl (1 mL final volume). DNA was

precipitated with centrifugation at maximum speed for 20 min. The pellet was washed once with ethanol, dried for several h at 37°C, and resuspended for several h in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA purity and concentration were confirmed using a Nanodrop 2000 (Thermo Fisher).

Cell culture

The HEK293FT cell line was purchased from Thermo Fisher/Life Technologies (RRID: CVCL_6911 [https://web.expasy.org/cellosaurus/CVCL_6911]) and was not further authenticated. The HEK293FT-LP cell line was a gift from Ron Weiss and was authenticated by flow cytometric analysis of EYFP expression, which was shown to be homogenous and stable over time—a pattern which is consistent with the original description of this cell line¹⁷³. Cells were cultured in DMEM (Gibco 31600-091) with 10% FBS (Gibco 16140-071), 6 mM L-glutamine (2 mM from Gibco 31600-091 and 4 mM from additional Gibco 25030-081), penicillin (100 U/μL), and streptomycin (100 μg/mL) (Gibco 15140122), in a 37°C incubator with 5% CO₂. Cells were subcultured at a 1:5 to 1:10 ratio every 2–3 d using Trypsin-EDTA (Gibco 25300-054). The HEK293FT cell line and the HEK293FT-LP cell line tested negative for mycoplasma with the MycoAlert Mycoplasma Detection Kit (Lonza Cat #LT07-318).

Transfection experiments

Experiments were conducted by transient transfection of HEK293FT cells using the calcium phosphate method. For transfection experiments, cells were plated at a minimum density of 1.5×10^5 cells/well in a 24-well plate in 0.5 mL of DMEM, supplemented as described above. After at least 6 h, by which time the cells had adhered to the plate, they were transfected via the calcium phosphate method. Plasmids for each experiment were mixed in H₂O, and 2 M CaCl₂ was added to a final concentration of 0.3 M CaCl₂. The exact DNA amounts added to the mix per well and plasmid details for each experiment are listed in the following sections and can be cross-referenced with **Supplementary Data 2 (Online¹)** for further details. This mixture was added dropwise to an equal-volume solution of 2x HEPES-Buffered Saline (280 mM NaCl, 0.5 M HEPES, 1.5 mM Na₂HPO₄) and gently pipetted up and down four times. After 2.5–4 min, the solution was mixed vigorously by pipetting eight times. 100 μL of this mixture was added dropwise

to the plated cells, and the plates were swirled gently. The next morning, the medium was aspirated and replaced with fresh medium. In some assays, fresh medium contained 0.05% DMSO or 0.05% DMSO with 0.1 μ M rapamycin. At 36–48 h post-transfection and at least 24 h post-media change, cells were harvested for flow cytometry with FACS Buffer (PBS pH 7.4 with 2–5 mM EDTA and 0.1% BSA) or with Trypsin-EDTA, which was then quenched with medium, and the resulting cell solution was added to at least 2 volumes of FACS buffer. Cells were spun at 150 x g for 5 min, FACS buffer was decanted, and fresh FACS buffer was added. All experiments were performed in biologic triplicate.

Western Blotting

For western blotting, HEK293FT cells were plated at 7.5×10^5 cells/well in 2 mL of DMEM and transfected as above, using 400 μ L of transfection reagent per well (the reaction scales with the volume of medium). At 36–48 h after transfection, the cells were lysed with 500 μ L of RIPA (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate) with protease inhibitor cocktail (Pierce/Thermo Fisher cat# A32953) and incubated on ice for 30 min. The lysate was cleared by centrifugation at 14,000 x g for 20 min at 4°C and the supernatant was harvested. A BCA assay was performed to determine protein concentration, and after a 10-minute incubation with Lamelli buffer (final concentration 60 mM Tris-HCl pH 6.8, 10% glycerol, 2% sodium dodecyl sulfate, 100 mM dithiothreitol, and 0.01% bromophenol blue) at 70°C, 0.5 μ g of total protein was loaded onto a 4-15% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad) and run at 50 V for 10 min followed by 100 V for at least 1 h. Wet transfer was performed onto an Immuno-Blot PVDF membrane (Bio-Rad) for 45 min at 100 V. Ponceau-S staining was used to confirm successful transfer. Membranes were blocked for 30 min with 3% milk in Tris-buffered saline pH 8.0 (TBS pH 8.0: 50 mM Tris, 138 mM NaCl, 2.7 mM KCl, HCl to pH 8.0), washed once with TBS pH 8.0 for 5 min, then incubated for 1 h at room temperature or overnight at 4°C in primary solution antibody (Mouse-anti-FLAG M2 (Sigma F1804, RRID: AB_262044 [http://antibodyregistry.org/AB_262044]), diluted 1:1000 in 3% milk in TBS pH 8.0). Primary antibody solution was decanted, and the membrane was washed once with TBS pH 8.0 then twice with TBS pH 8.0 with 0.05% Tween, for 5 min each. Secondary antibody (HRP-anti-Mouse (CST 7076, RRID: AB_330924

http://antibodyregistry.org/AB_330924), diluted 1:3000 in 5% milk in TBST pH 7.6 (TBST pH 7.6: 50 mM Tris, 150 mM NaCl, HCl to pH 7.6, 0.1% Tween)) was applied for 1 h at room temperature, and the membrane was washed three times for 5 min each time with TBST pH 7.6. The membrane was incubated with Clarity Western ECL Substrate (Bio-Rad) for 5 min, and then exposed to film, which was developed and scanned. Images were cropped with Photoshop CC (Adobe). No other image processing was employed. Original images are available on request.

The western blot shown in **Supplementary Fig. A2.12f** was conducted twice with comparable results. The first experiment included only the RaZFa component (no additional loading control) to confirm the presence of only one band in each lane (data not shown). In the second experiment, 40 ng of pPD798 (encoding a 3X-FLAG tagged NanoLuciferase) was co-transfected with the RaZFa components to provide a control for loading and transfection.

Analytical flow cytometry

Flow cytometry was run on a BD LSRII or BD LSR Fortessa Special Order Research Product (Robert H. Lurie Cancer Center Flow Cytometry Core). The lasers and filter sets used for data acquisition are listed in **Supplementary Table A2.3**. Approximately 2,000–3,000 single, transfected cells were analyzed per sample.

Flow Cytometry Data Analysis

Samples were analyzed using FlowJo v10 software (FlowJo, LLC). As illustrated in **Supplementary Fig. A2.15**, the HEK293FT cell population was identified by FSC-A vs. SSC-A gating, and singlets were identified by FSC-A vs. FSC-H gating. To distinguish transfected and non-transfected cells, a control sample of cells was generated by transfecting cells with a mass of pcDNA (empty vector) equivalent to the mass of DNA used in other samples in the experiment. For the single-cell subpopulation of the pcDNA-only sample, a gate was made to identify cells that were positive for the constitutively driven fluorescent protein used as a transfection control in other samples, such that the gate included no more than 1% of the non-fluorescent cells. The mean fluorescence intensity (MFI) of the single-cell transfected population was

calculated and exported for further analysis.

To calculate reporter expression, MFI in the FITC channel was averaged across three biologic replicates. From this number, the autofluorescence of the cells was subtracted. To calculate the autofluorescence of the cells, in each experiment, a control group of cells transfected with DNA encoding the fluorescent protein transfection control and pcDNA were used. The background-subtracted MFI was converted to Mean Equivalent of Fluorescein (MEFLs) by multiplying by a coefficient determined in each experiment, as described below. Standard error was propagated through all calculations.

Conversion of arbitrary units to standardized fluorescence units

As shown in **Supplementary Fig. A2.16**, to determine the conversion factor for MFI to MEFLs, Rainbow Calibration Particles (Spherotech, RCP-30-5) or UltraRainbow Calibration Particles (Spherotech URCP-100-2H) were run with each flow cytometry experiment. This reagent contains six (RCP) or nine (URCP) subpopulations of beads, each of a specific size and with a known number of various fluorophores. The total bead population was identified by SSC vs. FSC gating, and the subpopulations were identified through two fluorescent channels. The MEFL values corresponding to each subpopulation were supplied by the manufacturer. A calibration curve was generated for the experimentally determined MFI vs. manufacturer supplied MEFLs, and a linear regression was performed with the constraint that 0 MFI equals 0 MEFLs. The slope from the regression was used as the conversion factor, and error was propagated.

Integration of cargo into landing pad cell lines

From exponentially growing HEK293LP cells, 0.5×10^5 cells were plated per well (0.5 mL medium) in 24-well format, and cells were cultured for 24 h to allow cells to attach and spread. When cells reached 50–75% confluence, Bxb1 recombinase was co-transfected with the integration vector by lipofection with Lipofectamine LTX with PLUS Reagent (ThermoFisher 15338100). 300 ng of BxB1 expression vector was mixed with 300 ng of integration vector and 0.5 μ L of PLUS reagent in a 25 μ L total volume reaction, with the remainder of the volume being OptiMEM (ThermoFisher/Gibco 31985062). In a separate tube, 1.9 μ L of LTX reagent was mixed with 23.1 μ L of OptiMEM. The DNA/PLUS Reagent mix was added to the LTX

mix. pipetted up and down four times, and then incubated at room temperature for 5 min. 50 μ L of this transfection mix was added drop-wise to each well of cells, which was mixed by gentle swirling. Cells were cultured until the well was ready to split (typically 3 d), without any media changes.

Selection and expansion of landing pad cell lines

Cells were harvested from the 24-well plate when confluent by trypsinizing and transferring to a single well of a 6-well plate in 2 mL of medium, and then cells were cultured until they reached 50-70% confluence. Then, medium was aspirated and replaced with 2 mL of fresh media containing appropriate selection antibiotic 1 μ g/mL puromycin (Invivogen ant-pr) or 6 μ g/mL blasticidin (Alfa Aesar/ThermoFisher J61883). Medium was replaced daily with fresh medium containing antibiotics until cell death was no longer evident. Selection was first performed in puromycin for 7 d, then cells were expanded for 7 d without antibiotics. Cells were then cultured in both puromycin and blasticidin to maintain selective pressure until flow sorting.

Sorting of landing pad cell lines

Cells were harvested by trypsinizing, resuspended at approximately 10^7 cells per mL in pre-sort medium (DMEM with 10% FBS, 25 mM HEPES (Sigma H3375), and 100ug/mL gentamycin (Amresco 0304)), and held on ice until sorting was performed. Cells were sorted using a BD FACS Aria 4-laser Special Order Research Product (Robert H. Lurie Cancer Center Flow Cytometry Core) with the optical configuration listed in **Supplementary Table A2.4**.

The sorting strategy was as follows: single cells were first gated to exclude all EYFP positive cells (as EYFP positive cells still have an intact landing pad locus, suggesting a mis-integration event occurred) and to include only EBFP2+ cells. Then a gate was drawn on EBFP2 expression, utilizing the line that demonstrated the least amount of silencing (ZF1x12-C_mKate2 + ZF1a) to capture the 90th to 98th percentile of EBFP2 expressing cells (the top 2% were excluded to exclude cells suspected to possess two or more integrated copies of the cargo vector). The gate drawn using this line was used for all other lines as well. No gating was performed on mKate2 reporter expression. 15,000 cells were collected for each line

in post-sort medium (DMEM with 20% FBS, 25 mM HEPES, and 100 µg/mL gentamycin), and cells were held on ice until they could be centrifuged at 150 x g for 5 min and resuspended in DMEM. Cells were plated in a 24-well plate and expanded until used in experiments. Gentamycin was included in the culture medium for one week after sorting.

Experiments involving landing pad cell lines

Stable cell lines were plated in 0.5 mL of DMEM in triplicate in 24-well format at a density expected to generate 50% confluent wells. The day after plating (24 h), cells were harvested with Trypsin-EDTA, as described in **Transfection Experiments**. For transfection experiments designed to accompany landing pad line experiments (**Fig. 2.6a,b**), cells were plated and transfected 2 d prior to the assay and harvested as described in **Transfection Experiments**. Flow cytometry was run on a BD LSR Fortessa, as described in **Analytical flow cytometry**.

For characterization, approximately 10,000 single, EBFP2-expressing cells were analyzed per sample, where EBFP2 is a marker for locus activity in the stable cells and a transfection control for transfected cells. Stable cells were analyzed using higher laser voltages than those used for transfected cells to effectively capture the range of reporter expression conferred by the panel of COMET promoters; thus, the results from this experiment are displayed in separate panels of **Fig. 2.6** even though data collection occurred on the same day.

COMET model development and analysis

This section provides an integrated discussion of model development, calibration, and analysis to supplement the discussion in the main text. We first describe the development of the core model and then discuss elaborations and models used for comparison. In developing the core model to investigate and predict COMET behavior, we account for two phenomena: cell heterogeneity using a statistical model, and gene regulation using a dynamical model.

Statistical Analysis

Statistical details for each experiment are in the figure legends. Unless otherwise stated, there are three independent biologic replicates for each condition. The data shown reflect the mean across these biologic replicates of the mean fluorescence intensity (MFI) of approximately 2,000–3,000 single, transfected cells. Error bars represent the standard error of the mean (S.E.M.). For main figures with heat maps, data are also shown in the corresponding supplemental figure as a bar graph with the mean and S.E.M.

ANOVA tests were performed using the Data Analysis Toolpak in Microsoft Excel. Tukey's HSD tests were performed with $\alpha = 0.05$. Pairwise comparisons were made using a one-tailed Welch's *t*-test, which is a version of Student's *t*-test in which the variance between samples is treated as not necessarily equal. The comparisons involved reporter only vs. reporter + ZFa in **Fig. 2.1**, **Fig. 2.3**; inhibited vs. uninhibited, or more inhibited vs. less inhibited, in **Fig. 2.5**; no binding sites vs. one binding site in **Supplementary Fig. A2.10**; DMSO vs. rapamycin in **Fig. 2.7**; and summed individual cases vs. co-expression in **Fig. 2.8**. For each comparison, the null hypothesis was that two samples were equal, and the alternative was that the latter was greater. The threshold for significance was set at 0.05. To decrease the false discovery rate, the Benjamini-Hochberg (BH) procedure was applied to each set of tests per figure panel; in all tests, after the BH procedure, the null hypothesis was rejected for *p*-values < 0.05. The outcome of each statistical test is indicated in the figure captions.

2.4 Results

Identifying promoter design rules in mammalian cells

In nature, TFs based on ZF domains coordinate diverse functions.¹⁸³ For example, the evolutionarily ancient and widely expressed SP1 contains three Cys2-His2-type ZF motifs (generally considered a minimal ZF domain), and SP1 binding sites appear as tandem arrays in genes regulating cell growth, apoptosis, and immune function, as well as in compact, dynamically regulated viral promoters such as the long terminal repeat of HIV¹⁸⁴. To begin developing a toolkit for constructing transcriptional programs from basic parts, we first considered five synthetic ZF domains characterized in yeast by Khalil et al.¹⁶⁹ and investigated whether these tools could be adapted to regulate transcription in mammalian cells. In this mammalian library, each TF comprises a ZF DNA-binding domain fused to the VP16 activation domain (AD), forming a ZF activator (ZFa) that recruits RNA polymerase II (RNAPII) and induces transcription¹⁸⁵. A new cognate promoter was generated for each ZFa by placing one ZF binding site upstream of the YB_TATA minimal promoter (**Fig. 2.1a**), which confers low background and inducible expression in several cell types^{87, 186}. All five ZFa induced expression from their cognate reporters between 4 and 17-fold above background (ZFa-independent) (**Fig. 2.1b, Supplementary Fig. A2.1a**). Interestingly, the rank order of the magnitudes with which these ZFa induced their cognate reporters differed from that observed in a similar system in yeast¹⁶⁹.

Initial reporter output was relatively dim—on the order of 10^5 Molecules of Equivalent Fluorescein (MEFLs, an absolute unit of fluorescence¹⁸⁷) per cell—so we explored strategies for building stronger inducible promoters. An established principle is that inducible gene expression increases with the number of TF binding sites, so we tested a panel of ZF1a-responsive promoters containing multiple ZF1 sites in an array upstream from the minimal promoter (**Fig. 2.1c, Supplementary Fig. A2.1b**). In general, ZF1a-inducible reporter expression increased with the number of sites while background was unaffected. The ZF1 promoter with 12 sites (ZF1x12) yielded 113-fold induction—approximately 12 times greater than the ZF1x1 promoter.

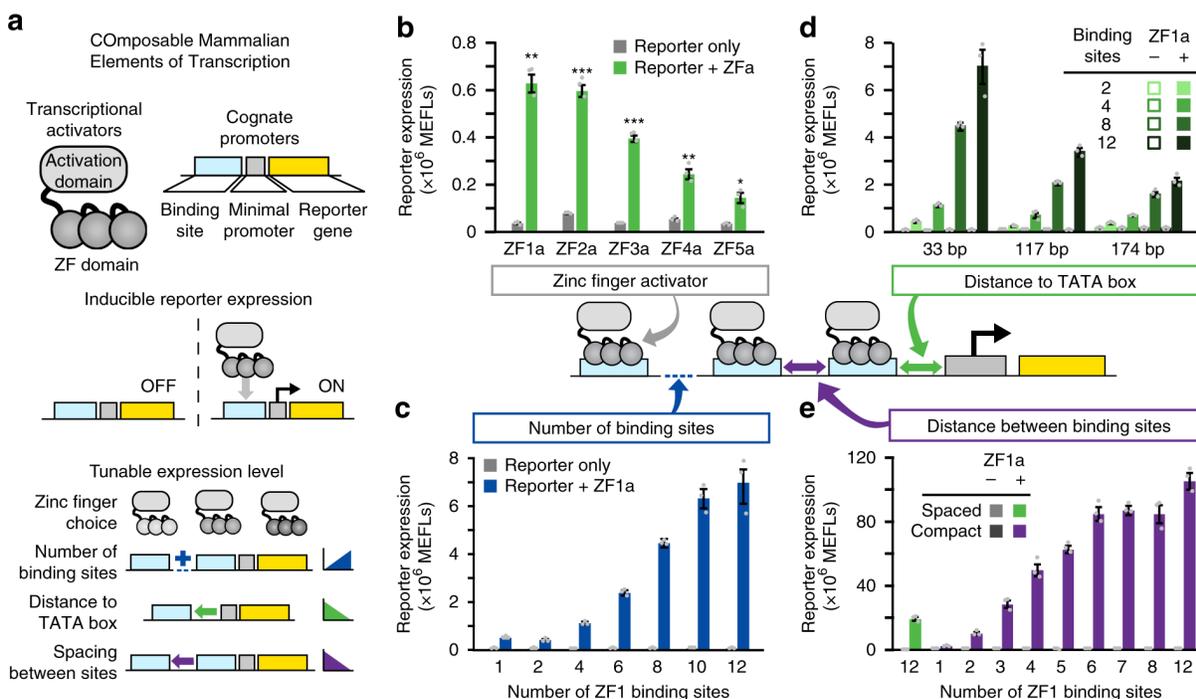


Fig 2.1 Investigation of COMET promoter design rules. **a** The schematic shows the modular, tunable features of COMET TFs and promoters. **b** Five ZFa with different ZF domains all induced reporter expression (one-tailed Welch's t -test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). **c** Increasing the number of ZF binding sites increased the level of gene expression in the presence of ZFa (ANOVA $p < 0.001$) but not without ZFa (ANOVA $p = 0.24$). Reporter expression increased significantly from 6 to 8 and from 8 to 10 binding sites but not on either side of this range (Tukey's HSD test with $\alpha = 0.05$). **d** Moving the ZF binding site array further upstream of the TATA box reduced reporter expression (two-factor ANOVA $p < 0.001$), and arrays with more binding sites showed more substantial decreases in reporter expression. **e** Compaction of ZFa binding sites enhanced ZFa-induced reporter expression, for an equivalent number of ZF binding sites (one-tailed Welch's t -test, $p = 0.002$), and across compact promoters, ZFa-induced reporter expression increased with the number of binding sites (ANOVA $p < 0.001$). Reporter expression increased significantly from 2 to 3, 3 to 4, 5 to 6, and 8 to 12 binding sites (Tukey's HSD test with $\alpha = 0.05$). Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Source data are provided in the **Source Data (Online¹)** file.

We hypothesized that expression might be influenced by the distance between the ZF binding site array and the TATA box, either favoring or blocking interactions with RNAPII, as was observed previously with synthetic promoters¹⁸⁸. To investigate, we constructed promoters with the binding site array moved from 33 bp upstream of the TATA box (the original position) to 117 bp or 174 bp upstream. Overall, increasing spacing led to decreased expression, and this effect was especially pronounced for promoters with many sites (**Fig. 2.1d**). Thus, when ZFa bind farther from the promoter, the AD is seemingly too distant to contribute to transcriptional activation. This mechanism would also explain diminishing returns observed when adding sites to large arrays (**Fig. 2.1c**). In summary, increasing ZF-TF binding site count enhances gene expression, but only if the sites are near the TATA box.

Given these findings, we next investigated whether compacting binding sites near the minimal promoter could potentiate transcriptional output. Our initial constructs had 16 to 38 bp spacers between each 9 bp binding site. To generate a more compact structure, constructs were generated with 6 bp spacers, such that ZFa would bind 15 bp apart in a rotating configuration around the DNA, as one turn of the double helix is 10.5 bp. We hypothesized that this configuration could avoid steric occlusion while increasing the local concentration of ZFa. A panel of compact promoters was generated, each containing 1–12 binding sites in an array beginning 33 bp upstream of the TATA box. These new promoters yielded strong output and 360-fold induction over background (**Fig. 2.1e, Supplementary Fig. A2.1c**). The output of the strongest compact promoter (ZF1x12-C) was over five-times greater than that of the comparable spaced promoter (ZF1x12-S). Background remained low across all constructs. Several of the strongest promoters exhibited mild squelching—a phenomenon in which inducing the expression of a TF (here, ZFa, which is expected to induce expression of the EYFP reporter) causes unexpected diminishment in the expression of a gene (here, the constitutively expressed EBFP2 transfection control) by competing for a limited pool of cellular resources that carry out transcription^{185, 189}. Here, squelching is apparent when cells with high EYFP expression have lower EBFP2 expression than do cells with lower EYFP expression (**Supplementary Fig. A2.1c**). Thus, COMET ZFa and promoters can be potentiated until they saturate the cellular capacity for transgene expression¹⁹⁰, and one can use simple rules to titrate transcriptional output.

Elucidating mechanisms of COMET gene expression

Several observations prompted investigation into the COMET mechanism. At high doses of ZFa plasmid, reporter output plateaued at different levels depending on promoter architecture (**Fig. 2.2a**, **Supplementary Fig. 2.2a–c**). This plateau did not increase by switching minimal promoters, although some choices led to higher background (**Supplementary Fig. 2.2d**). Reporter expression increased with total plasmid dose (while holding the ratio of the ZFa plasmid and reporter plasmid constant), suggesting that transcription, and not translation, limits reporter output (**Supplementary Fig. 2.2e**).

To help elucidate the mechanisms by which COMET operates, we developed a mathematical model of this system. As summarized in **Figure 2.2** and detailed in **Methods**, we first considered mechanistic steps of gene expression, wrote equations capturing these steps (writing such equations is tantamount to formulating a hypothesis as to how gene expression operates), identified a formulation consistent with experimental observations, and simplified this representation by removing details not required to describe observed trends in order to generate a concise model. Finally, we fit parameters of the concise model to data in order to quantitatively describe experimental observations. We hypothesized that this process should generate a set of experimentally grounded parameters representing interpretable features of TF-promoter activity. Throughout, our goal was not to predict TF or promoter sequences *de novo*, but rather to describe and provide insight into observed trends. The explanatory value of such a model often exceeds insights that are accessible by intuition alone, and ultimately this framework could be used to design new genetic functions based upon COMET parts.

We initiated this process by using first principles to produce a detailed model with features of transcriptional control¹⁹¹ including physical and functional interactions between the promoter, TFs, and proteins like RNAPII (**Fig. 2.2b**, **Methods**). This detailed model relates transcriptional output to TF concentration, TF-DNA binding affinity, TF-DNA binding cooperativity, RNAPII recruitment cooperativity, and maximum promoter activation. We then generated a series of theoretical landscapes analogous to the experimental landscapes in **Fig. 2.2a**, varying parameters across a biologically reasonable range, and observed that the landscapes fell within one of four categories defined with respect to the concavity and sigmoidicity of cross-sections along each axis (**Fig. 2.2c**). The experimental data most closely resembled

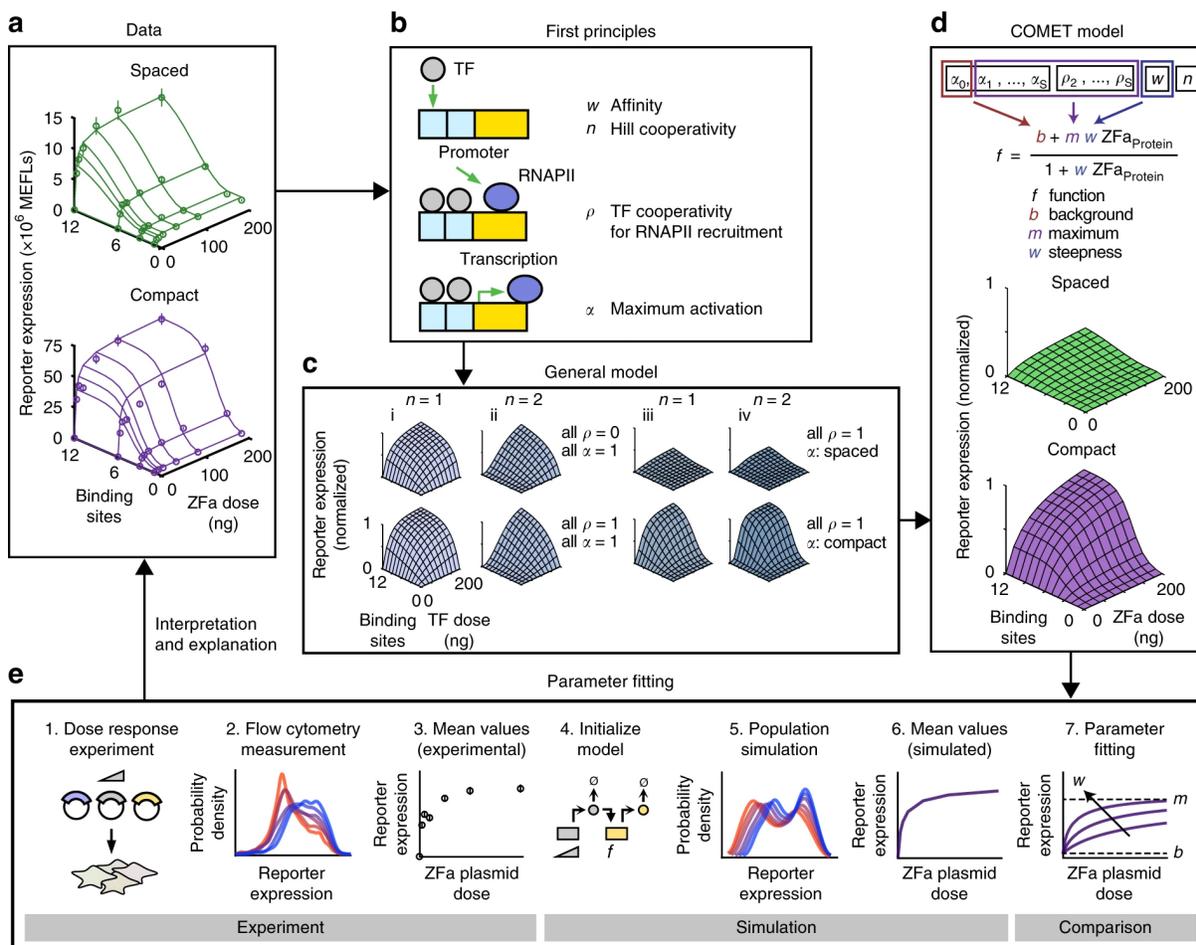


Fig 2.2 A model for COMET-mediated gene regulation. This figure summarizes the process of model development, refinement, and fitting. **a** The COMET model (model outputs are represented by the lines on each plot) explains experimentally observed trends (circles) for reporter expression as a function of ZFa dose and promoter features. This model uses a fitted response function for ZFa-induced gene expression (discussed in **b–e**) and simulates a cell population to account for variation in gene expression (**Supplementary Fig. A2.3**); lines depict the average outcome for the population. **b** We started with a detailed model of transcriptional activation in which reporter expression depends on TF concentration, a metric related to TF-DNA binding affinity (w), TF-DNA binding cooperativity ($n = 1$ for non-cooperative, $n > 1$ for cooperative), RNAPII recruitment cooperativity by each multiple-TF configuration at a promoter ($\rho = 0$ for non-cooperative, $\rho > 0$ for cooperative), and maximum promoter activation by each configuration ($0 \leq \alpha \leq 1$). **c** This model yielded four types of landscapes (i–iv) under different assumptions, and two representative examples of each type are shown. COMET most closely resembles (iii). **d,e** A model that represents ZFa-induced reporter expression by a response function was used to fit the data in **a** (workflow for parameter estimation shown in **e**). The terms in this concise model can be related to terms in the mechanistic model. Landscapes in **c,d** are simulations of a single cell (homogenous model), and those in **a** are simulated mean values for a heterogeneous population. The outputs of this final fitted model are represented alongside experimental data in **a**. Source data are provided in the **Source Data (Online¹)** file.

case (iii), indicating that TF-DNA binding is non-cooperative, but RNAPII recruitment is cooperative, and the maximum transcription rate (at a high ZFa dose) increases with both the number and compactness of binding sites. Therefore, the enhanced potency of the compact promoters stems from the cooperative recruitment of transcriptional machinery.

Based upon the observed ZFa dose response profiles (**Fig. 2.2a**) and these insights, we proposed a concise response function to represent the rate of transcription (f) as a function of ZFa dose with three parameters: background (TF-independent) transcription (b), a steepness metric (w) related to TF-DNA-binding affinity, and a metric for maximum transcription (m) (**Fig. 2.2d, Methods**). As indicated, the three parameters in this concise response function can be related to the additional parameters in the original detailed representation. For a given ZFa-promoter combination, m is experimentally determined and is based upon the number and spacing of binding sites in the promoter, and b is determined based on reporter expression without ZFa; w can be fit to ZFa dose response data by our previously developed method that improves parameter estimation by accounting for variation in gene expression¹¹⁷ (**Fig. 2.2e, Supplementary Fig. A2.3a–c**; fitted parameters are listed in **Supplementary Tables A2.1–A2.2**). Simulated data from the calibrated model provided close agreement with the experimental data, demonstrating that a concise representation can be used to analyze and describe COMET-mediated gene expression.

Comparison of the calibrated model and experimental data confirmed two trends that hold across conditions (**Supplementary Fig. A2.3d**). First, the dependence of relative reporter output on binding site number is independent of the dose of ZFa plasmid when the output is scaled to its maximum value in each binding site series. Second, the dependence of relative reporter output on ZFa dose is independent of the number of binding sites when the output is scaled to its maximum value in each dose series. Thus, inducible gene expression follows a pattern that holds across various promoter designs and that is captured by a concise model. The occurrence of these similar patterns, when paired with the properties elucidated by the model, makes ZFa-induced gene expression readily interpretable and ultimately usable—these are desirable features for a transcriptional toolkit.

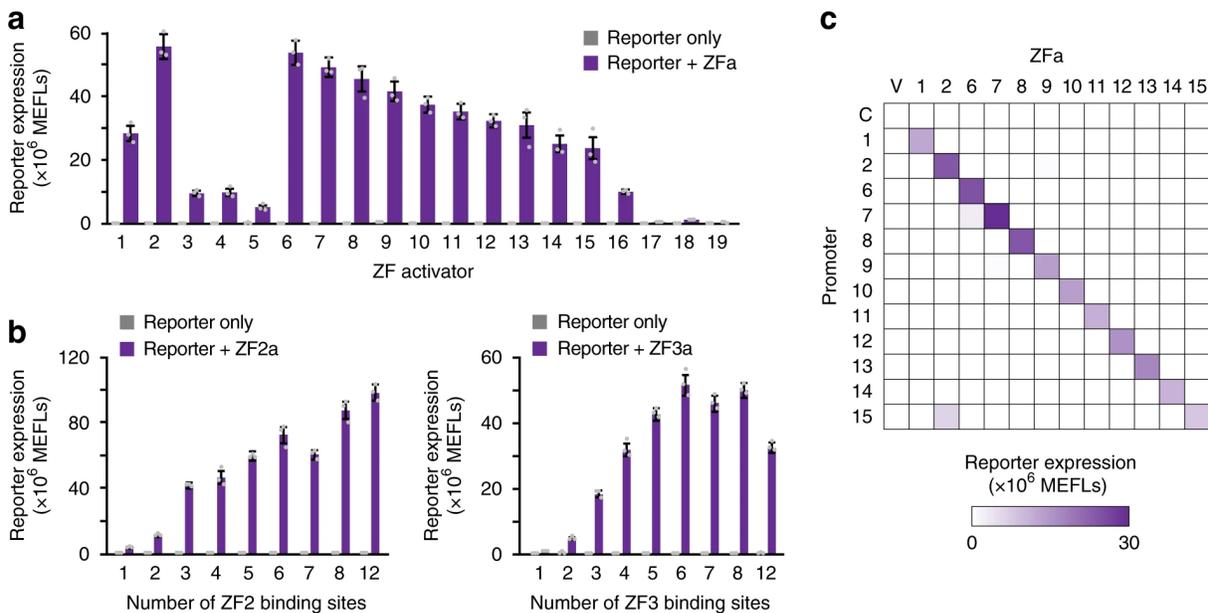


Fig 2.3 Characterizing an expanded panel of ZFa. **a** Nineteen ZFa were paired with cognate x6-C promoters, and all significantly induced gene expression (one-tailed Welch's t -test all $p < 0.02$). **b** ZFa-induced gene expression increased with the number of binding sites, on compact promoters, for ZF2 (ANOVA $p < 0.001$) and ZF3 (ANOVA $p < 0.001$). **c** Investigating the orthogonality between the 12 strongest ZFa using x6-C promoters. Abbreviations: V (Vector control, no ZFa gene), C (Control reporter, no ZF binding sites). Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Source data are provided in the **Source Data (Online¹)** file.

ZFa library characterization and orthogonality

Building upon our initial characterization of five ZFa (**Fig. 2.1b**), we evaluated whether 19 previously characterized ZFa¹⁶⁹ could activate gene expression in mammalian cells. We observed that all ZFa drove transcription from their x6-C cognate promoters to varying extents (**Fig. 2.3a, Supplementary Fig. A2.4a**). Dose response profiles for the strongest 12 ZFa revealed a set of uncorrelated m and w values (**Supplementary Table A2.1, Supplementary Fig. A2.5a–c**). Additionally, the magnitude of induced reporter expression varied substantially between ZFa, which we hypothesized might be due to differential ZF affinity for binding cognate DNA sequences. Since the base pair upstream and base pair downstream (flanking nucleotides) of each 9 bp binding site affect ZF affinity¹⁹², we revisited promoters for two ZFa with contrasting outcomes in **Fig. 2.3a** (ZF2a for high expression and ZF3a for low expression) and observed that changing the flanking nucleotides significantly affected outcomes (**Supplementary Fig. A2.5d**). In both cases, changes guided by prior knowledge^{169, 192} increased transcriptional activation, and thus choice of flanking nucleotides can be used to tune transcriptional activation. To test whether the magnitude of reporter induction mediated by ZF2a and ZF3a depends on the number of binding sites in a manner similar to that observed for ZF1a (**Fig. 2.1e**), we varied the number of sites using compact promoters, and observed a similar trend for up to eight sites (**Fig. 2.3b**). Interestingly, there was a small decrease in reporter expression as the number of binding sites increased from 6 to 7 for both ZF2 and ZF3. It is possible that some promoter architectures, such as ZFx7-C, involve mechanisms that result in slight deviations from overall trends.

To test whether ZFa-mediated activation of cognate promoters is orthogonal across ZFa-promoter combinations, we performed a series of pairwise evaluations using the twelve strongest ZFa and x6-C reporters. The highest expression from each promoter was observed with its cognate ZFa (**Fig. 2.3c, Supplementary Fig. A2.5e**). Of the 132 pairs of ZFa and non-cognate promoters, 80% showed less than 1% of the maximal expression from that promoter (i.e., off-target activation), and 97% showed less than 3% off-target activation. The highest off-target activation of a ZFa/non-cognate promoter pair (ZF2a/ZF15x6-C at 75%) may be explained by the similarities in the binding site sequences for ZF2 and ZF15 (7 of 9 bp in common). However, such sequence similarities were not noted for the next three highest off-target combinations (ZF6a/ZF7x6-C at 10%, ZF7a/ZF15x6-C at 6%, and ZF7a/ZF9x6-C at 4% off-target

activation). Overall, COMET ZFa are generally orthogonal from one another and are thus well-suited to composing genetic programs requiring multiple independent transcription units.

Tuning transcription through protein engineering

Having explored strategies for tuning gene expression by promoter engineering, we next investigated two strategies for tuning via protein engineering: altering the affinity of the ZF for the DNA and altering the strength of the AD. For the first strategy, we mutated four arginine residues in the ZF that interact with the DNA backbone (**Fig. 2.4a**). Arginine-to-alanine substitutions in these positions ablate favorable charge interactions between the ZF and negatively-charged phosphates in the DNA backbone and decrease ZF1a-induced expression in yeast^{169, 193, 194}. As expected, ZFa-mediated gene expression decreased with an increasing number of such substitutions (**Fig. 2.4b**, **Supplementary Fig. A2.6a,b**). Interestingly, while changing the promoter architecture affected only the maximum transcription (m) (**Fig. 2.2**), ZF mutations affected both the maximum transcription and the relative steepness of the ZFa dose response curve (m and w). Additionally, the changes in these values were correlated, revealing an axis along which ZFa R-to-A mutations tune TF strength. This result differs from our previous analysis of ZF domain choice, which affected m and w in an uncorrelated manner (**Supplementary Fig. A2.5c**). R-to-A mutations decreased ZFa-induced transcription in a manner that was similar across various numbers of binding sites in the promoter (**Supplementary Fig. A2.6c**), showing that this tuning can be applied across a variety of promoters. For the second tuning strategy, we tested two ADs in place of VP16: VP64¹⁹⁵ and VPR¹⁷¹ (**Fig. 2.4c**). When fused in place of VP16, VPR produced the highest expression across several promoters, and VP64 was modestly stronger than VP16 in some cases (**Fig. 2.4d**). The relative effect of AD choice diminished as the number of ZF binding sites increased, suggesting that cooperative transcriptional activation by multiple weakly activating TFs (e.g., those based upon VP16), can approach the same magnitude of transcriptional activation mediated by fewer potentially activating TFs (e.g., those based upon VPR). Replacing the AD on four other ZFa led to similar results (**Supplementary Fig. A2.6d**). Overall, these observations support the utility of multiple TF engineering strategies for tuning gene expression.

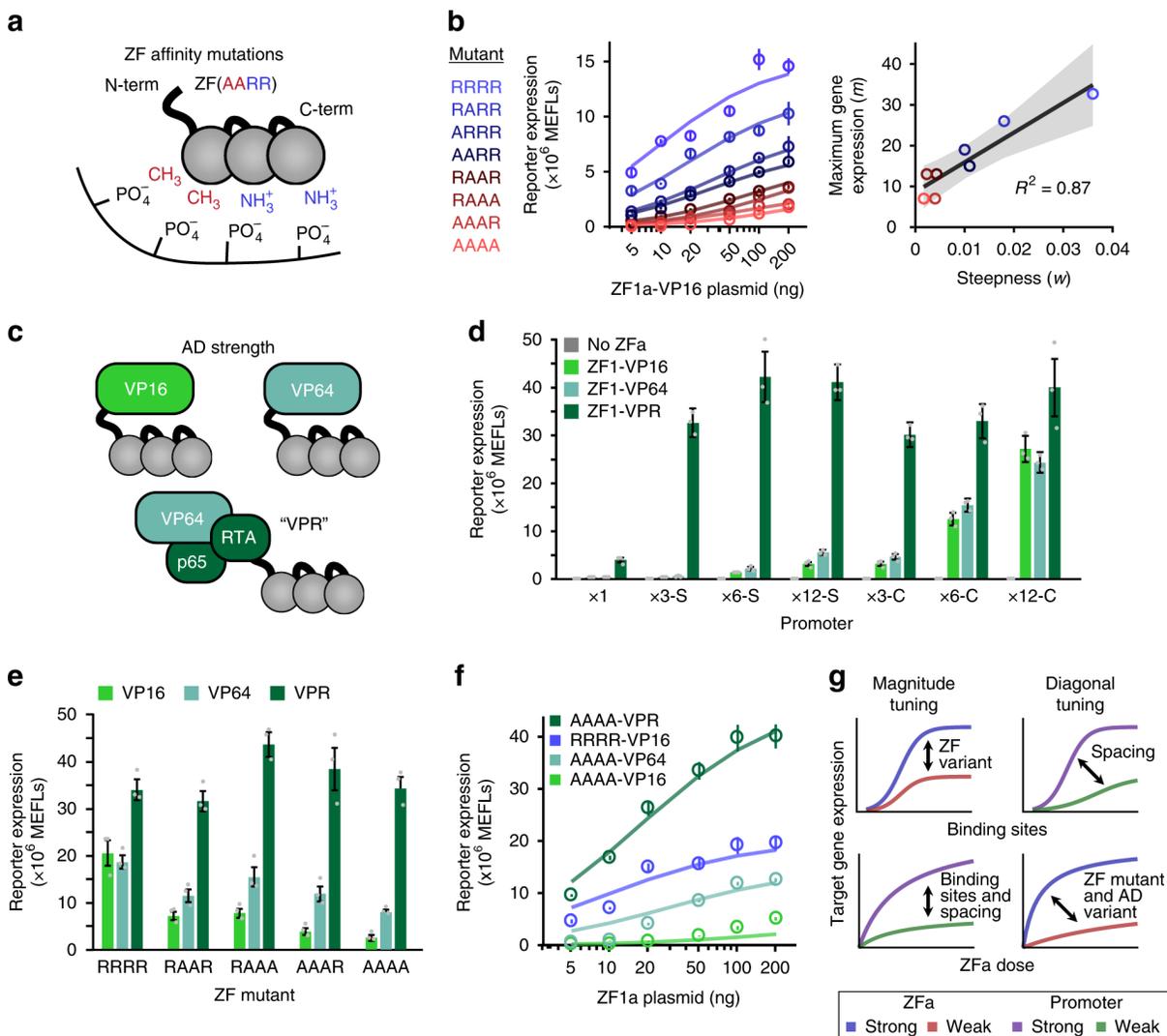


Fig 2.4 Tuning transcription through ZF mutants and AD variants. **a** The cartoon illustrates arginine-to-alanine (R-to-A) mutations in the ZF domain, which decrease the DNA-binding affinity. **b Left:** ZF mutations modulate the steepness and the maximum of the ZFa dose response profile. Circles represent experimental data and solid lines represent fitted response function models. **Right:** correlation between m and w parameters across mutants. The regression line is $m = 7.3 \times 10^2 w + 8.6$, and the shaded region is the 95% confidence interval (one-tailed permutation test $p < 0.001$). **c** The cartoon depicts evaluated ADs. **d** Effects of AD on inducible reporter expression with different promoters. Gene expression varied with the choice of promoter (two-factor ANOVA $p < 0.001$) and choice of AD ($p < 0.001$), and an interaction was observed between these two variables ($p < 0.001$). **e** Combined effects of AD variants and ZF mutations were identified. Gene expression was affected by both the ZF mutations (two-factor ANOVA $p < 0.001$) and the AD ($p < 0.001$), with an interaction seen between these two variables ($p < 0.001$). Each mutant ZFa induced more reporter expression with VP64 than with VP16 (one-tailed Welch's t -test, all $p < 0.05$) and with VPR than VP64 (one-tailed Welch's t -test, all $p < 0.01$). All VPR-ZFa induced similar expression regardless of the use of a WT or mutant ZF (Tukey's HSD test with $\alpha = 0.05$). **f** The choice of AD affects the steepness and the maximum of the dose response. Circles represent experimental data and solid lines represent fitted response function models.

g The cartoon summarizes expected trends in output gene expression that result from tuning each modular feature of the ZFa and promoters. These design choices can produce either a vertical shift or diagonal shift in response profiles with respect to the number of binding sites and the dose of ZFa. Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars depict S.E.M. Source data are provided in the **Source Data (Online¹)** file.

To explore interactions between the two TF protein engineering strategies, we investigated whether stronger ADs could enhance gene expression conferred by TFs with low-affinity ZFs. As ZF binding affinity decreased, ZFa-mediated gene expression decreased substantially with VP16, yet only moderately with VP64 and not at all with VPR (**Fig. 2.4e**). We then compared the dose response for the weakest-binding ZFa mutant (AAAA) with each AD to the VP16 ZFa bearing a wild-type (WT) ZF domain (**Fig. 2.4f**, **Supplementary Fig. A2.6e**). As AD strength increased, both m and w increased, as was observed when varying DNA binding affinity. Although the two domains of a ZFa are physically modular, since they affect the same parameters in the response function, we find that the domains are functionally intertwined. Therefore, the two TF protein engineering strategies should be considered jointly when tuning a ZFa. In summary, our observations illustrate how gene expression can be tuned through selection of physical features—ZF domain choice, mutations that affect DNA binding affinity, AD choice, and the number, spacing, and arrangement of binding sites in the promoter—and together this ensemble of designs provides a variety of realizable response profiles (**Fig. 2.4g**, **Supplementary Fig. A2.6f**).

Design of inhibitory TFs

Inhibitors comprise a key component of a versatile TF toolkit. We hypothesized that removing the AD from the ZFa would result in an inhibitor that binds DNA without inducing transcription (ZF inhibitor, ZFi) (**Fig. 2.5a**). We built a promoter with six compact binding sites for ZF1 and in which each ZF1 site overlapped with a ZF2 site to allow for pairwise testing of ZFi and ZFa with fully or partially overlapping sites (**Supplementary Fig. A2.7a**). Co-expressing ZF1a with ZF1i or ZF2i (equimolar plasmid doses) resulted in a ~50% decrease in inducible expression compared to ZF1a only, and inhibition mediated by partially overlapping ZF2i resembled that mediated by fully overlapping ZF1i (**Fig. 2.5b**, **Supplementary Fig. 2.7b**). This pattern held across various ZFi doses, and nearly complete inhibition was attained at high

ZFi doses (**Supplementary Fig. A2.8a**). We hypothesized that transcriptional inhibition could be increased by incorporating a bulky domain to sterically block ZFa from binding adjacent sites in the promoter or to block the recruitment of RNAPII or associated factors. To test this hypothesis, we fused DsRed-Express2 (abbreviated throughout as DsRed) to the ZF domain. Co-expression of ZFi-DsRed and ZFa (equimolar plasmid doses) reduced reporter expression by 90–95%, and at higher ZFi-DsRed doses the inhibition was essentially complete, even when using stronger promoters based upon the CMV minimal promoter (**Fig. 2.5b, Supplementary Figs. A2.8b-c**). Therefore, the choice of a fusion partner affords an additional design handle for substantially tuning ZFi performance characteristics.

To help understand the mechanism of ZFi-mediated transcriptional inhibition, we considered that within each cell, promoters occupy an ensemble of states that depend on the promoter architecture and the ZFa and ZFi that are present (**Fig. 2.5c**). As the relative dose of ZFi to ZFa increases, the distribution of the ensemble should shift toward states that are more inhibited; a trend towards more inhibition should also occur by increasing the relative DNA binding affinity of the ZFi versus that of the ZFa. Given our understanding of ZFa-mediated transcriptional activation, we speculated that the inhibitors should act via a dual mechanism with these properties: (i) competitive inhibition: since each site in the promoter can accommodate at most one TF, the binding of an inhibitor should prevent the binding of an activator; and (ii) decreased cooperativity: since inhibitors intersperse between activators, the spacing between activators should widen, and the effective m should resemble that of a promoter with lower cooperativity.

To explore this proposed mechanism of inhibition, we developed a model that describes the activity of ZFa and ZFi at a single-site promoter by representing physical interactions without a response function (**Supplementary Fig. A2.8d, Methods**). Simulated trends for ZFa dose responses with perturbations to DNA-binding affinity broadly agreed with experimental data (**Supplementary Fig. A2.8e, Fig. 2.4**). However, simulated trends for ZFa dose responses for varying AD strengths (at the simulated *single-site* promoter) differed qualitatively from the trends observed experimentally for a *multi-site* promoter (**Fig. 2.4f**). The difference in outcomes for the single-site and multi-site cases is consistent with our expectation that cooperative ZFa-mediated RNAPII recruitment would be observed only for the latter case (**Fig. 2.2**). Notably, the model also showed less responsiveness of reporter output to ZFi (at the simulated single-site

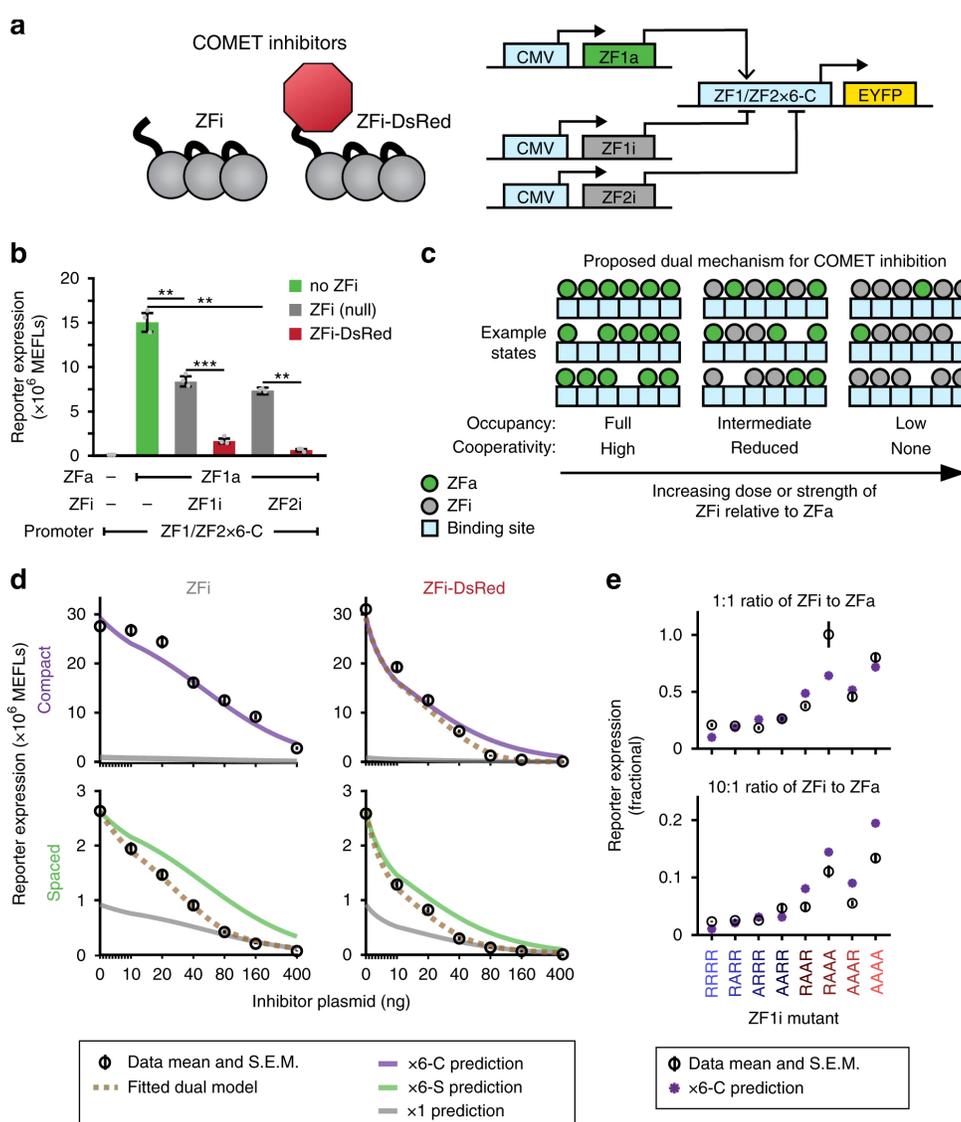


Fig 2.5 Transcriptional inhibition. **a** The schematic depicts two types of inhibitors that were evaluated. A ZF1/ZF2x6-C hybrid promoter is activated by ZF1a and inhibited by ZF1i or ZF2i. **b** ZFi and ZFi-DsRed differentially inhibit transcription (one-tailed Welch's t -test: ** $p < 0.01$, *** $p < 0.001$). **c** The cartoon summarizes the proposed conceptual model of ZFi-mediated inhibition. Within each cell, a promoter can occupy states with different configurations of ZFa and ZFi. Several example states are shown for three conditions of increasing dose or strength of inhibitor (i.e., DNA-binding affinity) relative to activator. **d** ZFi and ZFi-DsRed differ from standard competitive inhibitors. Predictions for competitive inhibition alone, for various promoter configurations, are depicted with solid lines (**Methods**). COMET inhibitors track the dotted lines, which represent fits to the dual mechanism model, except in the case of ZFi paired with x6-C, which tracks the competitive inhibition-only prediction. Each condition uses ZF1a at a dose of 40 ng. X-axes are scaled linearly from 0–10 ng and logarithmically above 10 ng. **e** Measured and predicted reporter expression were compared for a panel of ZFi mutants. Each condition uses ZF1a(RAAR) at a dose of 40 ng and the ZF1x6-C compact promoter. Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Source data are provided in the **Source Data (Online¹)** file.

promoter) than was experimentally observed for a multi-site promoter (**Supplementary Fig. A2.8f**), again suggesting that for multi-site promoters, ZFi can impair ZFa-mediated transcription by disrupting cooperative RNAPII recruitment.

To experimentally test the proposed dual mechanism of inhibition, we conducted dose responses for the ZFi and ZFi-DsRed inhibitors using the ZF1x6-S and ZF1x6-C promoters, with ZFa dose held constant (**Fig. 2.5d**). When ZFi was applied to the compact promoter, reporter expression matched the concise model for competitive inhibition alone. However, for the other three cases, observed reporter expression began to deviate with increasing doses of inhibitor, and by high doses it showed complete loss of cooperative RNAPII recruitment. The inhibitor dose at which the experiment began to deviate from the model was lower for ZFi-DsRed compared to ZFi and for spaced promoters compared to compact promoters. At intermediate doses of inhibitor, reporter expression ramped down toward single-site promoter behavior (**Fig. 2.5c** middle column, **Fig. 2.5d** dotted lines, **Methods**), and by high doses the ramp down was complete (**Fig. 2.5c** right column). The highest dose of ZFi-DsRed, used with the compact promoter, resulted in a profound 400-fold decrease in reporter expression. To further examine the case where the employed inhibitor did not disrupt cooperative RNAPII recruitment (i.e., ZFi used with the x6-C promoter), we paired a panel of ZFi varying in DNA-binding affinity with a reduced-affinity ZFa mutant (**Fig. 2.5e**). For all cases examined, ZFi-mediated inhibition was still predicted by competitive inhibition alone (**Methods**). We conclude that the compact promoter is more capable of cooperative RNAPII recruitment than is the spaced promoter, and that ZFi is a weaker inhibitor than is ZFi-DsRed, such that the dual inhibition mechanism applies to three of the four types of inhibitor-promoter pairings evaluated, and the pairing most responsive to inhibition is ZFi-DsRed with a compact promoter. Thus, the mechanism by which cooperative transcriptional machinery recruitment renders the compact promoter architecture highly activatable by a ZFa also causes such promoters to be substantially inhibited through disruption of this mechanism by a ZFi-DsRed.

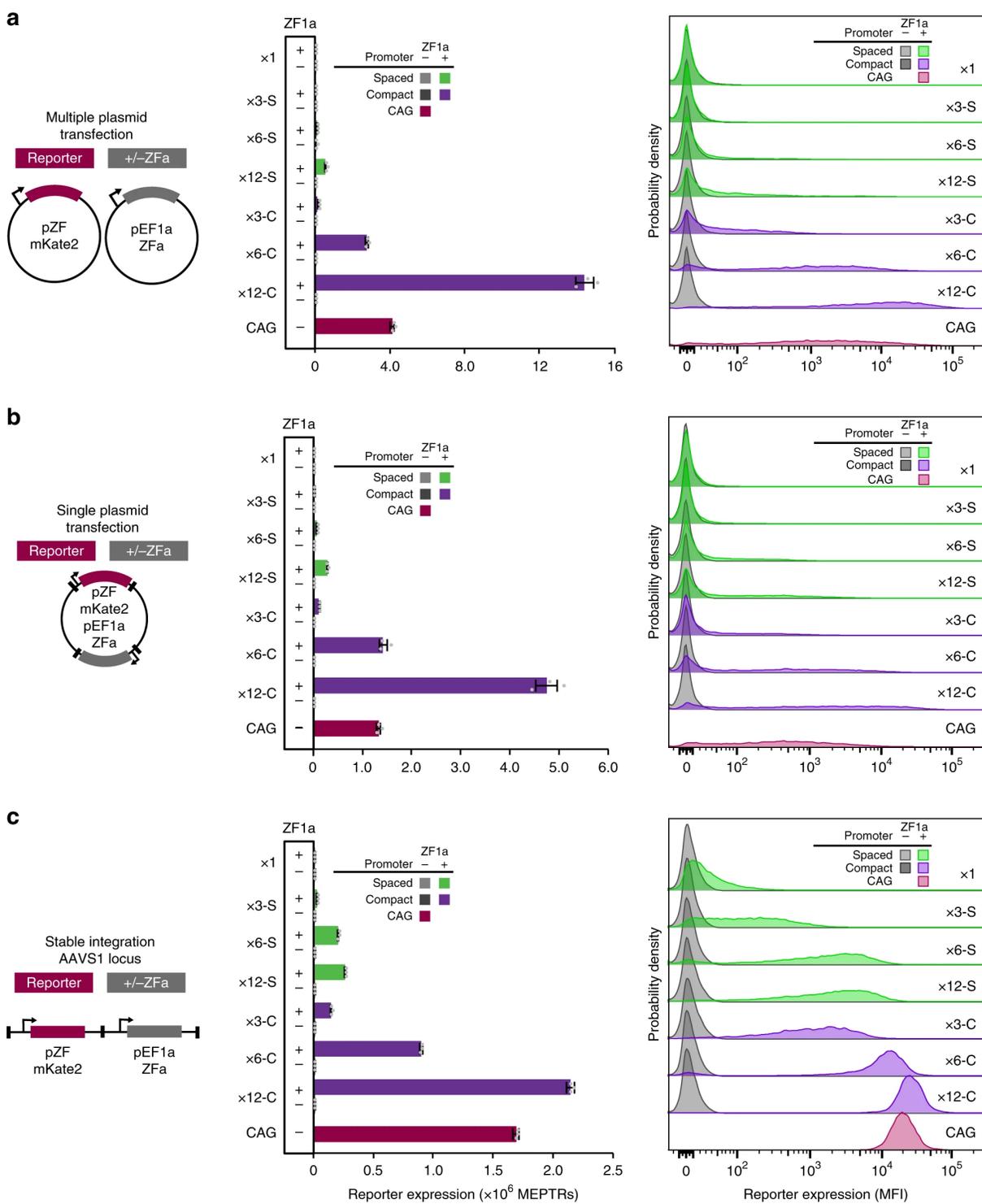


Fig 2.6 Characterization of promoter design rules in the genome. The cartoons summarize the systems used to evaluate promoter performance characteristics across three contexts: **a** multiple plasmid transient transfection, **b** single plasmid transient transfection, and **c** single-copy stable integration at a genomic safe harbor locus. The promoters included here comprise 1, 3, 6, or 12 ZF1 binding sites positioned using spaced or compact architectures upstream of the YB_TATA minimal promoter driving an mKate2 reporter gene. Constitutive EBFP2 was used as a transfection control in the transient transfection context and as a marker for genomic locus activity in the stable context. Bar graphs and histograms show reporter expression for EBFP2-expressing cells. In all contexts, ZFa-induced gene expression increased with the number of binding sites on spaced and compact promoters (ANOVA $p < 0.00001$). To profile the range of inducible expression conferred by each promoter, stable cell lines and transiently transfected cells were characterized using two distinct sets of flow cytometry settings (voltages), each of which was independently calibrated to yield comparable absolute fluorescence units (bar graphs). Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Source data are provided in the **Source Data (Online¹)** file.

Genomic integration of COMET TFs

Since some applications require stable integration of genetic programs, we investigated how COMET parts function upon stable integration into the genome, and in particular, whether COMET design rules gleaned from transient transfections might extend to performance in the genomic context. As a representative test set, we generated a panel of stable cell lines that each constitutively express a ZFa and contain one of several COMET promoters—varying the number of ZF binding sites and spacing between binding sites—that drive expression of a fluorescent reporter protein. To enable comparisons using a consistent site of genomic integration, we used site-specific Bxb1 recombinase-mediated integration into the AAVS1 safe harbor locus of HEK293FT landing pad cells.¹⁷³ In this process, COMET components were cloned into transcription unit positioning vectors (TUPVs) followed by one-step assembly into all-in-one integration vectors (IVs). The IVs used include a constitutive fluorescent protein marker and antibiotic resistance markers, a COMET promoter-driven mKate2 reporter, and either a constitutively expressed VP16-ZF1a or a blank control sequence (**Supplementary Fig. A2.9a–c**). Following gene delivery and selection (**Supplementary Fig. A2.9d**), we obtained cell lines that enable a comparison of COMET-driven gene expression in the stable genomic context (**Fig. 2.6c, Supplementary Fig. A2.10c**) to delivery by transfection of separate plasmids (**Fig. 2.6a, Supplementary Fig. A2.10a**) or transfection of all-in-one vectors (**Fig. 2.6b, Supplementary Fig. A2.10b**).

Overall, genomic COMET components drove gene expression following trends that are consistent with those observed in transient transfection: compact promoters drove more expression than did spaced

promoters and expression increased with the number of binding sites. Interestingly, for compact promoters, increasing the number of binding sites also led to more homogeneous reporter expression profiles spanning only a single order of magnitude—matching the tight distribution expected of a constitutively expressed gene in a landing pad.¹⁷³ For the strongest promoters (x6-C and x12-C), tight distributions of reporter expression contributed to high fold inductions (8,000 and 14,000, respectively, compared to corresponding reporter-only cell lines). The promoter containing a single ZF1 site, placed in a favorable position with respect to the TATA box (**Fig. 2.1d**), did confer modest but significant gene expression compared to the control promoter lacking any ZF1 site (**Supplementary Fig. A2.10d**), although the expression induced by this ZFa from a x12-C promoter was 800-fold higher (**Fig. 2.6c**). Thus, COMET TFs can drive expression from either the genome or a plasmid, and the design rules used to tune expression in transient transfections may be transferrable, at least qualitatively, to the genomic context.

Design and evaluation of small molecule-responsive TFs

Chemical inducibility is useful for conferring external and temporal control over gene expression. We designed a small molecule-responsive ZFa by fusing FBKP and FRB domains, which heterodimerize upon exposure to rapamycin¹⁹⁶, onto ZF and AD, respectively (**Fig. 2.7a**). We expected that without rapamycin, the ZF would bind DNA and not induce transcription, and that with rapamycin, FKBP and FRB would dimerize to reconstitute a functional ZFa. Indeed, rapamycin-activated ZFa (RaZFa) with each AD showed rapamycin-induced reporter expression (**Fig. 2.7b, Supplementary Fig. A2.11a**). Thus, COMET TFs can be adapted to achieve small molecule-induced gene expression.

We noted that fold-increase in reporter output was lower for the RaZFa (+/- rapamycin) than for the ZFa (+/- ZFa). For the RaZFa using VP16, this effect was attributable to low induced reporter expression. We hypothesized that if FKBP-ZF were present in excess, it might competitively inhibit the reconstituted RaZFa from binding the promoter. To investigate, we varied the doses and ratios of RaZFa components (**Fig. 2.7c, Supplementary Fig. A2.12a**). High FKBP-ZF levels diminished expression as a ZFi would, and excess VP16-FRB increased inducible expression, resulting in high fold induction when paired with lower doses of FKBP-ZF. However, VP64-based RaZFa and VPR-based RaZFa were less

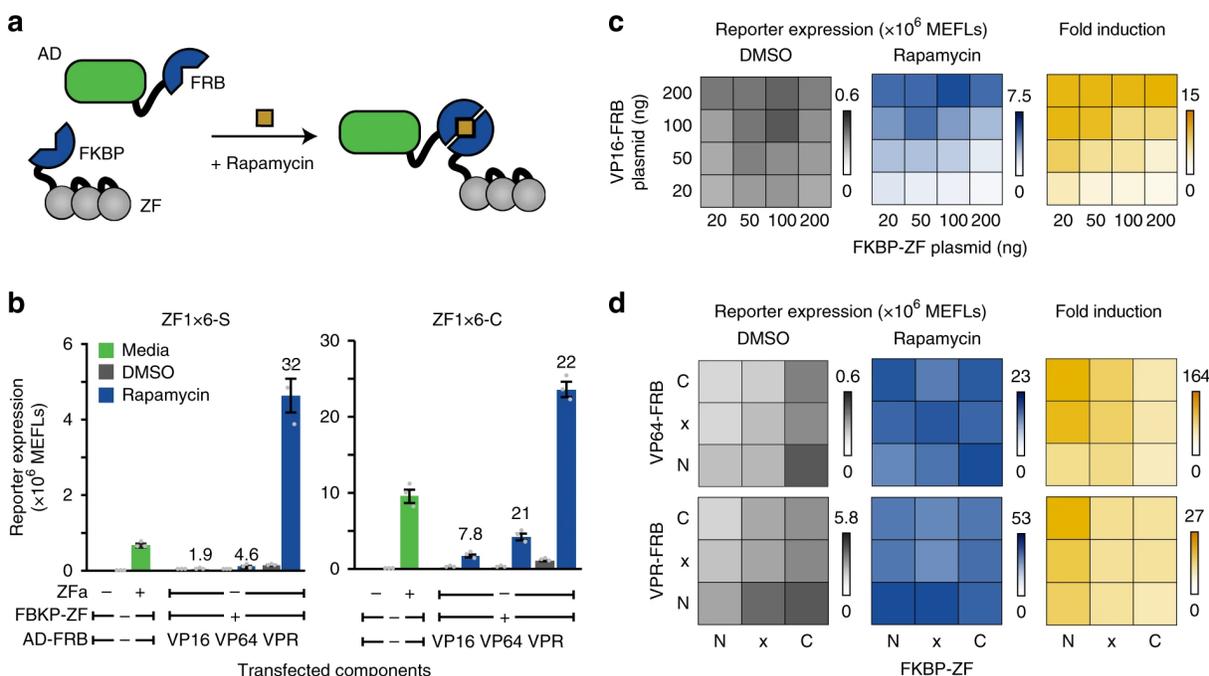


Fig 2.7 Engineering small molecule-responsive TFs. **a** The cartoon illustrates chemically-responsive control of gene expression using rapamycin-inducible ZFa (RaZFa). **b** The effects of promoter architecture and AD on RaZFa performance were evaluated. For all RaZFa on both promoters, reporter expression was significantly higher with rapamycin than DMSO (one-tailed Welch's t -test, all $p < 0.05$). Fold induction is shown above the rapamycin case for relevant conditions. **c** Gene expression in the absence of rapamycin was affected by VP16-FRB dose (two-factor ANOVA $p < 0.001$) and FKBP-ZF dose ($p < 0.001$), with no interaction between these variables ($p = 0.14$). Reporter expression after rapamycin addition was affected by VP16-FRB dose (two-factor ANOVA $p < 0.001$) and FKBP-ZF dose ($p < 0.001$) with a significant interaction between these variables ($p < 0.001$). **d** Effects of subcellular localization tags: N = nuclear, x = no localization, C = cytoplasmic. For VP64-based RaZFa, gene expression in the absence of rapamycin was affected by AD-FRB localization (two-factor ANOVA $p = 0.01$) and FKBP-ZF localization ($p < 0.001$), with no interaction between these variables ($p = 0.39$). For VP64-based RaZFa, gene expression after rapamycin addition was not affected by AD-FRB localization (two-factor ANOVA $p = 0.26$) but was affected by FKBP-ZF localization ($p = 0.02$), with an interaction ($p = 0.001$). For VPR-based RaZFa, gene expression in the absence of rapamycin was affected by AD-FRB localization (two-factor ANOVA $p < 0.001$) and FKBP-ZF localization ($p < 0.001$), with an interaction ($p = 0.03$). For VPR-based RaZFa, gene expression in the presence of rapamycin was affected by AD-FRB localization (two-factor ANOVA $p < 0.001$) but not by FKBP-ZF localization ($p = 0.29$), with no interaction ($p > 0.05$). Experiments in **c** and **d** use a ZF1x6-C promoter. Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Source data are provided in the **Source Data (Online¹)** file.

affected by component ratios (**Supplementary Fig. A2.12b,c**). Thus, it appears that the relative weakness of VP16-mediated transcriptional activation makes VP16-based RaZFa more sensitive to excess FKBP-ZF.

Since high background in the absence of rapamycin limited the fold induction for VP64-based and VPR-based RaZFa, we investigated strategies to decrease background. VPR-FRB alone promoted a very low amount of reporter expression, and this background was greater in the presence of ZF-fusion proteins, even in the absence of rapamycin (**Supplementary Fig. A2.12d**), suggesting that the ZF can bind the promoter in such a way that transient promoter-AD interactions induce some transcription. To circumvent this putative undesired mechanism, we removed the nuclear localization signal (NLS) from each RaZFa component or replaced the NLS with a nuclear export signal (NES) (**Fig. 2.7d, Supplementary Fig. A2.12e**). For both VP64-based and VPR-based RaZFa, NES tagging of AD-FRB and NLS tagging of FKBP-ZF decreased background while conferring little effect on rapamycin-induced reporter expression, such that fold induction improved. To explain why the addition of NES to FKBP-ZF increased background, we hypothesize that while low levels of nuclear FKBP-ZF are sufficient to allow AD-FRB to drive transcription from the promoter, at higher nuclear levels the FKBP-ZF can act as an inhibitor. The decrease in background associated with the NES tag on AD-FRB was not due to decreased expression (**Supplementary Fig. A2.12f**). Expression of VP64-FRB was low relative to other components, but increasing the dose of VP64-FRB plasmid—above levels used in **Supplementary Fig. A2.12b**—increased background and diminished inducible reporter expression (**Supplementary Fig. A2.12g**). Altogether, tuning these design variables led to improved rapamycin-inducible gene expression (greater fold induction) for each AD choice (**Supplementary Fig. A2.13a**), with responsiveness across several orders of magnitude of rapamycin concentration (**Supplementary Fig. A2.13b**) and yielding a useful system for chemically-induced expression.

Expanding the CID concept to engineer new chemically responsive TFs

We next sought to expand the CID concept to generate more chemically inducible TFs with COMET. During this study, we also aimed to elucidate some principles for engineering TFs. In the original COMET work, we only picked one orientation for each component—the AD was always fused to the N-

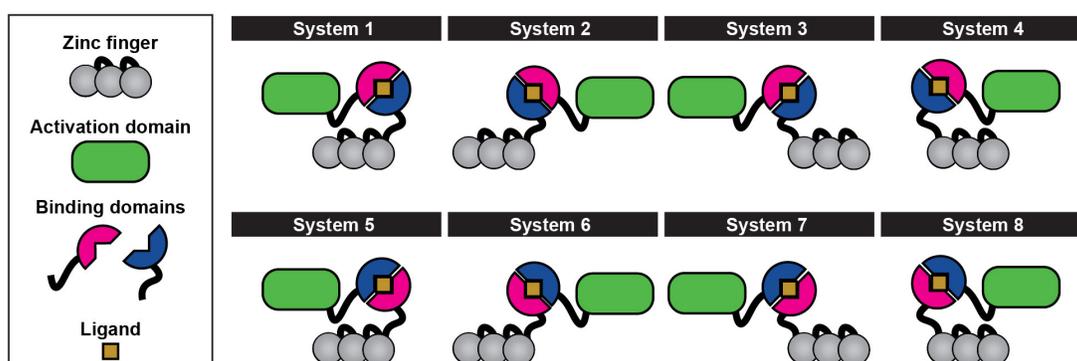


Fig 2.8. Strategies for design of CID systems. Legend at left describes each component. Systems 1-4 differ in whether the binding domains are fused to the N-termini or C-termini of each protein in the system, but keep each binding domain paired with the same component of the COMET TF (ZF or AD). Systems 5-8 represent the binding domains having been flipped between the components from systems 1-4.

terminus of FRB, and FKBP was always fused to the N-terminus of FKBP. We did not know whether all possible orientations (**Fig. 2.8**) would perform equally well, or whether differences might arise from either protein stability, non-optimal orientation of the AD relative to the promoter, or some other issue. We began by investigating eight strategies in the possible orientations and combinations thereof for building these CID TFs for rapamycin and testing them on both spaced and compact promoters (**Fig. 2.8, 2.9a**). All performed well, and we noticed several trends. Strategies 1-4, in which FKBP and the ZF were paired, yielded higher ligand-induced signal than strategies 5-8, in which the FRB and ZF were paired. Further, strategies 2 and 4, in which the FRB was fused to the N-terminus of the AD, had higher signal than strategies 1 and 3, in which the FRB was fused to the C-terminus of the AD. These differences were more pronounced on the spaced than the compact promoter.

We then evaluated the 8 strategies for four other systems (ligands and binding domains). Gibberellin is a plant hormone and brings the GID1 and GAI domains together. Abscisic acid, also a plant hormone, induces the dimerization of ABI and PYL. The small molecule drug ABT-737 induces binding of the scFV scAZ1 to a fragment of Bcl-xL. Fusicoccin, a product secreted by fungi, binds the CT52 to T14-3-3 Δ C domains. Though fusicoccin only induced weak signal in one strategy and many strategies showed non-negligible background signal, all other ligands induced significant gene expression (**Fig 2.9b-d**).

Interestingly, the other CID systems each showed several strategies that performed better than others, though no clear pattern held among the systems. For instance, with the gibberellin system (**Fig 2.9b**), strategies 5 and 7 showed lower induced signaling than the others, indicating that there was an issue with the VP64-GAI component of this system. For the ABT-737 system (**Fig 2.9d**), strategies 1-4, in which the ZF and Bcl-xL were fused, produced only low levels of signaling compared to the other strategies for this system, indicating that the ZF and Bcl-xL fusion protein is either not stable or has other issues that limit chemically induced transcription when it is used in a CID system.

Implementing Boolean logic with COMET

Finally, we explored whether COMET could be used to encode Boolean logic functions within individual promoters. Our exploration of promoter architecture (**Fig. 2.1c,e**) suggested a strategy for

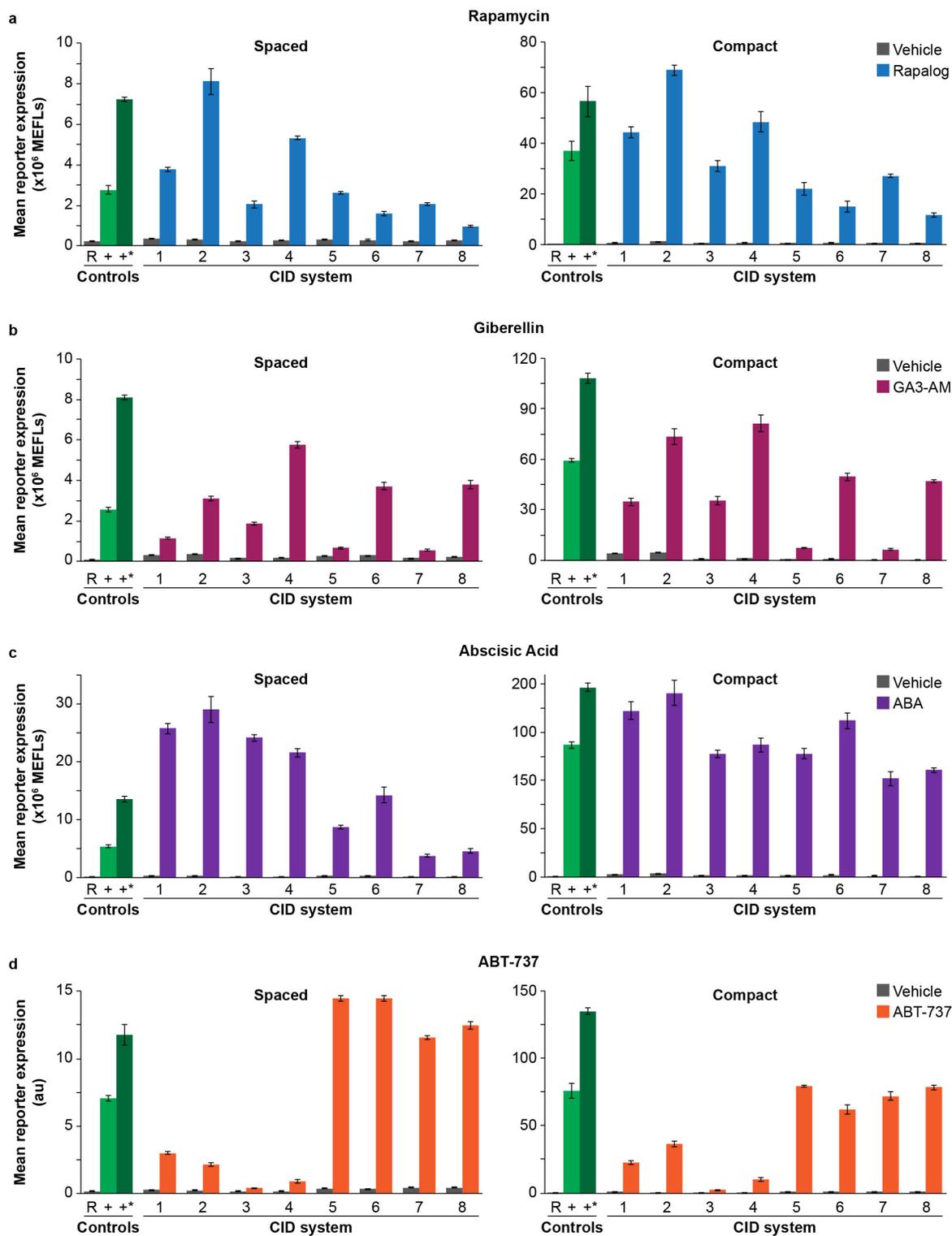


Fig. 2.9. Evaluation of new CID systems for COMET. Each system from Fig. 2.8 was built and tested for four CID systems against the ligands: a) rapamycin, b) gibberellin, c) abscisic acid, d) ABT-737. Each experiment includes comparison to cells transfected with reporter only (R), with reporter and a VP16-ZFa (+), and with reporter and a VP64-ZFa (+*). Experiments were conducted in biological triplicate, and error bars represent the S.E.M.

designing hybrid promoters with alternating sites for combinations of ZFa to implement AND logic (**Fig. 2.10a**). We hypothesized that cooperative activation on compact promoters would occur only when both species of ZFa were present, conferring AND gate behavior. Synergistic activation arising from closely arranged TF binding sites has been used to build AND gates in mammalian genetic programs¹⁸⁸, but arranging sites in alternating patterns does not necessarily guarantee the synergy required for an AND gate¹¹⁷. We tested promoters containing varying pairs of ZF2 and ZF3 sites (**Fig. 2.10b, Supplementary Fig. A2.14a**). In each case, maximal reporter expression occurred when both ZFa were present, and this expression was greater than the sum of those induced by each ZFa individually—this defines AND gate behavior. For the three-pair hybrid promoter, AND gate behavior was observed even at low ZFa levels; 5 ng of each plasmid encoding ZF2a and ZF3a together produced more reporter expression than did 200 ng of plasmid encoding either ZFa alone (**Fig. 2.10c, Supplementary Fig. A2.14b**). The steep OFF-ON transition along the perimeter of the dose response landscape is due to the effective transition between x3-S and x6-C architectures—an advantageous behavior of COMET that differs from previously reported transcriptional AND gates utilizing tTA and Gal4 (**Fig. 2.10d, Supplementary Fig. A2.14c, Methods**)¹¹⁷.

We extended this hybrid promoter strategy to generate candidate three-input AND gates for ZF1a, ZF2a, and ZF3a. A promoter with one site for each ZFa did not produce AND gate behavior (**Supplementary Fig. A2.14d**), which is consistent with the expected similarity in reporter expression for promoters recruiting two versus three ZFa (**Fig. 2.1c, e**). However, a promoter with two sites per ZFa did produce AND gate behavior; reporter expression when all three ZFa were present was higher than the sum of the levels when any two ZFa were present plus the level conferred by the third (**Fig. 2.10e**). COMET's modular features enable the composition of single-promoter AND gates.

Finally, we investigated whether inhibitors could be combined with activators to build complex logic functions using design rules elucidated in this study. As a test case, we designed a four-input logic function

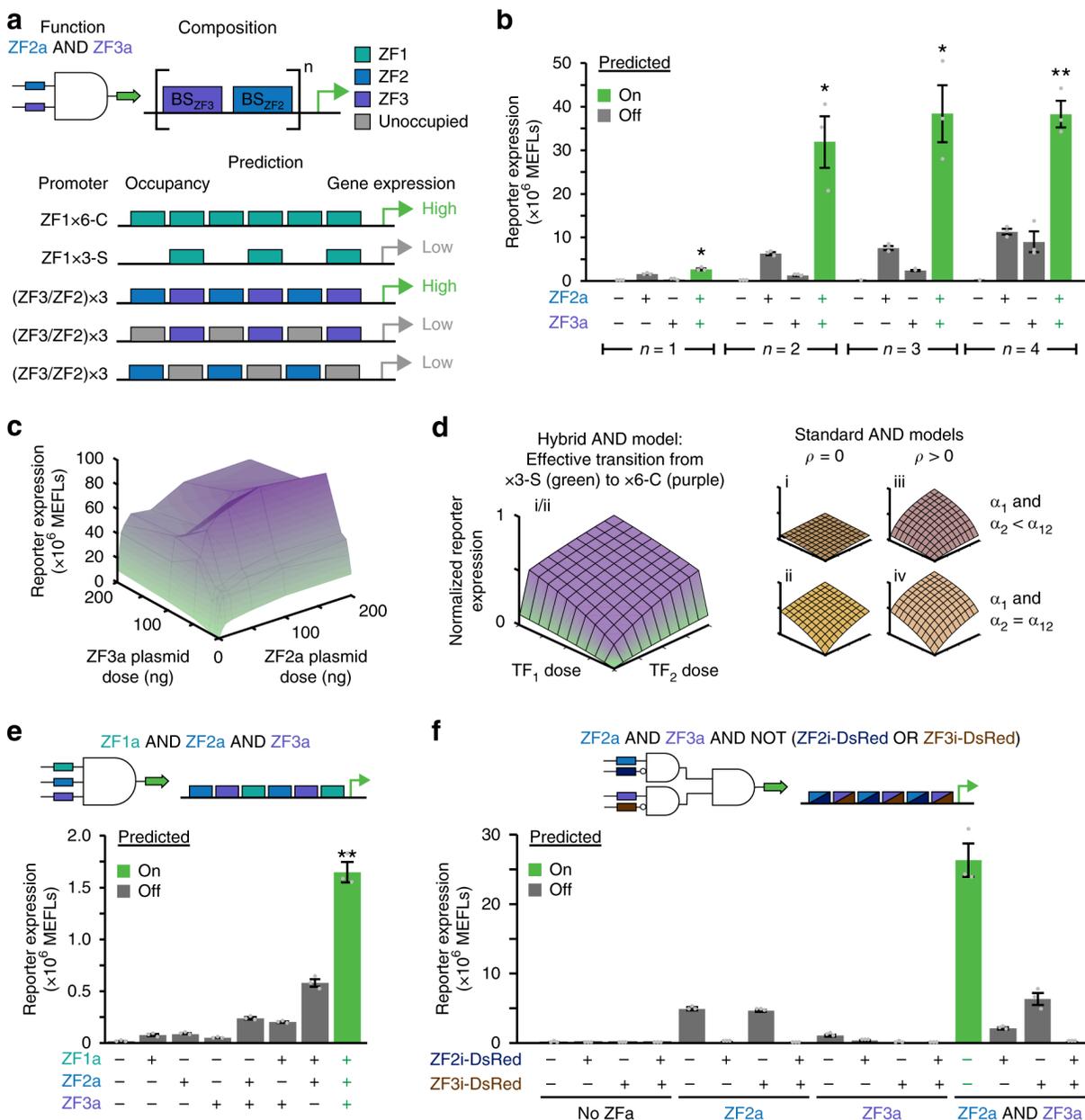


Fig 2.10 Composing Boolean logic. **a** The cartoon summarizes a strategy for single-layer, promoter-based logic gates with ZF-TFs. We hypothesized that AND gate promoters could be designed by using multiple repeats of a paired ZF3/ZF2 motif. Full occupancy of this promoter by both ZF2a and ZF3a mimics a fully occupied x6-C promoter, and partial occupancy (with either ZFa alone) mimics an x3-S promoter. Thus, there is a large increase in gene expression when the promoter is occupied by two types of ZFa compared to one type. **b** Candidate two-input AND gates were constructed using one to four repeats of paired binding sites in the promoter. AND gate behavior is considered significant if reporter expression with both ZFa is greater than the sum of reporter expression with each ZFa individually (one-tailed Welch's *t*-test: **p* < 0.05, ***p* < 0.01). **c** Two-input dose response for the AND gate with three repeats of paired binding sites. The landscape is shaded from green to purple to facilitate visualization in the z-axis direction.

d A theoretical model of COMET AND behavior is compared with other models of transcriptional AND gates; the latter vary in whether activators have multiplicative cooperativity (ρ) and whether maximum activation (α) is equivalent for TFs individually and together (**Methods**). **e** A three-input AND gate was constructed using two repeats of a triplet binding site motif. AND gate behavior is considered significant if reporter expression with all three ZFa is greater than the sum of reporter expression with each ZFa individually, and also greater than the sum from each of the three combinations with two co-expressed ZFa and the other ZFa individually (one-tailed Welch's *t*-test, $**p < 0.01$ for all four of these tests). **f** A four-input gate was constructed using the binding site arrangement shown. Experiments were conducted in biologic triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Source data are provided in the **Source Data (Online¹)** file.

that takes both activators and inhibitors as inputs (**Fig. 2.10f, Methods**). We first characterized individual interactions between activators and inhibitors and found that ZF2i-DsRed and ZF3i-DsRed were the most effective at inhibiting expression (**Supplementary Fig. A2.14e,f**). In the full genetic program, all cases produced the expected outcomes (**Fig. 2.10f, Methods**). Thus, COMET components and design principles can be employed to compose complex functions including single-layer logic.

2.4 Discussion

We anticipate that COMET will be a useful resource for building genetic programs. Currently, engineering mammalian cellular functions is slow and involves multiple iterations of the design-built-test-learn cycle. In prokaryotes, the design and construction of genetic programs has been streamlined by the development of large libraries of well-characterized and orthogonal components in concert with computational tools such as Cello¹⁴⁰. COMET similarly provides a large library of TFs and promoters with tunable features, and the characterization of these components provided a foundation for a mathematical model. We used the model to elucidate mechanisms by which the activators and inhibitors operate at promoters and fitted parameters to describe how these activities vary across the design choices examined. This integrated approach transcends the identification of general qualitative trends (e.g., increasing the number of binding sites in a promoter generally increases inducible gene expression) to yield quantitative and often mechanistic understanding as to how design choices affect TF-promoter activity. This insight could not have been deduced from prior knowledge, including biophysical intuition or even characterization of similar ZFa and promoters in yeast¹⁶⁹. Whether the design rules elucidated here ultimately enable large-scale model-driven design is an important question worthy of subsequent investigation.

COMET comprises an extensible toolkit that readily accommodates new parts. The current COMET library includes 44 activating and 12 inhibitory TFs and 83 cognate promoters. Of the 44 ZFs, 19 were ported from a toolkit originally characterized in yeast¹⁶⁹ with only minor changes in the linker between protein domains. Generating the remaining activators and inhibitors involved combining ZF domains with functional domains. This highlights COMET's modularity, in that new elements can be characterized, modeled (**Fig. 2.2, Methods**), and then utilized for customized gene regulatory functions.

Our combined experimental and computational investigation revealed properties and design rules that guide the use of COMET parts. By selecting TF-promoter combinations, one can select a magnitude of output gene expression from a range spanning three orders of magnitude. Design rules explain, at a high level, many functional consequences of choices such as ZF domain, mutations in the ZF domain that impact binding affinity, the AD, competition between activating and inhibitory TFs, and the number, spacing, and arrangement of binding sites in the promoter. COMET-mediated gene expression confers dose response landscapes that differ from those of tTA and Gal4¹¹⁷, and COMET could be better suited for applications such as building hybrid promoters. COMET is also amenable to incorporation of other functional modalities such as chemically inducible gene expression.

A key insight is that COMET promoter strength arises from cooperative recruitment of transcriptional machinery, which is an effect that varies with the spacing between binding sites. This mechanism differs from that of previously characterized ZF-TF systems in which cooperativity is directly engineered into TFs through protein-protein interaction domains such as PDZ or leucine zippers^{134, 169, 197}. While these previous strategies usefully enable tuning performance characteristics such as dose response curves, they are potentially limited by the availability, orthogonality (with respect to both synthetic and endogenous components), and geometric requirements of the protein-protein interaction domains employed. In contrast, the scalability of COMET thus far appears limited only by the availability of orthogonal ZFs; these domains can be constructed using technologies such as OPEN¹⁷⁰ as well as other methods, and this remains an active area of research.

COMET promoters' design-based cooperativity is useful. First, it confers both low background expression and high fold induction, even though these two objectives typically present a trade-off⁸⁷. Second,

it enables the implementation of logic gates that have attractive features. Unlike other previously described logic gates that require different architectures for activation and inhibition¹³⁴, a single COMET promoter can be used in either activating or inhibitory logic gates. Many gates function as predicted without extensive tuning (**Fig. 8**). These properties simplify the design process and enable integrating multiple inputs at a single promoter, ultimately decreasing the number of components required to construct genetic programs. Inhibitory COMET TFs modulate effective cooperativity to completely inhibit COMET TF-mediated transcription (**Fig. 2.5, Supplementary Fig. A2.8**). This mechanism is fast and reversible, which could be advantageous over mechanisms that employ slower KRAB-mediated chromatin repression and subsequent reactivation¹⁷⁴.

Another advantage of promoter-based cooperativity is that it should enhance the specificity with which ZFa activate target promoters. A limitation to the minimal three-finger ZF-TF strategy investigated here is that any single 9 bp ZF binding sequence might occur many times in a genome. However, the probability that two binding sites would occur at the same locus is unlikely, and the chance that three or more sites would co-occur is vanishingly small. Moreover, the potent activation reported in **Fig. 2.1** also required the ZF binding array to be proximal to a transcriptional start site, which should further boost the distinction between on-target and off-target transcription. Indeed, in a genomic context (**Fig. 2.6**), although ZF1a drove modest expression from a x1 promoter (in which the ZF binding site was placed favorably close to the TATA box), the expression from a x12-C promoter was 800-fold greater. The protein engineering design rules elucidated here also suggest that specificity could be further increased, if desired, by the choice of AD and ZF domain. For example, selection of a weaker AD could necessitate that multiple ZFs bind in a compact configuration at a promoter in order to drive transcription (**Fig. 2.4d**). Reducing the affinity with which a ZF binds DNA could also be combined with selection of a weaker AD to shift the dose response curve, such that a promoter is activated only at high concentrations of ZFa (**Fig. 2.4f**). Thus, a potential advantage of pairing weaker ZFa with multi-site promoters is the possibility of dramatically boosting the effective specificity of the ZFa for driving transcription from a target promoter. Chromatin state, and thus cell type, likely impacts the tradeoff between on-target and off-target gene regulation, and this question is worthy of exploration in the future use of COMET for specific applications.

Several COMET properties are not easily explained by simple design rules. It is not yet clear why some ZFa combinations exhibit limited crosstalk when no sequence similarity in ZF binding sites is apparent (**Fig. 2.3c**); our empirical evaluation identifies how such crosstalk can be avoided by informed selection. Also, some non-specific transcriptional activation was conferred by the most potent ADs (e.g., VPR) when ZF domains were separately expressed but not driven to physically associate (i.e., by addition of rapamycin), suggesting a noncanonical mechanism. Operationally, these phenomena present minor complications that can be circumvented by system selection and attentiveness to potential artifacts during development and design of new functions.

It will be interesting to evaluate how the trends observed here are conserved or diverge as the COMET toolkit grows and is applied to new applications. For example, we cannot predict *a priori* the magnitude of gene expression that a new ZFa will confer on its cognate promoter, nor can we predict orthogonality, but our analysis suggests that new parts may be screened, tuned, and combined following the same principles used in this study. We expect that the specific quantitative parameters determined in this study could be limited to the implementations used here, including the methods for DNA delivery and the cell type in which the characterizations were performed. However, since the fundamental mechanisms of transcription are maintained across contexts, we expect that the observed trends will extend across cell types and delivery methods. For instance, the rank order of promoter strength across the number of binding sites was conserved between transient transfection and genomic integration (**Fig. 2.6**). In practice, it is straightforward to test a focused library of COMET parts to empirically identify which combinations provide the function needed for an application, and if needed, tune system performance using strategies described in this study.

A particularly exciting prospect is using COMET with other synthetic biology technologies. For example, COMET could be integrated into synthetic receptors that utilize orthogonal TFs as outputs, such as MESA or synNotch, to generate cellular programs for sensing, processing, and responding to environmental cues^{2, 110, 115, 117}. Alternatively, COMET could be used to regulate the expression of synthetic components, such as GEMS receptors which interface with endogenous regulation¹¹². We expect that COMET will be useful for prototyping and implementing sophisticated cellular functions for both

fundamental research and cellular engineering applications.

2.6 Data and code availability

All reported experimental data are included as **Source Data (Online¹)**. The raw datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Plasmid maps for all plasmids reported in this study are provided as annotated GenBank files in **Source Data (Online¹)**. The majority of the plasmids used in this study are deposited with and distributed by Addgene, including complete and annotated GenBank files. The exceptions are plasmids pPD610, pPD611-pPD619, pPD630—these are not deposited with Addgene. Plasmids pPD610 (BxB1 Recombinase Expression Vector), pPD612 (pLink2), pPD614 (pLink4), and pPD618 (pLink8), and pPD630 (Destination Vector) were obtained through a Material Transfer Agreement with the Massachusetts Institute of Technology (MIT) and are available from Ron Weiss at MIT upon reasonable request (Weiss Lab plasmid names are given in parentheses, above). The series pPD611-pPD619 comprise linker vectors for mMoClo that have been superseded by an extended set that is deposited with Addgene; pPD611, pPD613, pPD615, pPD616, pPD617, and pPD619 are available from the corresponding author on reasonable request.

This study uses data obtained from the following Addgene plasmids, as described in more detail in **Methods**: #63798 [<https://www.addgene.org/63798/>], #14893 [<https://www.addgene.org/14893/>], #58855 [<https://www.addgene.org/58855/>], #58877 [<https://www.addgene.org/58877/>], #58876 [<https://www.addgene.org/58876/>], #78099 [<https://www.addgene.org/78099/>], #74285 [<https://www.addgene.org/74285/>], #61425 [<https://www.addgene.org/61425/>].

MATLAB code for COMET can be found at <https://github.com/leonardlab/COMET>. v1.0.1, which was used in this manuscript, is provided in a ZIP file titled **Supplementary Software (Online¹)** and on Zenodo¹⁹⁸. All code is provided under an open source license

2.7 Acknowledgements

This work was supported in part by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under Award Number F30CA203325; the National Institute of Biomedical Imaging and

Bioengineering of the NIH under Award Number 1R01EB026510; the National Institute of General Medical Sciences (NIGMS) of the NIH under Award Number T32GM008152 (to Hossein Ardehali); the Northwestern University Flow Cytometry Core Facility supported by Cancer Center Support Grant (NCI 5P30CA060553); the NUSeq Core of Northwestern's Center for Genetic Medicine, and a seed grant from the Northwestern Institute for Cellular Engineering Technologies (iCET). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors would like to thank Everett Allchin, Lauren Battaglia, Pear Dhantravan, Benjamin Leibowitz, Sophia Li, Siyuan Feng, and Viswajit Kandula for their assistance with cloning in this paper, and Amy Hong, Cameron McDonald, Justin Finkle, and Sebastian Bernasek for useful discussion on the computational model. We thank Ahmad Khalil (Boston University) for sharing plasmids encoding the ZFa from his 2012 study¹⁶⁹.

This chapter was previously published as described (**2.1 Preface**) and modified for this thesis. It is used under the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)

2.8 Competing Interests

P.S.D and J.N.L are co-inventors on patent-pending intellectual property that covers the COMET technology (PCT/US18/23989 filed with review pending; this patent covers the core transcription regulators, promoters, and their usage described in this manuscript. Applicant: Northwestern University).

Chapter 3. Elucidation and refinement of synthetic receptor mechanisms

3.1 Preface

A version of this chapter was previously published as¹⁹⁹:

Hailey I Edelstein, Patrick S Donahue, Joseph J Muldoon, Anthony K Kang, Taylor B Dolberg, Lauren M Battaglia, Everett R Allchin, Mihe Hong, Joshua N Leonard, Elucidation And Refinement Of Synthetic Receptor Mechanisms, *Synthetic Biology*, <https://doi.org/10.1093/synbio/ysaa017>

This is the fifth paper describing the MESA technology from our lab. In the first, Daringer, *et al.* established MESA as a technology for sensing extracellular ligands¹¹⁵. In the second, Schwarz *et al.* extended MESA to sense a biological ligand and rewire the response to this ligand to a therapeutic cytokine². In the third, Hartfield *et al.*, placed multiple MESA in the same cell to enable the sensing of multiple ligands and identified, via computational model, several targets for future investigations to improving the performance of the MESA system¹¹⁷. In the fourth study, Dolberg *et al.* developed a new intracellular signaling mechanism for MESA³. This study is the fifth description of the MESA technology from our lab, in which we address the areas identified in the third study.

This was a highly collaborative work. Joseph and I designed the PC kinetics study, I planned the cloning, and Joseph and Amy built and characterized the constructs. I developed the luciferase assay. I built and characterized constructs for the PCIL study. I selected native TMDs with Joseph, designed the cloning, mentored Lauren through cloning them, and conducted most of the TMD-related experiments. I selected synthetic TMDs, designed the cloning, mentored Everett through cloning some of them, cloned other plasmids myself, and carried out the synthetic TMD experiments. Hailey and Anthony built constructs for and carried out the FRET experiments; I assisted them with the design and interpretation of these studies. Taylor designed and carried out the experiments involving GFP-MESA and split-TEV. Joseph and Amy built and characterized the new MESA systems against ABA and GA3, some of which were based on my designs and cloning plans. Hailey, I, Joseph, and Joshua wrote and edited the manuscript, and all authors contributed to the editing and approved of the final version.

3.2 Abstract

Synthetic receptors are powerful tools for engineering mammalian cell-based devices. These biosensors enable cell-based therapies to perform complex tasks such as regulating therapeutic gene expression in response to sensing physiological cues. Although multiple synthetic receptor systems now exist, many aspects of receptor performance are poorly understood. In general, it would be useful to understand how receptor design choices influence performance characteristics. In this study, we examined the modular extracellular sensor architecture (MESA) and systematically evaluated previously unexamined design choices, yielding substantially improved receptors. A key finding that might extend to other receptor systems is that the choice of transmembrane domain (TMD) is important for generating high-performing receptors. To provide mechanistic insights, we adopted and employed a Förster resonance energy transfer (FRET)-based assay to elucidate how TMDs affect receptor complex formation and connected these observations to functional performance. To build further insight into these phenomena, we developed a library of new MESA receptors that sense an expanded set of ligands. Based upon these explorations, we conclude that TMDs affect signaling primarily by modulating intracellular domain geometry. Finally, to guide the design of future receptors, we propose general principles for linking design choices to biophysical mechanisms and performance characteristics.

3.3 Introduction

Engineered cell-based therapies are a promising strategy for the targeted treatment of many diseases^{200, 201, 202, 203}. Central to this approach is the use of genetically encoded sense-and-response programs, which cause the cell to enact a therapeutic function upon detection of specified cues. Designing and implementing a customized functional program generally requires integrating native and engineered cellular components, including receptors, signal transduction pathways, and genetic regulators. Developing principles and tools for doing so is an active frontier in the field of synthetic biology. This study focuses on elucidating principles that serve the broad goal of building, refining, and utilizing synthetic receptor systems.

Synthetic receptor systems can be designed either to interact with or to be independent of endogenous signaling^{204, 205}. One strategy is to couple synthetic receptor-mediated signaling to

endogenous pathways such as those involving JAK/STAT, MAPK/ERK, PLCG, PI3K/AKT, NFAT, and mediators downstream of GPCRs^{206, 207, 208, 209}. When native signal mediators are paired with downstream engineered promoters, these signals can be redirected to new transcriptional outputs^{112, 210} or cell behaviors^{211, 212}. A second strategy involves using engineered components to redirect the signaling outputs from native receptors via modified phosphorylation^{114, 213, 214, 215}, dimerization¹³⁸, or protein recruitment events²¹⁶. A third strategy is to employ *orthogonal* systems with mechanisms that are essentially independent of endogenous signaling and regulation. Examples of this approach include synthetic Notch (synNotch) receptors, which sense surface-bound ligands^{110, 111, 217, 218}, and the modular extracellular sensor architecture (MESA), which detects soluble ligands^{2, 3, 115, 117}. Both of these receptor systems can regulate either endogenous or exogenous genes directly through ligand binding-induced release of a transcription factor (TF). Orthogonal receptor systems have several potential advantages, including evasion of inadvertent activation or repression by cellular factors, the potential for use in diverse cell types, and the potential to multiplex receptors to implement sophisticated functions. As a result, orthogonal systems are of great interest for building cell-based devices, and their intrinsic modularity should facilitate extensions to new ligand inputs and functional outputs²¹⁹. Notably, these systems have not been tuned through evolution nor studied deeply in the biological literature, yet their modular structure renders them uniquely suited to iterative improvement. Therefore, there exist unique opportunities for elucidating and improving the performance characteristics of orthogonal synthetic receptors.

This study focuses on exploring and improving MESA receptors—a class of self-contained synthetic receptors that signal in a manner that is independent of endogenous cellular pathways^{2, 3, 115, 117}. Each MESA receptor comprises two transmembrane chains that dimerize upon ligand binding, triggering an intracellular proteolytic *trans*-cleavage reaction that releases an initially sequestered TF (**Fig. 3.1a**). Here, we chose to use the well-studied tetracycline-controlled transactivator (tTA) as the TF¹²⁶. Across these studies, we demonstrated that obtaining desirable performance characteristics—low ligand-independent (background) signaling and high fold induction of signaling upon ligand addition—required tuning both the absolute and relative levels at which each receptor chain is expressed. While this phenomenon is not entirely different from what one observes with native receptors and other systems, it

would be desirable to minimize this sensitivity to receptor expression level, for example to facilitate translational applications. Moreover, our computational analysis¹¹⁷ indicated that if design changes could improve the two performance characteristics noted above, this could render biosensor function robust to variation in receptor expression level. These observations motivate this investigation into refining the MESA design.

In this investigation, we build mechanistic understanding and improve the functional performance of MESA receptors by systematically exploring variations upon MESA receptor design that we hypothesized might overcome the aforementioned limitations. We identify several features that may be rationally modified to tune performance (i.e., design handles), and we employ this knowledge to generate new high-performing receptors, expand the MESA repertoire to sense new ligands, and enhance a recently developed MESA variant that employs a distinct mechanism³. Finally, we synthesize these observations to propose a framework of biophysically motivated design considerations for building novel synthetic receptors.

3.4 Materials and methods

General DNA assembly

Plasmid cloning was performed primarily using standard PCR and restriction enzyme cloning with Vent DNA Polymerase (New England Biolabs (NEB)), *Taq* DNA Polymerase (NEB), Phusion DNA Polymerase (NEB), restriction enzymes (NEB; Thermo Fisher), T4 DNA Ligase (NEB), Antarctic Phosphatase (NEB), and T4 PNK (NEB). Golden gate assembly and Gibson assembly were also utilized. The pBI-EYFP reporter was described previously (Addgene #58855)¹¹⁵. GBP2-containing and GBP7-containing source plasmids were a generous gift from Constance Cepko²²⁰. Plasmids were transformed into chemically competent TOP10 *E. coli* (Thermo Fisher), and cells were grown at 37°C.

Cloning MESA receptors

MESA receptors were cloned into pcDNA backbones to confer high expression in HEK293FT cells. These plasmid backbones are versions of the pcDNA3.1/Hygro(+) Mammalian Expression Vector (Thermo Fisher #V87020), modified by our laboratory in previous work (Addgene #138749)¹. In general, restriction

sites were chosen to facilitate modular swapping of parts via restriction enzyme cloning. A complete list of plasmids used in this study is provided in **Supplementary Data 1 (Online¹⁹⁹)**. Plasmid maps are included as GenBank files in **Supplementary Data 2 (Online¹⁹⁹)**.

Plasmid preparation

TOP10 *E. coli* were grown overnight in 100 mL LB medium with an appropriate selective antibiotic. The following morning, cells were pelleted at 3,000×g for 10 min and then resuspended in 4 mL of a solution of 25 mM Tris pH 8.0, 10 mM EDTA, 15% sucrose, and 5 mg/mL lysozyme (Fisher Scientific #AAJ6070114). Cells were lysed for 15 min by addition of 8 mL of a solution of 0.2 M NaOH and 1% SDS, followed by neutralization with 5 mL 3 M sodium acetate (pH 5.2). The precipitate was pelleted by centrifugation at 9,000×g for 20 min. Supernatant was decanted and treated with RNase A for 1 h at 37°C. 5 mL phenol chloroform was added, and the solution was mixed and then centrifuged at 7,500×g for 20 min. The aqueous layer was removed and subjected to another round of phenol chloroform extraction with 7 mL phenol chloroform (Fisher Scientific #BP1752I). The aqueous layer was then decanted and subjected to an isopropanol precipitation (41% final volume isopropanol, 10 min at room temperature—approximately 22°C, 9,000×g for 20 min), and the pellet was briefly dried and resuspended in 420 µL water. The DNA mixture was incubated on ice for at least 12 h in a solution of 6.5% PEG 20,000 and 0.4 M NaCl (1 mL final volume). DNA was precipitated by centrifugation at 21,000×g for 20 min. The pellet was washed once with ethanol, dried for several h at 37°C, and resuspended for several h in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA purity and concentration were confirmed using a Nanodrop 2000 (Thermo Fisher).

Cell culture

The HEK293FT cell line was purchased from Thermo Fisher/Life Technologies (RRID: CVCL_6911 [https://web.expasy.org/cellosaurus/CVCL_6911]) and was not further authenticated. Cells were cultured in DMEM (Gibco #31600-091) with 4.5 g/L glucose (1 g/L, Gibco #31600-091; 3.5 g/L additional, Sigma #G7021), 3.7 g/L sodium bicarbonate (Fisher Scientific #S233), 10% FBS (Gibco #16140-071), 6 mM L-glutamine (2 mM, Gibco #31600-091; 4 mM additional, Gibco #25030-081), penicillin (100 U/µL), and

streptomycin (100 µg/mL) (Gibco #15140122), in a 37°C incubator with 5% CO₂. Cells were subcultured at a 1:5 to 1:10 ratio every 2–3 d using Trypsin-EDTA (Gibco #25300-054). The HEK293FT cell line tested negative for mycoplasma with the MycoAlert Mycoplasma Detection Kit (Lonza #LT07-318).

Transfection

Transient transfection of HEK293FT cells was conducted using the calcium phosphate method. Cells were plated at a minimum density of 1.5×10^5 cells per well in a 24-well plate in 0.5 mL DMEM, supplemented as described above. For surface staining experiments, cells were plated at a minimum density of 3.0×10^5 cells per well in a 12-well plate in 1 mL DMEM. For microscopy experiments, glass coverslips placed in 6-well plates were coated in a 0.1 mg/mL solution of poly-L-lysine hydrobromide (Sigma #P6282) for 5 min and left to dry overnight before plating 6×10^5 cells per well in 2 mL DMEM. After at least 6 h, by which time the cells had adhered to the plate, the cells were transfected. For transfection, plasmids were mixed in H₂O, and 2 M CaCl₂ was added to a final concentration of 0.3 M CaCl₂. The exact DNA amounts added to the mix per well and plasmid details for each experiment are listed in **Supplementary Data 3 (Online¹⁹⁹)**. This mixture was added dropwise to an equal-volume solution of 2× HEPES-Buffered Saline (280 mM NaCl, 0.5 M HEPES, 1.5 mM Na₂HPO₄) and gently pipetted up and down four times. After 2.5–4 min, the solution was mixed vigorously by pipetting ten times. 100 µL of this mixture was added dropwise to the plated cells in 24-well plates, 200 µL was added to the plated cells in 12-well plates, or 400 µL was added to the plated cells in 6-well plates, and the plates were gently swirled. The next morning, the medium was aspirated and replaced with fresh medium. In some assays, fresh medium contained ligand and/or vehicle as described in **Supplementary Table A3.1** and indicated in figure legends.

Typically at 36–48 h post-transfection and at least 24 h post-media change, cells were harvested. As noted in figure captions, some experiments involved treatment with ligand or vehicle at later time points. In these cases, medium was still replaced the morning after transfection, and ligand diluted in serum-free DMEM was added as indicated. Cells were harvested for flow cytometry using FACS buffer (PBS pH 7.4, 2–5 mM EDTA, 0.1% BSA) or using 0.05% Trypsin-EDTA (Thermo Fisher Scientific #25300120) for 5 min followed by quenching with medium. The resulting cell solution was added to at least 2 volumes of FACS

buffer. Cells were spun at 150×g for 5 min, supernatant was decanted, and fresh FACS buffer was added.

Luciferase assays

Some functional assays used a luciferase readout (Dual-Glo, Promega #E2940) measured by a microplate reader (BioTek Synergy H1). Cells were transfected in biological triplicate with MESA receptor chains, an inducible Firefly luciferase, a constitutive *Renilla* luciferase, an inducible EYFP, a constitutive EBFP2, and an empty vector (as needed to maintain equal total plasmid mass across conditions). The day after transfection, vehicle and ligand treatments were applied during the medium change. Two days after transfection, EBFP2 and EYFP served as confirmatory microscopy readouts, and the two luciferase signals were quantified. Dual-Glo kit reagents were stored and prepared per the manufacturer-supplied instructions and equilibrated to room temperature; all steps were carried out at room temperature. At the time of assaying, medium was aspirated from cell cultures, and cells were washed with PBS. Passive lysis buffer stock (5X, Promega #E1941) was diluted in water, and the diluted buffer (120 μ L, 1X) was added to each well. Plates were placed on a rocker for 15 min, after which the lysates were transferred into a 1.5 mL Eppendorf tube. Each tube was vortexed (15 s), and tubes were centrifuged (15,000×g, 30 s). 30 μ L of supernatant was pipetted from each tube into a well of a 96-well plate (Thermo Fisher Scientific #655906). Luciferase reagent (30 μ L, a volume equal to that of cell lysate) was pipetted into each well and mixed. Plates were incubated by rocking in the dark for 15 min.

Data were collected using the microplate reader's luminescence fiber. For Firefly luciferase signal acquisition, autogain was set to scale the brightest wells to a value of 5,000,000 RLU. Integration time was set to 1 s and three technical replicate measurements per well were obtained. Stop & Glo reagent (30 μ L, equal to that of the previous reagent) was added to each well, and plates were incubated by rocking in the dark for 15 min. For *Renilla* luciferase signal, the gain was set to 200, the integration time was set to 5 s, and three technical replicate measurements per well were obtained. For each well, the mean of the three technical replicates was calculated for each of the two luciferase readouts. Subsequently, the autoluminescence signal (calculated from the mean of the three vector-only transfected wells) was subtracted from each condition to background-normalize for each of the two luciferase readouts,

respectively. Then for each biological replicate, Firefly luciferase signal was divided by *Renilla* luciferase signal. Quotients were linearly scaled such that the mean of the quotients for a condition transfected with only the reporter was equal to 1 a.u. Error was propagated accordingly.

Western blotting

For Western blotting, HEK293FT cells were plated at 7.5×10^5 cells per well in 2 mL DMEM in 6-well plates and transfected as above, using 400 μ L transfection reagent per well (the reaction scales with the volume of medium). At 36–48 h after transfection, cells were lysed with 500 μ L RIPA (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate) with protease inhibitor cocktail (Pierce/Thermo Fisher #A32953) and incubated on ice for 30 min. Lysate was cleared by centrifugation at $14,000 \times g$ for 20 min at 4°C, and supernatant was harvested. A BCA assay was performed to determine protein concentration, and after a 10 min incubation in Lamelli buffer (final concentration 60 mM Tris-HCl pH 6.8, 10% glycerol, 2% sodium dodecyl sulfate, 100 mM dithiothreitol, and 0.01% bromophenol blue) at 70°C (or 100°C for experiments that involved multiple co-transfected MESA receptors), protein (0.5 μ g for experiments that were imaged with film, 10 to 25 μ g for experiments that were imaged digitally) was loaded onto a 4–15% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad) and run either at 50 V for 10 min followed by 100 V for at least 1 h, or at 100 V for at least 1 h. Wet transfer was performed onto an Immuno-Blot PVDF membrane (Bio-Rad) for 45 min at 100 V. Ponceau-S staining was used to confirm protein transfer. Membranes were blocked for 30 min with 3% milk in Tris-buffered saline pH 8.0 (TBS pH 8.0: 50 mM Tris, 138 mM NaCl, 2.7 mM KCl, HCl to pH 8.0), washed once with TBS pH 8.0 for 5 min, and incubated for 1 h at room temperature or overnight at 4°C in primary antibody (Mouse-anti-FLAG M2 [Sigma #F1804, RRID: AB_262044 [http://antibodyregistry.org/AB_262044])), diluted 1:1000 in 3% milk in TBS pH 8.0. Primary antibody solution was decanted, and the membrane was washed once with TBS pH 8.0 and then twice with TBS pH 8.0 with 0.05% Tween, for 5 min each. Secondary antibody (HRP-anti-Mouse [CST 7076, RRID: AB_330924 [http://antibodyregistry.org/AB_330924])), diluted 1:3000 in 5% milk in TBST pH 7.6 (TBST pH 7.6: 50 mM Tris, 150 mM NaCl, HCl to pH 7.6, 0.1% Tween), was applied for 1 h at room temperature, and the membrane was washed three times for 5 min each time with

TBST pH 7.6. The membrane was incubated with Clarity Western ECL Substrate (Bio-Rad) for 5 min, and then either exposed to film, which was developed and scanned, or digitally captured using an Azure c280 (Azure Biosystems). Images were cropped with Photoshop CC (Adobe). No other image processing was employed. Original western blot images are provided in **Source Data 2 (Online²²¹)**, which can be found on Zenodo (<https://doi.org/10.5281/zenodo.4026861>)²²¹.

Expression normalization of MESA chains

Scanned Western blot images were imported into ImageJ and analyzed using the analyze gel feature. The intensity for each MESA chain band was quantified and reported as the percent of the total signal from all MESA bands on the blot; the same calculation was performed for all of the NanoLuciferase bands. This analysis was repeated for multiple images captured for each blot (including a range of exposure times to minimize bias) with non-detectable and saturated bands excluded from the analysis. The calculated intensity was averaged across all exposure times, and then MESA intensity (expression level) was divided by the NanoLuciferase intensity (expression level). This value was compared to the intensity calculated for the CD28-TMD rapamycin-binding TC sample, which was included on each blot as an internal cross-comparison control. This calculated ratio was used to adjust the doses of plasmids used for transfections in subsequent rounds of experiments, which were again evaluated by Western blots. The set of plasmid doses used in each round is in **Supplementary Table A3.2**.

Immunohistochemistry

For surface staining, HEK293FT cells were plated at 3×10^5 cells per well in 1 mL DMEM in 12-well plates and transfected as described above, using 200 μ L transfection reagent per well. At 36–48 h after transfection, cells were harvested with 500 μ L FACS buffer and spun at $150 \times g$ at 4°C for 5 min. For the experiment in which multiple harvest methods were compared, some samples were harvested using 0.05% Trypsin-EDTA (3 min or 10 min incubation, 37°C), which were then quenched with medium and added to two volumes of FACS buffer. Supernatant was decanted, and 50 μ L fresh FACS buffer and 10 μ L human IgG (Human IgG Isotype Control, ThermoFisher Scientific #02-7102, RRID: AB_2532958, stock

concentration 1 mg/mL) was added. Cells were incubated in this mixture at 4°C for 5 min. 5 µL FLAG tag antibody (Anti-DDDDK-PE, Abcam ab72469, RRID: AB_1268475, or Anti-DDDDK-APC, Abcam ab72569, RRID: AB_1310127) was added at a concentration of 0.5 µg per sample and cells incubated at 4°C for 30 min. Following incubation, 1 mL FACS buffer was added, cells were spun at 150×g at 4°C for 5 min, and supernatant was decanted. This wash step was repeated two more times to total three washes. After decanting supernatant in the final wash, 1–3 drops of FACS buffer were added.

Analytical flow cytometry

Flow cytometry was run on a BD LSR Fortessa Special Order Research Product (Robert H. Lurie Cancer Center Flow Cytometry Core). Lasers and filter sets used for data acquisition are listed in **Supplementary Table A3.3** (for experiments involving reporter expression), **Supplementary Table A3.4** (for experiments quantifying receptor expression on the cell surface), and **Table 3.1** (for experiments involving FRET). Approximately 2,000–3,000 single transfected cells were analyzed per sample, using a single transfection control or, when available, multiple transfected fluorophores for gating (e.g., mCerulean+/mVenus+ cells were classified as the subset of transfected cells of interest for FRET experiments).

Samples were analyzed using FlowJo v10 software (FlowJo, LLC). Fluorescence data were compensated for spectral bleed-through. Additionally, spectral bleed-through compensation in FRET experiments included compensation of the fluorescence of either mCerulean or mVenus out of the AmCyan channel. The HEK293FT cell population was identified by SSC-A vs. FSC-A gating, and singlets were identified by FSC-A vs. FSC-H gating. To distinguish transfected from non-transfected cells, a control sample of cells was generated by transfecting cells with a mass of pcDNA (empty vector) equivalent to the mass of DNA used in other samples in the experiment. For the single-cell subpopulation of the pcDNA-only sample, a gate was made to identify cells that were positive for the constitutive fluorescent protein used as a transfection control in other samples, such that the gate included no more than 1% of the non-fluorescent cells.

Fluorophore	Channel name	Excitation laser	Filter set
mCerulean	Pacific Blue	405 nm (Violet)	450/50
mVenus	FITC	488 nm (Blue)	505LP, 530/30
mVenus FRET ¹	AmCyan	405 nm (Violet)	505LP, 530/30
miRFP670	APC	640 nm (Red)	670/30

Table 3.1 Flow cytometry setup for FRET analysis

Quantification of reporter output

MESA signaling was quantified by measuring the expression of a downstream fluorescent reporter protein (**Fig. 3.1a**). For most experiments, the reporter protein was EYFP. For experiments in which the ligand was sGFP, DsRed2 was used as the reporter (instead of EYFP) to avoid spectral overlap with sGFP. The mean fluorescence intensity (MFI) for each relevant channel (as defined in **Supplementary Table A3.3**) of the single-cell transfected population was calculated and exported for further analysis. To calculate reporter expression, MFI in the FITC channel (for EYFP reporter) or PE-Texas Red channel (for DsRed2 reporter) was averaged across three biological replicates. From this number, cell autofluorescence was subtracted. To calculate cell autofluorescence, in each experiment, a control group of cells transfected with DNA encoding the fluorescent protein transfection control and pcDNA was used. The background-subtracted MFI was converted to Molecules of Equivalent Fluorescein (MEFLs) or Molecules of Equivalent PE-Texas Red (MEPTRs); to determine conversion factors for MFI to MEFLs and for MFI to MEPTRs, Rainbow Calibration Particles (Spherotech #RCP-30-5) or UltraRainbow Calibration Particles (Spherotech #URCP-100-2H) were run with each flow cytometry experiment. These reagents contain six (RCP) or nine (URCP) subpopulations of beads, each with a known number of various fluorophores. The total bead population was identified by FSC-A vs. SSC-A gating, and bead subpopulations were identified through two fluorescent channels. MEFL and MEPTR values corresponding to each subpopulation were supplied by the manufacturer. A calibration curve was generated for the experimentally determined MFI vs. the manufacturer-supplied MEFLs or MEPTRs, and a linear regression was performed with the constraint that 0 MFI equals 0 MEFLs or MEPTRs. The slope from the regression was used as the conversion factor, and error was propagated. Fold differences were calculated by dividing reporter expression with ligand treatment by the reporter expression without ligand treatment (vehicle). Standard error was propagated through all calculations.

Quantification of FRET by flow cytometry

Detailed laser and filter setups for FRET data collection are listed in **Table 3.1**. Briefly, the donor fluorescence intensity was quantified using the Pacific Blue channel ($\lambda_{\text{ex}} = 405\text{nm}$, $\lambda_{\text{em}} = 450/50\text{nm}$), the

acceptor fluorescence intensity was quantified using the FITC channel ($\lambda_{\text{ex}} = 488\text{nm}$, $\lambda_{\text{em}} = 530/30\text{nm}$), and the FRET fluorescence intensity was quantified using the AmCyan channel ($\lambda_{\text{ex}} = 405\text{nm}$, $\lambda_{\text{em}} = 530/30\text{nm}$). The mCerulean+/mVenus+ population was distinguished from samples transfected with the transfection control and pcDNA only, mCerulean only, and mVenus only. This gate was drawn such that less than 1% of the listed single-color samples were included. The normalized FRET (NFRET) parameter²²² was defined in the FlowJo workspace by dividing the compensated fluorescence intensity (FI) of a cell in the AmCyan channel by the square root of the product of the compensated FI of that cell in the Pacific Blue and FITC channels, as described by equation (1):

$$NFRET = \frac{\text{Comp AmCyan FI}}{\sqrt{\text{Comp Pacific Blue FI} \times \text{Comp FITC FI}}} \quad (1)$$

Average NFRET metrics were calculated for controls included in each experiment, including a negative control (cytosolic mCerulean and mVenus co-transfected) and positive control (membrane-tethered mCerulean-mVenus fusion protein). A calibrated NFRET parameter was defined based on the negative and positive FRET controls such that the NFRET of all samples is scaled linearly between these controls, as described by equation (2):

$$NFRET_{\text{Calibrated}} = \frac{(NFRET - NFRET_{\text{negative control}})}{(NFRET_{\text{positive control}} - NFRET_{\text{negative control}})} \quad (2)$$

The calibrated NFRET metrics were exported, along with each mean FI in the relevant channels of the single-cell, transfected, mCerulean+/mVenus+ population for further analysis. In this study, all plotted NFRET metrics are calibrated values, and all FRET FI are compensated. Fold differences were calculated by dividing NFRET in the presence of ligand by NFRET in the absence of ligand (presence of vehicle only). Standard error was propagated through all calculations.

Confocal microscopy

Confocal microscopy was performed on a Leica SP5 II laser scanning confocal microscope (Northwestern Chemistry of Life Processes Institute, Biological Imaging Facility) with a 100× (1.44 NA) oil-immersion objective. Coverslips were removed from media and mounted on a glass slide immediately before imaging. Ten fields of view were captured for each sample. Fields of view were selected using the

brightfield channel to identify areas with adherent cells, and the z-axis was adjusted to focus on the central plane of most cells within the field of view before exposing the cells to lasers. All images were captured at a 512×512 image resolution with a scanning speed of 400 Hz. Excitation and emission settings were selected depending on which fluorophore (donor (D) or acceptor (A)) was excited and which emission was captured. The excitation and emission setup is described in **Supplementary Table A3.5**. Hybrid detector voltage settings were held constant for all samples within each experiment.

Quantification of FRET by image processing

Images were exported as stacks and separated into single-channel images using Fiji software. A custom image processing script was produced in MATLAB to import single-channel images into matrices. The script used in this study can be found at <https://github.com/leonardlab/MESAFRET> and on Zenodo (<https://doi.org/10.5281/zenodo.4026851>)²²³. All code is provided under an MIT open source license. To describe channels throughout this section, DD indicates that the donor was excited and emission from the donor was captured, AA indicates that the acceptor was excited and emission from the acceptor was captured, and DA indicates that the donor was excited and emission from the acceptor was captured. Empty vector-only (pcDNA) transfected samples were used to identify the upper limit of autofluorescence in each channel (DD, AA, DA), and thresholds were defined as the 99.9th percentile of fluorescence of pixels in the vector-only samples in each channel. Pixels below this threshold were set to an intensity of zero. Saturated pixels were also removed from the matrices. Next, respective donor-only and acceptor-only controls (single-receptor transfections) were used to calculate spectral bleed-through parameters. These parameters, as defined in equations (3–6), were calculated by averaging across all pixels in ten fields of view, excluding pixels with infinite or undefined values. Intensities (I) are subscripted with DD, AA, or DA to indicate the excitation and emission conditions, and with (D) or (A) to indicate that this parameter was calculated using donor-only or acceptor-only samples, respectively.

$$a = \frac{I_{DA(A)}}{I_{AA(A)}} \quad (3)$$

$$b = \frac{I_{DD(A)}}{I_{AA(A)}} \quad (4)$$

$$c = \frac{I_{AA(D)}}{I_{DD(D)}} \quad (5)$$

$$d = \frac{I_{DA(D)}}{I_{DD(D)}} \quad (6)$$

The parameters calculated for each pair of receptors were used to subtract the contribution of donor and acceptor fluorescence from the FRET channel fluorescence on a pixel-by-pixel basis as described in equation (7) and as reported previously²²⁴. This step produced a corrected FRET fluorescence intensity (F_c):

$$F_c = I_{DA} - a(I_{AA} - cI_{DD}) - d(I_{DD} - bI_{AA}) \quad (7)$$

NFRET is calculated by normalizing F_c to the square root of the product of the donor and acceptor fluorescence intensities on a pixel-by-pixel basis as described in equation (8) and as reported previously²²².

$$NFRET = \frac{F_c}{\sqrt{I_{AA} \times I_{DD}}} \quad (8)$$

NFRET values were exported across pixels within each field of view and across all fields of view, then averaged to generate a mean NFRET metric for the entire sample. Corrected FRET intensity and NFRET matrices were plotted and visualized as processed images.

Ectodomain distance estimations

Published crystal structures for ligand-bound ectodomains (ECDs) were analyzed using Chimera (FKBP/FRB PDB: 3FAP²²⁵, GID1/DELLA PDB: 2ZSH²²⁶, ABI1/PYL1 PDB: 3KDJ²²⁷, G6-311 VEGF scFv PDB: 2FJG²²⁸, B20-4.1 VEGF scFv PDB: 2FJH²²⁸). Residues that correspond to the C-termini used in our experimental receptor system were identified, and inter-termini distances were measured.

Statistical analyses

Unless otherwise stated, three independent biological replicates were evaluated for each condition. The data shown reflect the mean across these biological replicates of the MFI of approximately 2,000–3,000 single, transfected cells or the mean NFRET or NFRET fold difference of approximately 2,000–3,000 single, transfected, mCerulean+/mVenus+ cells. Error bars represent the standard error of the mean (S.E.M.).

ANOVA tests and Tukey's HSD tests were performed using RStudio. Tukey's HSD tests were performed with $\alpha = 0.05$. Pairwise comparisons were made using a two-tailed Welch's *t*-test, which is a

version of Student's *t*-test in which the variance between samples is treated as not necessarily equal. Two-tailed Welch's *t*-tests were performed in GraphPad. To decrease the false discovery rate, the Benjamini-Hochberg procedure was applied to each set of tests per figure panel; in all tests, after the Benjamini-Hochberg procedure, the null hypothesis was rejected for *p*-values < 0.05. The outcomes for each statistical test are provided in the figure captions, and additional details for some panels are in referenced supplementary notes. **Supplementary Note A3.1** includes the outcomes for one-way ANOVA tests followed by Tukey's HSD tests for **Fig. 3.1c** and **Supplementary Fig. A3.3e**. **Supplementary Note A3.2** includes the outcomes for two-way ANOVA tests followed by Tukey's HSD tests for **Fig. 3.4a,b** and **Fig. 3.6a–f**. **Supplementary Note A3.3** includes the outcomes for three-way ANOVA tests followed by Tukey's HSD tests for **Fig. 3.4c**, **Fig. 3.6a–f**, and **Supplementary Fig. 28b**.

3.5 Results

Protease tuning reduces background

We initially focused on the goal of decreasing MESA receptor background signaling by investigating two strategies that we hypothesized could modulate the kinetics of proteolytic *trans*-cleavage. Our first strategy was motivated by extending a published observation. In our initial development of MESA¹¹⁵, we used prior biochemical analyses of the tobacco etch virus protease (TEVp)^{4, 229, 230, 231} to vary cleavage kinetics by mutating the amino acid residue that immediately follows the cleavage site (the P1' position) within the protease recognition sequence (PRS: ENLYFQX, where X is the P1' position) on the target chain (TC). This search enabled the identification of a kinetic regime in which dimerization-inducible signaling was feasible. However, in the initial study, we did not explore all kinetic regimes that are accessible using known TEVp biochemistry. Here, we investigated whether TEVp kinetics could be subtly tuned to preferentially reduce background signaling more so than ligand-induced signaling. We introduced mutations to the TEVp on the protease chain (PC) that reduce cleavage kinetics to varying degrees^{4, 229, 230, 231}, and we subsequently combined these with mutations to the P1' position of the PRS on the TC. We found many combinations of TC and PC variants that exhibited lower background signaling compared to the base case, as desired. However, the fold difference in reporter expression (the ratio of reporter expressed with vs.

without ligand) did not improve; these receptor variants exhibited a decrease in ligand-induced signaling that was proportional to or greater than the decrease in background signaling, and therefore we chose not to pursue the strategy of tuning TEVp kinetics (alone) to improve receptor performance.

We next explored a second strategy motivated by the observation that TEVp naturally includes a C-terminal auto-inhibitory peptide (AIP) that is often omitted from biochemical tools that use TEVp (including MESA). Crystallographic evidence suggests that the AIP can reside in the TEVp active site²³¹, and since the PRS and AIP are similar in sequence, it is possible that these peptides compete for binding to the active site of TEVp. We hypothesized that by placing variants of the PRS or AIP on the TEVp C-terminus, these peptides might reduce background signaling by reversibly occupying the TEVp active site such that TC-PC *trans*-cleavage is inhibited during transient diffusive encounters (in the absence of ligand), but cleavage would eventually occur during sustained ligand binding-induced chain dimerization (**Fig. 3.1b**). Appending the full AIP to TEVp decreased both background and ligand-induced signaling, and a similar outcome was observed when this effect was combined with the previously evaluated mutations in TEVp or the PRS. Appending most AIP and PRS variants also produced this pattern (**Fig. 3.1c**), but four appended peptides (ELVYSQ, ELVYSQM, ELVYSQA, ELVYSQK) slightly increased fold difference compared to the base case. Altogether, we conclude that adding active site-occupying peptides can modestly reduce background and improve fold inducibility, but other strategies are needed to more substantially improve receptor performance.

Protease chain expression can be stabilized by linker selection

In prior work, we observed that PC surface expression was often lower than TC surface expression^{2, 115}, and we hypothesized that some aspects of PC design might render this chain less stable. Western blot analysis showed expression of the full-length PC as well as a smaller fragment, the size of which was consistent with juxtamembrane cleavage (**Fig. 3.1d,e**). Since this pattern was observed even when TEVp was mutated to be catalytically inactive via D81N substitution⁴ (**Fig. 3.1d,e**), the observed cleavage can be attributed to endogenous cellular processes, rather than the catalytic activity of TEVp. Cleavage of the PC is potentially problematic because the residual membrane-tethered ectodomain (ECD)

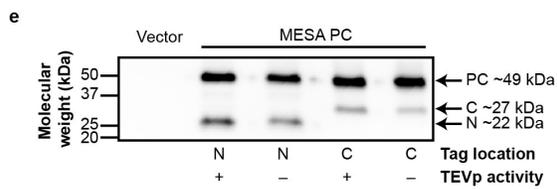
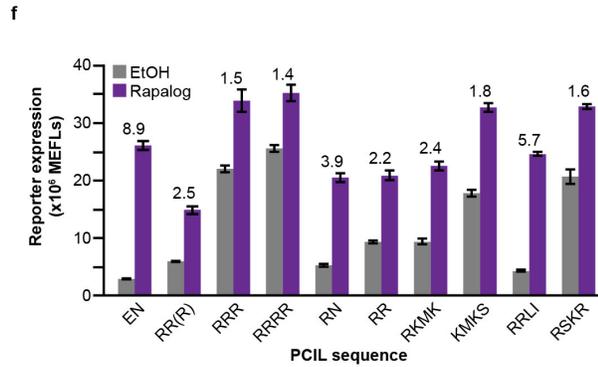
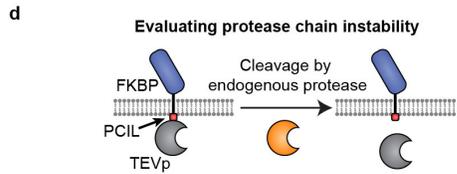
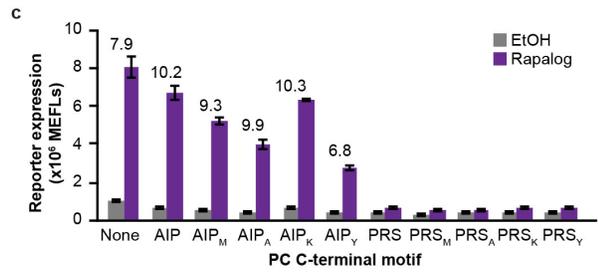
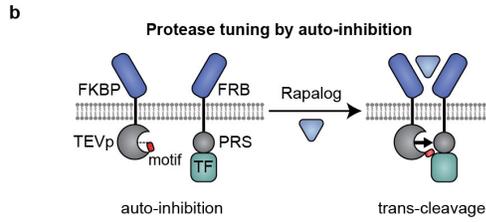
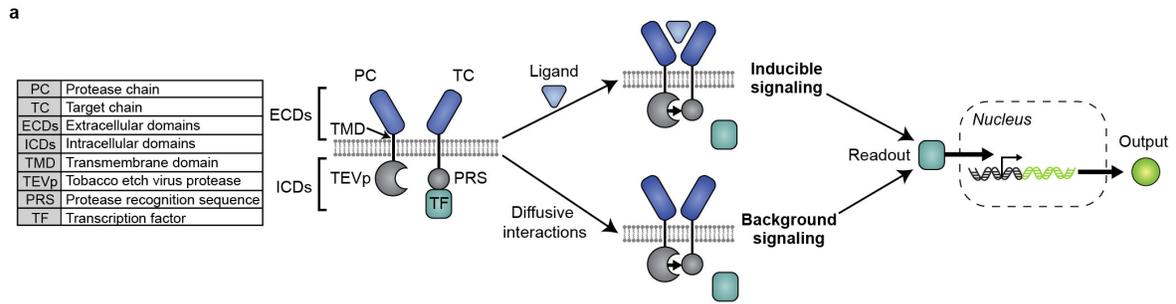


Fig 3.1 Protease chain tuning to improve MESA receptor performance. **a** This schematic depicts the MESA signaling mechanism. Ligand-induced receptor dimerization results in TEVp-mediated *trans*-cleavage to release a TF, which then enters the nucleus and induces target gene expression. Ligand-independent (background) receptor interactions can also result in TF release. **b,c** This proposed TEVp auto-inhibition strategy (**b**) was explored by functional evaluation (**c**) of MESA receptor variants in which a peptide—either a modified auto-inhibitory peptide (AIP: ELVYSQX) or protease recognition sequence (PRS: ENLYFQX), where X is a variable amino acid (e.g., AIP_M: ELVYSQM)—has been appended onto the C-terminus of the protease chain (PC) TEVp. The leftmost condition is the base case (no peptide appended). Each condition uses a target chain (TC) with M at the P1' site of the PRS. Numbers above bars indicate fold difference in reporter signal between samples treated with rapalog (dissolved in EtOH) vs. EtOH (vehicle-only control). Fold difference values are shown for samples in which ligand treatment induced significant signal above background (two-way ANOVA, $p < 0.05$). **d,e** Juxtamembrane cleavage of the PC (**d**) is suggested by Western blot analysis (**e**) of PCs tagged with 3x-FLAG on either the N-terminus or C-terminus; the PC appears to be cleaved into two fragments having sizes consistent with cleavage near the transmembrane domain (N = N-terminal product, C = C-terminal product). PCs with the TEVp^{D81N} mutation, which renders TEVp catalytically inactive⁴, were cleaved similarly to the catalytically active PCs; thus we conclude that the observed cleavage can be attributed to endogenous cellular processes. **f** PCIL substitution generally led to decreased receptor performance through increased background and/or decreased induced signal (two-way ANOVA, $p < 0.05$). Fold difference values are shown above bars for samples in which ligand treatment induced significant signal above background (two-way ANOVA, $p < 0.05$). Bars depict the mean of three biological replicates, and error bars represent the S.E.M. Outcomes from ANOVA and Tukey's HSD tests for **c,f** are in **Supplementary Notes A3.1 and A3.2**. In this experiment and subsequent experiments, we employ a rapamycin analog (rapalog) as a model ligand, which induces higher ligand-induced reporter expression than does rapamycin (**Supplementary Fig. A3.1a**).

could function as a competitive inhibitor of intact receptors, and the TEVp released into the cytosol could contribute to background signaling. To explore alterations that might prevent cleavage of the PC, we first varied the sequence of the PC inner linker (PCIL) that connects the TEVp and transmembrane domain (TMD) by introducing positively charged residues and sequences from native receptors. These substitutions improved protein stability, reducing the appearance of the originally observed juxtamembrane cleavage product (**Supplementary Fig. A3.2a**). However, none of these substitutions improved receptor performance, and some substitutions diminished functional performance by increasing background signaling and/or decreasing induced signaling and therefore decreasing the ligand-induced fold difference (**Fig. 3.1f**). This effect could not be overcome by decreasing the PC plasmid dose (to compensate for the increased levels of intact PC) without diminishing ligand-induced signaling (**Supplementary Fig. A3.2b,c**), and additionally these substitutions reduced surface expression of the PC (**Supplementary Fig. A3.2d**). Although it is not clear why each substitution conferred these undesirable effects, it was clear that PCIL substitution alone did not address PC stability or background signaling via a useful mechanism, and so we turned to other modifications as alternative approaches.

Transmembrane domain substitution can improve receptor performance and stabilizes the protease chain

TMD choice is an as-of-yet unexplored and potentially important aspect of MESA design. Previous MESA receptor designs^{2, 115} employ a form of the CD28-TMD commonly used in chimeric antigen receptors (CARs) that differs somewhat from the native CD28-TMD sequence. When engineering CARs, TMD choice has proven to be a useful handle for tuning interactions between receptor chains and modulating the strength of target antigen binding-induced receptor signaling^{232, 233}. Since the native CD28 receptor clusters as a member of the immunological synapse formed between a T cell and an antigen-presenting cell^{234, 235}, we hypothesized that when the CD28-TMD is employed in a synthetic receptor system such as MESA, some residual clustering (mediated by the TMD alone) could lead to ligand-independent signaling. We speculated that using an alternative TMD might avoid this problem. In natural receptor tyrosine kinases (RTKs), the TMDs regulate diverse aspects of receptor signaling mechanisms, including dimerization propensity and geometry, ligand binding-induced rotational conformation changes, and clustering²³⁶.

Therefore, we decided to investigate whether replacing the CD28-TMD in MESA with other TMD variants might improve receptor performance.

We selected a panel of seven natural TMDs from RTKs, sampling a range of reported dimerization propensities (48), and two synthetic TMDs²³⁷, and we substituted these for the CD28-TMD (**Fig. 2a**). TMD substitution had a substantial effect on receptor expression, and in some cases this effect differed for the TC and PC (**Supplementary Fig. A3.3a**). Since MESA performance depends upon both chain expression level and the ratio of expression of the two chains¹¹⁷, we sought to isolate the mechanistic effects of TMD choice from the effects upon expression levels. Therefore, we normalized protein expression by iteratively varying plasmid dose (**Methods, Supplementary Fig. A3.3**). In this context, we observed that all of the TMD substitutions except for the FGFR1-TMD stabilized the PC (i.e., resolved the juxtamembrane cleavage observed with the CD28-TMD) (**Fig. 3.2b, Supplementary Fig. A3.2**), and all substitutions altered surface expression (**Supplementary Fig. A3.4a**). In subsequent functional evaluations, using TC:PC protein expression ratios of approximately 1:1, TMD substitution conferred substantial, expression level-independent effects on performance (**Fig. 3.2c, Supplementary Fig. A3.4b**). Notably, employing the TMDs from GpA and FGFR4 increased the ligand-induced signaling compared to the CD28-TMD without increasing background signaling, leading to high-performing systems. Conversely, utilizing the TMD from FGFR1 increased both ligand-induced and background signaling, and the TMDs from FGFR2, FGFR3, EphA4, and VEGFR1 did not result in receptors that were capable of signaling.

To determine whether using different TMDs on the TC and PC could yield further improvements, we evaluated the 100 (10×10) pairwise TC-PC combinations (**Fig. 3.2d, Supplementary Fig. A3.5a,b**). In our initial screen, most TCs that conferred little or no detectable ligand-induced signaling when paired with a PC bearing the same TMD (matched pairs) also showed little or no ligand-induced signaling when paired with a different TMD PC (mixed pairs) (**Supplementary Fig. A3.5a,b**). An exception to this trend is the VEGFR1-TMD, which showed some ligand-induced signaling when paired with a PC containing the CD28-, GpA-, FGFR1-, FGFR3-, FGFR4-, or Valine-TMD. Notably, many other mixed TMD pairs also showed substantially improved performance compared to the matched CD28-TMD base case, including reduced background signal and/or increased ligand-induced signal, leading to fold differences as high as 97 (**Fig.**

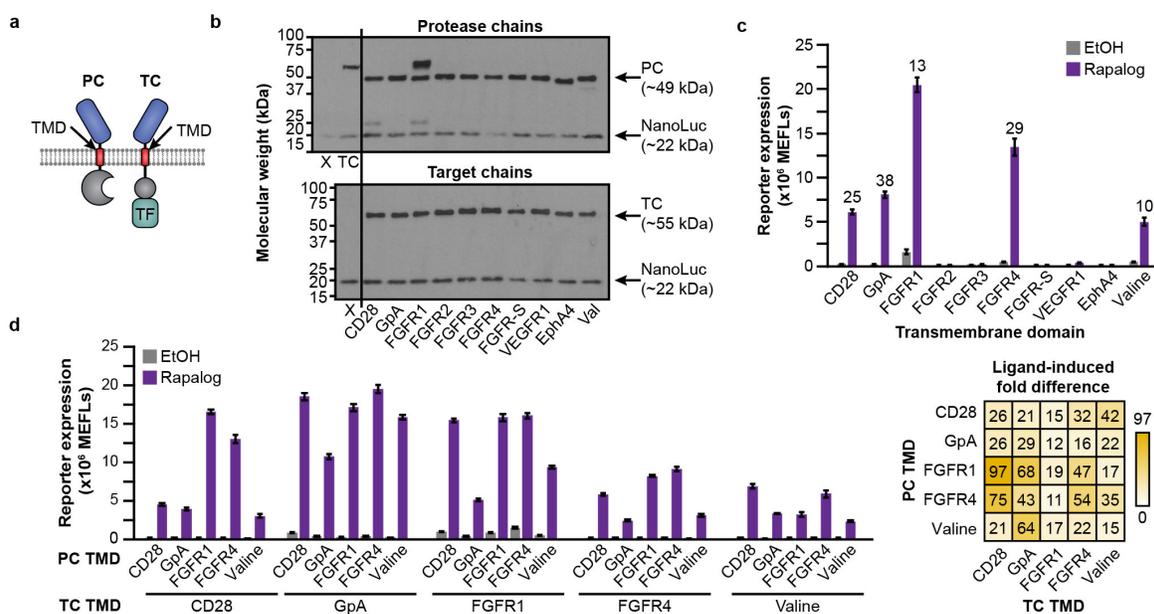


Fig 3.2 TMD contributions to MESA receptor signaling. **a** This schematic identifies the design choice examined here—the TMD sequence. **b** Effect of TMD choice on the expression of expected bands (PC, TC, co-transfected NanoLuc loading control) versus cleavage products (CD28 and FGFR1 cases). For this experiment, chain expression levels were first normalized to that of CD28-TMD TC expression (M, upper panel) by varying transfected plasmid dose through iterative Western blot analyses (**Supplementary Fig. A3.3**). The X denotes a vector-only negative control (including NanoLuc); TC denotes a CD28-TMD TC. **c** Paired TMD substitution conferred varying effects on receptor performance. Labels indicate the TMD that was used on both the TC and PC. Numbers above bars indicate fold difference when the ligand induced a significant signal above background (two-way ANOVA, $p < 0.05$). **d** Combinatorial TMD substitution further improved receptor performance. Fold difference is reported in the heatmap at right. All combinations exhibit a significant increase in reporter expression upon ligand treatment (three-way ANOVA, $p < 0.001$). Bars depict the mean of three biological replicates, and error bars represent the S.E.M. Outcomes from ANOVAs and Tukey's HSD tests for **c,d** are in **Supplementary Notes A3.1, A3.2, and A3.3**.

3.2d, Supplementary Fig. A3.5c). Together, these results suggest that TMD choice is a key determinant of receptor performance and a rich target for tuning.

Transmembrane domain choice does not substantially impact receptor dimerization propensity

Given the promising results obtained with certain TMD choices, we next sought mechanistic insight into the roles of these domains in MESA signaling. For native receptors, TMDs can affect both localization and function²³⁶, but how this choice affects synthetic receptor function is unexplored. Since some TMD sequence motifs mediate receptor homodimerization, we hypothesized that TMD choice might affect MESA receptor performance by modulating the propensity for chains to associate. We first evaluated TMD association computationally using TMDOCK, a tool that uses amino acid sequence to predict TMD association by simulating alpha helix packing arrangements and conducting local energy minimization²³⁸. This analysis predicted differences in matched TMD interactions, although the predicted trends only partially agreed with our experimental observations. For example, TMDOCK predicted the FGFR1-TMD to exhibit a high propensity to homodimerize, which is consistent with the observed high background signal (**Fig. 3.2c**). In contrast, TMDOCK also predicted the GpA-TMD to homodimerize (with more stability than the CD28-TMD), which is consistent with prior biochemical analyses^{239, 240}, yet we observed no evidence of enhanced homodimerization (i.e., in the form of high background) in functional assays. One possible explanation for this discrepancy is that TMDOCK evaluates TMD interactions in isolation, omitting any effects that might be conferred by the intracellular or extracellular receptor domains, and thus this tool alone is insufficient to explain the mechanistic consequences of TMD choice. Therefore, we next sought to directly investigate how TMD choice affects MESA receptor association by experimental characterization of full-length proteins.

To develop an assay for experimentally quantifying MESA inter-chain association, we employed a Förster resonance energy transfer (FRET)-based method^{222, 241} (**Fig. 3.3a**). We replaced the TC and PC intracellular domains (ICDs) with mVenus or mCerulean—fluorescent proteins that exhibit FRET in a manner dependent on spatial co-localization (FRET occurs within a donor-acceptor distance of approximately 10 nm^{242, 243}). Using this assay, ligand-independent association and ligand-dependent association are quantified by measuring acceptor (mVenus) fluorescence upon donor (mCerulean)

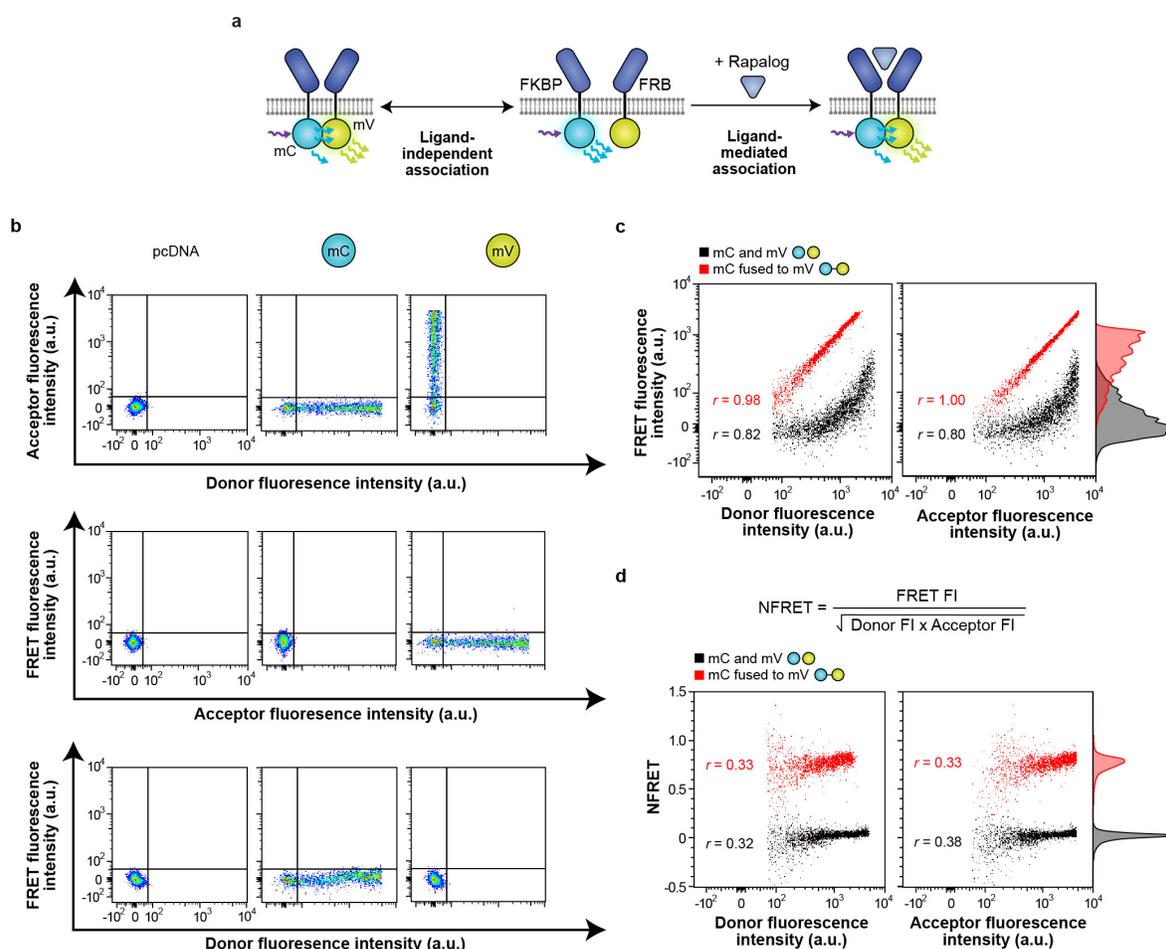


Fig 3.3 Development of a flow cytometric FRET approach to probe receptor chain association. **a** This schematic illustrates our strategy for quantifying ligand-independent (left) and ligand-mediated (right) receptor associations using Förster resonance energy transfer (FRET). Rapamycin-sensing MESA receptor ICDs were replaced with mCerulean (donor) and mVenus (acceptor) fluorophores. **b** Single-fluorophore samples were used to linearly compensate bleed-through from individual fluorophores into both the other fluorophore channel and the FRET channel. These plots also illustrate the gating used to identify cells expressing both the donor and acceptor fluorophores (mCerulean+/mVenus+). Abbreviations: mC, mCerulean; mV, mVenus. **c** Cytosolically expressed control constructs that are expected to display low FRET (separate soluble donor and acceptor proteins) or high FRET (donor-acceptor fusion protein) differ by a vertical shift in fluorescence in the FRET channel. Fluorescence in the FRET channel is linearly correlated with donor and acceptor fluorescence, respectively. The cells shown are singlets that are transfected (miRFP670+) and that express the donor and acceptor (mCerulean+/mVenus+).

d When FRET fluorescence is normalized to donor and acceptor fluorescence intensities by the calculated NFRET metric (equation shown), the cytosolic controls still display a vertical shift in NFRET, but NFRET only has a low correlation with donor and acceptor fluorescence, respectively; NFRET is more independent of expression differences observed across the cell population (compared to FRET fluorescence intensity in **c**). The NFRET metric better distinguishes low and high FRET controls than does unprocessed FRET fluorescence. The cells shown are singlets that are transfected (miRFP670+) and that express the donor and acceptor (mCerulean+/mVenus+). Experiments were conducted in biological triplicate, and individual representative samples are shown. Adjunct histograms represent probability density and are scaled to unit area. Data were analyzed as described in **Methods**.

excitation. To establish a high-throughput workflow that yields single-cell resolution data, we first adapted a reported approach to quantify FRET by flow cytometry²⁴¹. Importantly, this workflow includes normalization of FRET signal to fluorophore levels on a single-cell basis, which is necessary to account for heterogeneity in protein expression. In this approach, single-cell mCerulean donor fluorescence, mVenus acceptor fluorescence, mVenus FRET fluorescence, and fluorescence from constitutively expressed miRFP670 are quantified by flow cytometry (**Table 3.1**). Single-color samples are analyzed to apply post-hoc linear compensation for spectral bleed-through across fluorophore and FRET channels using an expression range that encompasses the receptor samples (**Fig. 3.3b**); only transfected cells (miRFP670+) that are both mCerulean+ and mVenus+ are analyzed.

To validate our FRET assay, we first used a model system: cells expressed either a positive control (an mVenus-mCerulean fusion protein that is expected to exhibit strong FRET²⁴⁴) or a negative control (mVenus and mCerulean expressed as separate proteins that are expected to exhibit minimal FRET). We evaluated two possible metrics to quantify FRET efficiency. The first metric, FRET fluorescence intensity, was quantified using unprocessed fluorescence intensity in the FRET channel. As expected, this metric correlated linearly with donor fluorescence and acceptor fluorescence (proxies for the expression level of each fluorescent protein), respectively, and as expected, FRET fluorescence intensity was higher for the fusion protein than for the non-fused control (**Fig. 3.3c**). The second metric, normalized FRET (NFRET)^{222, 245}, was calculated by normalizing the FRET fluorescence intensity of each cell to that cell's donor and acceptor fluorophore expression levels to provide a whole-cell FRET metric that accounts for fluorescent protein expression level (**Fig. 3.3d**). Like FRET fluorescence intensity, NFRET was higher for the fusion protein than for the non-fused control. However, NFRET demonstrated the anticipated low correlation with

fluorophore expression level and enabled comparisons across two orders of magnitude in donor and acceptor fluorescence (**Fig. 3.3d**). Although each of these metrics can report *whether* FRET occurs in this model system, NFRET provides better quantitative resolution separating the two control populations (**Fig. 3.3c,d**), and since it inherently controls for variation in protein expression, we utilized this metric exclusively going forward. As a final validation of our analytical pipeline, and to enable comparison of our methods with classic microscopy-based FRET analyses, we adopted our method to a confocal microcytometry workflow (**Methods**), which confirmed the patterns observed by flow cytometry.

Having validated this FRET assay, we next employed it to interrogate the contribution of TMD choice to MESA inter-chain interactions. To investigate the mechanistic questions and apparent contradictions discussed above, we selected three TMDs for investigation: the CD28-TMD was selected for its predicted large number of energetically favorable modes of association, propensity to cluster²⁴⁰, and use in previous MESA receptors^{2, 115}; the GpA-TMD was selected for its documented high propensity to homodimerize^{239, 240}, and surprising lack of elevated background signaling (**Fig. 2c**); and the FGFR4-TMD was selected for its documented²⁴⁰ and predicted low propensity to dimerize and high performance in functional assays (**Fig. 3.2c**). We hypothesized that if TMD choice impacts receptor function *primarily* through modulating inter-chain affinity, then functional characteristics (e.g., low vs. high background signaling) would correlate with FRET trends (e.g., low vs. high basal FRET signal, respectively). We started by evaluating MESA with matched TMD pairs and found that background NFRET was slightly (but significantly) lower for the GpA-TMD and FGFR4-TMD based MESA than for the CD28-TMD based MESA, although ligand-induced increases in NFRET (i.e., fold difference) were significant and similar across all three TMD choices (**Fig. 3.4a**). These receptors also exhibited similar ligand dose responses (**Fig. 3.4b**). We next examined MESA receptors with mixed TMD pairs and again observed that constructs exhibited similar and significant ligand-induced NFRET (**Fig. 3.4c**). This pattern held upon swapping the fluorophore domains, testing different cell harvest methods, and measuring NFRET by confocal microscopy. To investigate whether TMD choice impacts association kinetics, we performed a time course assay (**Fig. 3.4d**), which revealed that MESA receptors containing the CD28-TMD and/or the FGFR4-TMD exhibited similar NFRET induction kinetics. Onset was generally observed within 15 min of ligand addition (**Fig. 3.4d**),

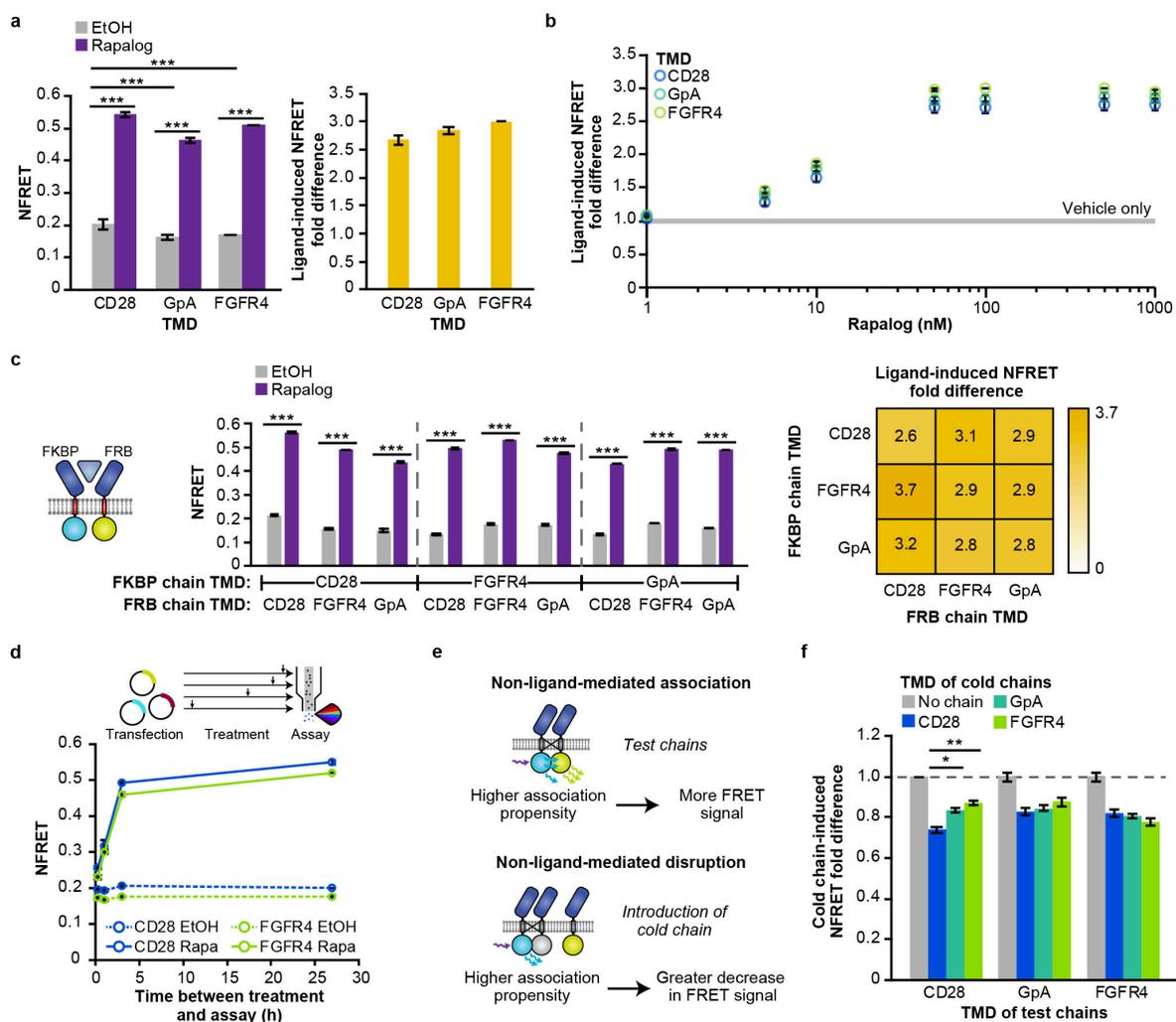


Fig 3.4 Effect of TMD choice on receptor chain association. **a** Receptor pairs with matched TMDs exhibit a ligand-induced increase in NFRET (27 h incubation, 100 nM rapalog) (two-way ANOVA, $*** p < 0.001$) (left). The CD28-TMD matched receptor pair exhibits slightly higher NFRET in the absence of ligand compared to the GpA-TMD pair and FGFR4-TMD pair (two-way ANOVA, $*** p < 0.001$). Fractional change in NFRET upon ligand treatment (ligand-induced NFRET fold difference) is comparable across TMDs (two-tailed Welch's t -test, all $p > 0.05$) (right). In all cases, the donor fluorophore is on the FKBP chain and the acceptor fluorophore is on the FRB chain. **b** NFRET induction varies with rapalog dose (measurement at 27 h incubation). **c** Pairs of receptors with mixed and matched TMDs exhibit a significant ligand-induced increase in NFRET (27 h incubation, 100 nM rapalog) (three-way ANOVA, $p < 0.001$). The ligand-induced NFRET increase is comparable across mixed and matched TMD pairs. **d** Dynamics of NFRET response to ligand. By 3 h post-ligand treatment, the NFRET increase is nearly maximal (87% relative to the NFRET at 27 h). Abbreviation: Rapa, rapalog.

e,f In a cold (non-fluorescent) chain competition assay with matched TMD fluorescent receptors, CD28-TMD exhibited a slightly higher propensity to associate: the NFRET decrease conferred by introduction of a CD28-TMD cold chain was greater in the CD28-TMD matched case than in other cases (two-tailed Welch's *t*-test, all $p > 0.05$ except for comparison between CD28-mediated and GpA-mediated disruption of CD28 FRET and CD28-mediated vs. FGFR4-mediated disruption of CD28 FRET, * $p < 0.05$, ** $p < 0.01$). All chains contain the same ECD (FRB). Experiments were conducted in biological triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. Outcomes from ANOVAs and Tukey's HSD tests for **a–c** are in **Supplementary Notes A3.2 and A3.3**.

indicating that this signal is attributable to the ligand-induced association of existing chains, rather than potential ligand-induced stabilization of newly synthesized chains. Altogether, these evaluations revealed no substantial TMD-dependent effects, which argues against the hypothesis that TMD choice impacts MESA function primarily through modulating inter-chain affinity.

Given the surprising finding that TMD choice did not appear to impact inter-chain association, we next used a separate and potentially more sensitive FRET assay to seek confirmation of this finding. For this evaluation, we performed a cold chain competition (**Fig. 3.4e**), in which fluorescently tagged MESA chains (called test chains) compete with MESA chains employing a non-fluorescent mVenus mutant (called cold chains) for TMD-mediated interactions. We expect that introduction of a cold chain will lead to a decrease in FRET between test chains, and that the magnitude of this decrease will trend with the affinity with which test and cold chains associate in a ligand-independent manner. As expected, we observed that the introduction of cold chains reduced NFRET between all test chains, and notably this reduction was approximately 20% in all but one case (**Fig. 3.4f**): CD28-TMD-containing test chains showed a slightly but significantly greater decrease in NFRET when co-expressed with CD28-TMD cold chains rather than with any other cold chains. This result indicates that MESA with the CD28-TMD have a slightly higher association propensity than do MESA with the GpA-TMD or FGFR4-TMD, which is consistent with the previously observed slight elevation in background NFRET for MESA with the CD28-TMD, compared to MESA with other TMD choices (**Fig. 3.4a**). Together, these results indicate a slightly higher propensity for CD28-TMD MESA to self-associate compared to MESA with the FGFR4-TMD and GpA-TMD. Interestingly, the increased propensity for CD28-TMD MESA to self-associate did not manifest in elevated background signal compared to MESA with other TMDs, and these results are insufficient to explain the observed TMD-dependent differences observed with GpA-TMD and FGFR4-TMD MESA in functional assays (**Fig. 2**).

Therefore, we conclude that TMD choice primarily affects MESA receptor performance by modulating a property other than TMD-dependent inter-chain association propensity.

Transmembrane domain choice governs receptor dimerization geometry and trans-cleavage efficiency

Considering the preceding observations, we hypothesized that TMD choice might impact the geometry of receptor dimerization in a manner that affects *trans*-cleavage efficiency without affecting the propensity for the chains to associate. To test this hypothesis, we used a panel of synthetic TMDs that were designed to dimerize with varying geometries and were previously used to study geometric constraints on juxtamembrane regions of RTKs that stemmed from the TMD²⁴⁶. In this panel, changes in TMD sequence are used to systematically vary the orientation of dimerization (**Fig. 3.5a**). TMDOCK analysis predicted that this panel of TMDs can homodimerize in configurations with distances between TMD C-termini (at the inner leaflet of the membrane) ranging from 5.4–16.0 Å and in a variety of rotational orientations. Notably, TMD-dependent differences in inter-chain orientation on this scale would not necessarily generate TMD-specific FRET signals. We hypothesized that if TMD interaction geometry does indeed impact *trans*-cleavage efficiency, then MESA receptors built using this panel of synthetic TMDs would exhibit differential signaling in a functional assay. We found that ligand-induced signaling indeed varied across the panel and decreased substantially as the residues mediating dimerization were moved away from the inner leaflet of the membrane (**Fig. 3.5b**). Notably, some synthetic TMD choices yielded very little ligand-induced signal, similar to what was observed with several native TMDs (**Fig. 3.2c**). The aforementioned effects of TMD choice could not be explained by differences in expression level of the receptors—the chains with the lowest magnitude of ligand-induced signaling were the most highly expressed (**Supplementary Fig. A3.6a**)—nor by differences in ligand-induced MESA protein accumulation, which was modest and similar across the panel (**Supplementary Fig. A3.6b**). Altogether, these data support the conclusion that TMD choice substantially influences the efficiency of MESA *trans*-cleavage via a mechanism that involves the intracellular geometry in which MESA chains dimerize upon ligand-binding.

Design rules extend to the development of new synthetic receptors

Given our new understanding of the role of TMD choice, we next investigated whether the observed trends would extend to MESA systems with different ECDs and ligands. To test this hypothesis, we first built two new MESA receptors to sense small molecules—gibberellin (GA3-AM is a cell-permeable analog) and abscisic acid (ABA)—and a third new MESA receptor to sense green fluorescent protein (GFP; here we use co-expressed secreted GFP, sGFP, as an expedient testing system), and explored design considerations that are typically expected to be ECD-specific, such as how linker length affects expression and cell-surface localization^{247, 248}. Interestingly, functional assays revealed TMD-associated trends that were largely consistent across ECDs (**Fig. 3.6a–d**). The TMD choice for each chain significantly affected background signaling and induced signaling, and the interaction between TMD choices was also significant, indicating that the choice of TC TMD or PC TMD alone does not fully explain the trends (**Supplementary Note A3.3**). Additionally, the TMD choices together account for most of the variance in background and induced signaling observed (**Supplementary Note A3.2**). These observations indicate that satisfying any one design objective (e.g., maximize fold difference, minimize background, etc.) requires choosing a pair of TMDs suited to that design goal. Additionally, some general trends held across the rapamycin MESA receptors and these three new receptors, suggesting that these might represent generalizable principles guiding the design of future MESA receptors. For example, high background and modest induced signaling were observed for pairs including FGFR1-TMD TC, resulting in generally low fold difference. Conversely, FGFR4-TMD-containing pairs often exhibited low background signaling and high fold difference (**Fig. 3.6a–d**). Overall, we conclude that some effects of TMD choice extend across multiple MESA receptors, and that a limited experimental evaluation of these few TMD choices enables one to generate multiple new high-performing receptors.

We next investigated whether the observed effects of TMD choice might extend to synthetic receptors that operate by a distinct mechanism. To test this, we employed a recently reported MESA system in which ligand binding induces reconstitution of a mutant split TEVp³. The reconstituted protease then releases a TF from the receptor, enabling the TF to enter the nucleus and drive reporter expression (**Fig. 3.6e**). We replaced the CD28-TMD used in the reported rapamycin-sensing base case³ with other TMDs

and also built GFP-sensing variants of these receptors. In functional assays, we again found that the TMD proved to be a useful handle for tuning performance (**Fig. 3.6e,f**). Background and ligand-induced signaling were again significantly affected by TMD choice for each chain and the interaction between the TMD choices (**Supplementary Note A3.3**). Strikingly, across all mixed and matched TMD pairs, the ligand-induced signal was relatively consistent, which differs from the trends observed for the *trans*-cleavage mechanism. Moreover, for these receptors, using the CD28-TMD on only one chain resulted in very low background signaling, but using the CD28-TMD on both chains (as previously reported³) resulted in the highest background of all combinations tested, with the interaction between TMD choices accounting for most of the variance in background signaling (**Supplementary Note A3.2**). When considered together with the FRET experiments (**Fig. 3.4a,f**), these results again suggest that CD28-TMD MESA exhibits increased ligand-independent association propensity compared to MESA employing other TMDs. This effect can be problematic if both chains include this TMD, yet this same property can be useful if only one chain bears this TMD. Comparing these two distinct MESA receptor mechanisms suggests that whereas subtle geometric effects conferred by specific TMD pairs have a substantial impact on MESA receptors employing the original *trans*-cleavage mechanism (**Fig. 3.6a-d**), these nuances can have less impact on the performance of receptors employing the distinct split-TEVp reconstitution mechanism (**Fig. 3.6e,f**). The observation that the effects conferred by TMD choice are dependent on signaling mechanism is also consistent with our hypothesis that TMD choice is most important for determining the geometry of dimerized receptors. If, conversely, the main effect of TMD choice were modulating inter-chain association propensity, then we might expect TMD choice to yield similar effects when used with either mechanism, yet we see this consistency only in CD28-TMD-mediated elevation of background signaling. A final important finding is that for both MESA receptor mechanisms examined, our initial model systems generated insights enabling the design of novel receptors.

Finally, given these new insights into the important connection between MESA receptor geometry and functional performance, we sought to utilize these tools to guide the selection of new ligands and ligand-binding domains. For MESA receptors engineered to sense rapamycin, gibberellin, abscisic acid, or in previous work, vascular endothelial growth factor (VEGF)²⁰⁴, there exist crystal structures enabling us to

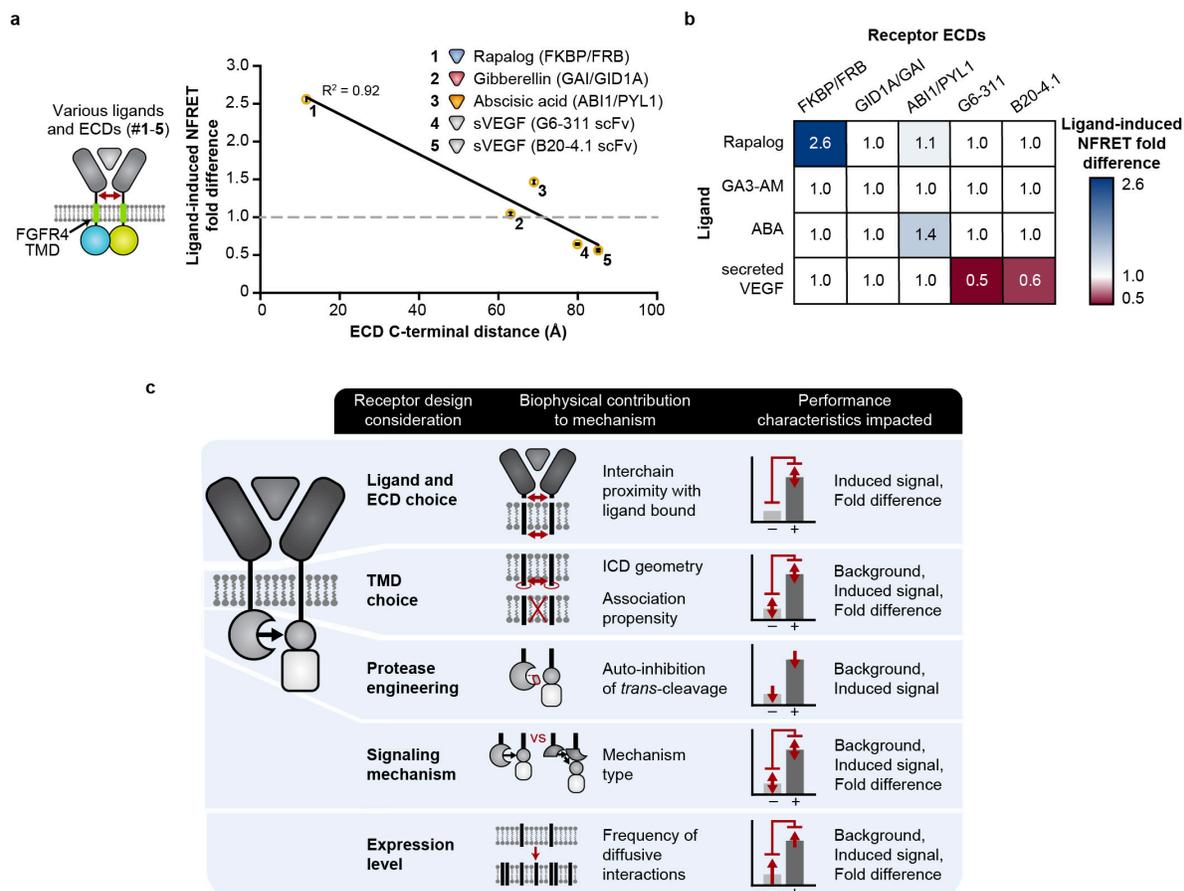


Fig 3.7 Generalizing principles for receptor engineering. **a** Ligand-induced inter-chain association (NFRET fold difference) varies with ECD C-terminal distance, with a negative linear relationship ($y = -0.027x + 2.9$, where y is NFRET fold difference and x is C-terminal distance in Å, $R^2 = 0.92$, two-tailed Student's t -test, $p = 0.01$). The ligands are rapalog, GA3-AM, ABA, and secreted VEGF. The red arrow indicates the ECD C-terminal distance—the spatial displacement between C-termini of the ligand-binding domains. **b** Validation of receptor-ligand orthogonality: substantial fold differences in NFRET were observed only when each ligand was paired with its cognate ligand-binding domains. Experiments were conducted in biological triplicate, and data were analyzed as described in **Methods**. Error bars represent the S.E.M. **c** This schematic synthesizes the findings of this investigation with prior MESA receptor development^{2,3} by relating receptor design choices to the proposed biophysical consequences and performance characteristics affected.

estimate the displacement between the C-termini of the ECDs when they are dimerized in the ligand-bound state^{225, 226, 227, 228}. To determine whether this spatial variation is sufficient to impact intracellular receptor geometry, we again employed FRET analysis (employing the well-behaved FGFR4-TMD) to best decouple our investigation of this question from the effects of receptor geometry on *trans*-cleavage (**Fig. 3.7a**). Interestingly, ligand-induced NFRET fold difference showed a strong negative correlation with the ECD C-terminal distance, suggesting that one can predict some aspects of receptor structure and function from prior knowledge of the ECDs alone—an attractive feature for a modular design strategy. We confirmed that these effects were indeed attributable to receptor complex formation by performing a ligand-ECD orthogonality analysis (**Fig. 3.7b**). Notably, the VEGF-binding ECDs—which correspond to the largest ECD C-terminal distances evaluated—showed a *diminishment* in NFRET upon ligand treatment in both experiments, suggesting that ligand-induced dimerization might diminish, on average, transient inter-chain ICD interactions compared to the ligand-free state. For a given TMD choice, the ligand-induced fold difference values exhibit a positive correlation (in rank order of effects) between NFRET and reporter expression. Comparing all FGFR4-TMD receptors used in both FRET and functional assays, the highest values for both metrics were observed for the rapamycin-sensing system, with lower but similar values observed for the gibberellin and abscisic acid-sensing systems (**Fig. 3.6a–c, Fig. 3.7a**). These observations could help to explain why MESA receptors with shorter ECD C-terminal distances (e.g., for rapamycin-sensing) generally exhibit strong signaling. They also suggest that functional sensing of gibberellin, abscisic acid, and VEGF (including receptor chain *trans*-cleavage) can be attributed to inter-chain interactions that occur with lower frequency than those that are required to confer FRET. Altogether, the analyses presented in this study provide powerful new insights into the connections among tunable protein design choices, biophysical consequences, and impacts on the performance of synthetic receptors (**Fig. 3.7c**).

3.6 Discussion

A key finding of this study is that systematic re-evaluation of the MESA synthetic receptor system enabled identification of design modifications that substantially improve receptor performance. While some modifications had modest effects, TMD substitution proved to be a particularly useful handle for receptor

tuning and optimization (**Fig. 3.2**). Some TMD combinations greatly decreased background signaling while maintaining or increasing ligand-induced signaling, yielding high-performing receptors with fold differences on the order of 100. By this metric, these rank among the best synthetic receptors reported to date^{111, 112, 215}. Moreover, evaluating a relatively small library of TMD variants enabled us to generate high-performing MESA receptors for three new ligands (**Fig. 3.6**). It is notable that even for the fully synthetic mechanism employed by MESA receptors, it was not clear ahead of time which design modification strategies would be most fruitful. These improvements were achieved largely by investigating MESA receptor biophysics using the approaches and strategies typically employed to study natural receptors. There likely exists substantial room for improving synthetic receptor systems—a recent report evidences the utility of applying this approach to improve synNotch²¹⁸—and powerful tools developed by the receptor biophysics community comprise substantial yet underutilized potential.

An integral part of this investigation was the use of flow cytometric FRET. This technique enabled us to link disparate observations and build understanding (**Fig. 3.4**). FRET has generally proven useful for elucidating native receptor mechanisms, and these experiments typically employ confocal microscopy^{249, 250, 251}. We employed a substantially more high-throughput assay that provides single-cell resolution (thousands of individual cells per sample) by adapting a reported flow cytometry workflow²⁴¹ to this investigation (**Fig. 3.3**). A key utility of the NFRET metric is that it enables comparisons across a heterogeneous population of cells by expression-normalizing FRET. This feature of the NFRET metric makes it useful for evaluating receptor dimerization when receptors contain new protein domain choices (such as ECD), as it decouples the effects of these new domains on expression level from their effects on dimerization (with or without ligand). To our knowledge, this is the first application of flow-FRET to characterize synthetic receptors, and this approach may be useful for studying and optimizing systems beyond MESA. A consideration for future use is that our assay used the mCerulean/mVenus fluorescent protein pair and standard flow cytometry filters, which sufficed for our analysis, but it is possible that FRET signal could be increased by selecting alternative fluorophores and/or bespoke filters.

Altogether, this investigation provides new mechanistic insights into how TMD choice impacts MESA receptor performance. Only the CD28-TMD used in previous MESA receptors^{2, 3, 115} showed some

modest propensity to cluster in the absence of ligand (**Fig. 3.4a,f**), which is consistent with roles played by CD28 in T cell signaling at the immunological synapse²³⁵. In general, TMD choice conferred little effect (CD28-TMD) or no effect (all other tested TMDs) on the propensity of MESA chains to associate in the presence or absence of ligand. Thus, there must exist an alternative explanation as to why TMD choice profoundly impacts background signaling, ligand-induced signaling, and even PRS cleavability. We propose that TMDs primarily contribute to receptor performance by influencing the geometry in which intracellular domains interact. Several lines of evidence support this hypothesis. Systematically varying the geometry with which TMDs associate either facilitates or constrains MESA receptor signaling (**Fig. 3.5**), and some TMDs can render a TC resistant to cleavage by a PC (**Fig. 3.2c, Fig. 3.5b, Supplementary Fig. A3.5**). TMD-associated trends also held to some extent across various choices of ligand and ligand-binding domains (**Fig. 3.6a–d**). Although it is not yet clear *why* each TMD pair impacts the geometry with which intracellular domains associate, experimentally screening the TMD library reported here enabled the improvement of existing MESA receptors based upon each of the two distinct mechanisms (**Fig. 3.6**)^{2, 3, 115}. It remains to be seen whether further TMD screening or engineering may represent an opportunity for further enhancement of MESA receptors, and whether TMD substitution will improve other synthetic receptors.

Ultimately, we hope that this investigation will serve as a guide for building and tuning synthetic receptors by connecting design choices to biophysical consequences and performance characteristics of interest (**Fig. 3.7c**). When designing new MESA receptors, some choices must be made at the outset. First, the geometry with which ligand-binding domains are separated in physical space when bound to ligand fundamentally limits inter-chain interactions (**Fig. 3.7a,b**), so minimizing the apposition of these domains may benefit MESA receptor performance (**Figs. 3.6, 3.7**). Since the magnitude of NFRET induction upon ligand addition roughly correlates with the fold difference in ligand-induced signaling (**Fig. 3.6a–c, Fig. 3.7a**), our FRET assay may be a useful tool for selecting candidate ligand-binding domains when a crystal structure cannot be used as a guide; since FRET phenomena depend weakly on TMD choice, such an initial evaluation need not include a combinatorial TMD screen. Second, since TMD choice substantially influences MESA performance, we recommend that functional evaluations of new receptors include a

limited combinatorial sampling of TMDs (e.g., using a panel based upon those included in **Fig. 3.6**). Finally, expression tuning also provides a fairly facile handle for optimizing (empirically, if needed) specific receptor performance characteristics, although this might be less important when using MESA that employ the split TEVp mechanism, which is less sensitive to receptor expression level³. The methods and insights developed here should facilitate the construction of novel, high-performing receptors for diverse ligands, including both MESA receptors and potentially synthetic receptor systems in general.

3.7 Data availability

Notable sequences for plasmids used in this study are provided in **Supplementary Data 1 (Online¹⁹⁹)**. Plasmid maps are provided as annotated GenBank files in **Supplementary Data 2 (Online¹⁹⁹)**. The majority of the plasmids used in this study are deposited with and distributed by Addgene, including complete and annotated GenBank files, at https://www.addgene.org/Joshua_Leonard/. This study additionally uses Addgene plasmids #58855, #138749 (described in **Materials and Methods**). The exact doses of DNA used in each experiment are listed in **Supplementary Data 3 (Online¹⁹⁹)**. All reported experimental data are provided in **Source Data 1 (Online²²¹)** and unprocessed western blot images are available in **Source Data 2 (Online²²¹)**, both of which can be found on Zenodo (<https://doi.org/10.5281/zenodo.4026861>)²²¹.

3.8 Acknowledgements

This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH) under award number 1R01EB026510, the National Cancer Institute (NCI) of the NIH under award number F30CA203325, the National Institute of General Medical Sciences (NIGMS) of the NIH under award number T32GM008152 (to Hossein Ardehali), the National Science Foundation under award number MCB-1745753, the Northwestern University Flow Cytometry Core Facility supported by Cancer Center Support Grant (NCI 5P30CA060553), and the NUSeq Core of the Northwestern Center for Genetic Medicine. Microscopy was performed at the Biological Imaging Facility at Northwestern University, graciously supported by the Chemistry of Life Processes Institute, the NU Office

for Research, and the Rice Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

This work was previously published as described (**3.1 Preface**) and modified for this thesis. It is used under an Open Access Creative Commons CC BY License.

Chapter 4. Developing robust cell-based biosensors for hypoxia

4.1 Preface

This chapter will ultimately be published as two papers, with the first covering the initial development of the HBS and methodology for studying for hypoxia and the second covering the feedback circuits for the sensor. I carried out this project with several talented undergraduate and masters students: Benjamin Leibowitz, Katie Zhu, Marya Ornelas, and Kate Chambers. Hailey Edelstein and I developed methods for working with landing pad cells and mMoClo in our laboratory. This project will be continued by several student including Jon Boucher and Kate Chambers leading the experimental work and Katie Dreyer leading the modeling efforts.

4.2 Abstract

While tumor-targeted, cell-based therapies have been successful in treating many hematologic malignancies, translating these successes to the treatment of solid tumors has been difficult for many reasons, including a lack of targetable tumor-specific antigens. One solution is to develop biosensors against features of the tumor environment common across many malignancies, for example hypoxia. Further, biosensors intended for these therapeutic purposes may provide utility beyond this original design, for instance, to study the development of hypoxia in a tumor longitudinally or in response to treatment via *in vivo imaging*. While several such biosensors exist, they rely on endogenous signaling pathways that are dysregulated in many cancers and therefore may not be robust to the choice of cell type and therefore not easily portable between different contexts. Ultimately, genetic circuits using synthetic transcription factors, such as the recently developed COMET toolkit (**Chapter 2**) could be used to address this challenge. Here, we characterize a DNA-based hypoxia biosensor in both transient transfection and in a stable landing pad context, in both the HEK293FT chassis line and the B16F10 murine melanoma line. By studying the effects of minimal promoter choice, responses to varying degrees of hypoxia, and the timing of the response, we identified targets for enhancing performance. We demonstrate that reliance on endogenous regulators of the hypoxic response results in difficulty porting the biosensor between cell types. To address this, we propose strategies for enhancing the robustness of synthetic biosensors with COMET. Ultimately, a high-performing hypoxia biosensor could be used for discovery, diagnostics, and therapeutics.

4.3 Introduction

Cell-based therapies have been successful in treating many hematologic malignancies, leading to substantial increases in survival for many patients. However, translating these successes to the treatment of solid tumors has been difficult for many reasons. Among these is the challenge of discovering tumor-specific antigens and developing biosensors against these targets. While this approach often leads to tumor-specific therapeutics, the diversity of antigens across cancers results in only incremental progress with each new therapeutic. An alternative approach is to develop biosensors against features of the tumor environment that are common across many malignancies, for example immunosuppression and hypoxia.

Hypoxia is a pathologic condition in which a tissue does not have enough oxygen and is present in many tumors. As growing tumors require a steady supply of nutrients for their continued proliferation, they produce vascular endothelial growth factor (VEGF) to induce the growth of blood vessels. However, tumor growth often surpasses the rate at which new blood vessels can grow^{79, 80}, and the vasculature in tumors is markedly abnormal, both of which lead to poor perfusion and resultant hypoxia⁷⁸. Tumor vasculature is elongated and tortuous, limiting the delivery of fresh blood, in contrast to well-organized physiologic vascular networks⁸¹. Further, the tight barrier formed by endothelial and smooth muscle cells is frequently insufficient, resulting in unusually permeable vasculature, which affects the ability to maintain adequate perfusion⁸². Together, these effects lead to regions of hypoxia throughout the tumor. While healthy tissues have multiple mechanisms to counteract hypoxia, including altering cellular metabolism and rapidly increasing blood flow through arterial dilatation, tumors lack the ability to respond likewise⁸³. This makes hypoxia a good marker for tumors in otherwise healthy humans, particularly those lacking ischemic disease.

The cell's response to hypoxia includes the stabilization of two hypoxia inducible factors (HIFs), HIF1 α and HIF2 α , as extracellular oxygen decreases^{252, 253}. These transcription factors (TFs) can each heterodimerize with a constitutively stable HIF1 β , bind to hypoxia response elements (HREs), and activate sets of genes that promote adaptation to and resolution of hypoxia. This pathway is also responsible for many adaptations that allow tumors to grow rapidly and is dysregulated in many tumors^{254, 255, 256, 257, 258, 259}. This signaling process forms the basis for DNA-based hypoxia biosensors (HBSs). By placing these HREs upstream of a minimal promoter, a downstream gene of interest can be conditionally expressed only when

the cell is experiencing hypoxic conditions^{84, 85, 86}. Additional elements can be placed upstream to increase the response to hypoxia, as well as small molecule mimetics of hypoxia. These sensors have enabled *in vivo* imaging of the response to hypoxia in mice^{85, 86}. It has also been shown that the choice of minimal promoter can influence magnitude of this response, as well as the amount of gene expression under normoxic conditions⁸⁷. However, biosensors that rely on this endogenous response may not be robust to dysregulation of the hypoxia response, which occurs in many tumors as both a survival mechanism and a consequence of the tumor microenvironment (TME)^{260, 261}.

Recently, we reported the development Composable Mammalian Elements of Transcription (COMET) toolkit for engineering genetic programs in mammalian cells¹. We subsequently demonstrated the ability of COMET to seamlessly integrate with and modulate the signal from synthetic biosensors²⁶². We hypothesize that COMET could be employed to modulate the output of an HBS in manners that confers robustness to the dysregulation of the hypoxia response among cell types. Here, we evaluate several HBS designs and their response to various levels and durations of hypoxia, in two cell types—the chassis HEK293FT humanoid cell line and the B16F10 model line for murine melanoma. We do so with the landing pad (LP) system, which was developed for the rapid prototyping of biosensors in a genomic context. An LP is a targeted integration locus, pre-engineered in a safe harbor locus, in which large amounts of DNA can be readily inserted using a transposase^{173, 263, 264}. LPs have several advantages over other methodologies, such as lentiviral transduction, including a much higher limit on cargo size. Additionally, as the cells with cargo integrated into the landing pad locus are genetically identical after integration, this methodology removes the confounding factor of biosensor integration locus and makes the resulting population more homogenous. During our initial evaluation of the HBS designs in the LP context, we discovered several opportunities for improving biosensor performance by modulating the signal with genetic circuits, including those based on endogenous and synthetic TFs. We designed several such circuits, tested several of them *in vitro*, and prepared to test the remainder in a manner that will allow for computational modeling of the circuits and prediction of the behavior of other HBS-modulating circuits *in silico*. These efforts will ultimately result in HBSs that are robust to the dysregulation of the hypoxic response and useful for fundamental and translational research, diagnostics, and therapeutics.

4.4 Materials and methods

Cloning, plasmid, and oligonucleotide sources

All constructs were initially characterized in the pPD005 backbone, which is a version of pcDNA3.1(+), modified as described previously¹. All HBS components were transferred into the mMoClo system¹⁷³, with Leonard Lab modifications, as described previously¹. Coding sequences are generally flanked by NheI and NotI restriction sites. Promoter regions are generally flanked by BglII or MluI on the 5' end and NheI on the 3' end. The HBS was synthesized with overlapping oligonucleotides from a previously published study⁸⁴, as were the YB_TATA and CMV_min promoters⁸⁷. The SV40_min was a gift from Yvonne Chen and PCR amplified prior to use⁸⁷. The stable HIF1 α mutant was sourced from pcDNA3 mHIF-1 α MYC (P402A/P577A/N813A) (Addgene #44028) was a gift from Celeste Simon²⁶⁵. EF1 α and TetON3G were sourced from pLVX-Tet3G (Clontech), and TRE3GV was sourced from pLVX-TRE3G (Clontech). BlastR was sourced from lenti dCAS-VP64_Blast, which was a gift from Feng Zhang (Addgene #61425)¹⁷⁷. The CHS4 insulator was sourced from PhiC31-Neo-ins-5xTetO-pEF-H2B-Citrine-ins, which was a gift from Michael Elowitz (Addgene #78099)¹⁷⁴. DsRed-Express2 was obtained by site directed mutagenesis of pDsRed2-N1, which was a gift from David Schaffer (University of California, Berkeley), and an internal Bpil restriction site in the coding region was ablated by making a sense mutation with site directed mutagenesis. EBFP2 was sourced from pEBFP2-Nuc, which was a gift from Robert Campbell (Addgene #14893)¹⁷². EYFP, was sourced from plasmids we previously described (Addgene #58855)¹¹⁵. mTagBFP2²⁶⁶, mNeonGreen²⁶⁷, mRuby3²⁶⁸, miRFP670^{269, 270}, and miRFP720²⁷¹ were synthesized as Gene Strings by ThermoFisher.

Construction of plasmids for B16F10-LP line generation

pU6-(BbsI)_CBh-Cas9-T2A-BFP-P2A-Ad4E4orf6 (Addgene #64220; referred to as pPD782) and pU6-sgRosa26-1_CBh-Cas9-T2A-BFP-P2A-Ad4E1B (Addgene #64219; referred to as pPD720) were gifts from Ralf Kuehn²⁷². The region encoding the BFP-P2A-Ad4E1B region (flanked by NheI/EcoRI) in pPD720 was replaced by a fragment encoding BFP-P2A-Ad4E4orf6 (flanked by NheI/EcoRI) to generate pPD783 U6 sgRNA Rosa26 Cas9-T2A-BFP-P2A-Ad4E4orf6.

The CAG promoter and homology arms for the Rosa26 locus were obtained from pR26 CAG/GFP Asc (Addgene #74285), which was a gift from Ralf Kuehn¹⁷⁵. The LP for the B16F10 cells was built by assembling the components as follows: TU1 contained the left Rosa26 homology arm, TU2 was assembled from the pPart series (CHS4x2 in pInsulator, CAG in pPro, a Placeholder UTR in p5'UTR, EYFP-P2A-HygroR in pGene, an inert 3' UTR in p3'UTR, and rbGlob PA terminator CHS4x2 in pPolyA) and then an attP site was inserted in place of the Placeholder 5'UTR, TU3 contained a Placeholder Homology arm, pLink3 into the Destination Vector. After a Golden Gate assembly, the Placeholder Homology arm was replaced with a Rosa26 Right Homology Arm. The final vector was named pPD864 and contained the LP flanked by Rosa26 homology arms. The pPart series vectors, pTU series vectors, pLink series vectors, and Destination Vectors used in this assembly were all gifts from Ron Weiss¹⁷³. The CAG promoter was then replaced with the EF1a promoter with restriction enzyme digest.

mMoClo Assembly of Integration Vectors

The mMoClo integration vectors were assembled through a Bpil-mediated Golden Gate reaction. Each 20 μ L reaction comprised 2 μ L 10x T4 ligase buffer, 2 μ L 10x BSA (1 mg/mL stock), 0.8 μ L Bpil-FD, 0.8 μ L T4 DNA Ligase (400 U/ μ L stock), 20 fmol integration vector backbone (pPD630), and 40 fmol of each transcription unit and linker plasmid to be inserted. The reaction was incubated at 37°C for 15 min, then subjected to 55 iterations of thermocycling (37°C for 5 min, 16°C for 3 min, repeat), followed by 37°C for 15 min, 50°C for 5 min, 80°C for 10 min to terminate the reactions; then the mixture was cooled to room temperature (optionally held at 4°C if the reaction ran overnight) and placed on ice prior to immediate transformation into bacteria.

Plasmid preparation

TOP10 E. coli were grown overnight in 100 mL of LB with the appropriate selective antibiotic. The following morning, cells were pelleted at 3000 x g for 10 min and then resuspended in 4 mL of a solution of 25 mM Tris pH 8.0, 10 mM EDTA, and 15% sucrose. Cells were lysed for 15 min by addition of 8 mL of a solution of 0.2 M NaOH and 1% SDS, followed by neutralization with 5 mL of 3 M sodium acetate (pH 5.2).

Precipitate was pelleted by centrifugation at 9000 x g for 20 min. Supernatant was decanted and treated with RNase A for 1 h at 37°C. 5 mL of phenol chloroform was added, and the solution was mixed and then centrifuged at 7500 x g for 20 min. The aqueous layer was removed and subjected to another round of phenol chloroform extraction with 7 mL of phenol chloroform. The aqueous layer was then subjected to an isopropanol precipitation (41% final volume isopropanol, 10 min at room temperature, 9000 x g for 20 min), and the pellet was briefly dried and resuspended in 420 µL of water. The DNA mixture was incubated on ice for at least 12 h in a solution of 6.5% PEG 20,000 and 0.4 M NaCl (1 mL final volume). DNA was precipitated with centrifugation at maximum speed for 20 min. The pellet was washed once with ethanol, dried for several h at 37°C, and resuspended for several h in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA purity and concentration were confirmed using a Nanodrop 2000 (Thermo Fisher).

Oxygen sensing system

Oxygen pressure at the bottom of the well was assessed by the PreSens SDR SensorDish system (Applikon Biotechnology), using a 24-well plate.

Microscopy

Microscopy was conducted on a Keyence BZ-X800E fluorescence microscope. Cells were maintained at 37°C and supplied with air containing 5% CO₂, 21% O₂, and 74% N₂. Filter sets were purchased from Chroma in Keyence BZX cubes (#91056). Filter sets for microscopy are described in **Supplementary Table A4.1**. Experiments for microscopy were conducted in Phenol Red-free DMEM (Sigma DMEM Powder D2909, supplemented with 3.7 g/L sodium bicarbonate, 3.5 g/L glucose, and 1 mL/L of a pyridoxine-HCl (4 mg/mL) and sodium phosphate (16 mg/mL) solution; then Fetal Bovine Serum, L-glutamine, and Penicillin-Streptomycin as described in **HEK293FT cell culture**) and, if transfected, were transfected with Lipofectamine LTX with PLUS Reagent (ThermoFisher 15338100).

HEK293FT cell culture

The HEK293FT cell line was purchased from Thermo Fisher/Life Technologies (RRID: CVCL_6911

[\[https://web.expasy.org/cellosaurus/CVCL_6911\]](https://web.expasy.org/cellosaurus/CVCL_6911)). Cells were cultured in DMEM (Gibco #31600-091) with 4.5 g/L glucose (1 g/L, Gibco #31600-091; 3.5 g/L additional, Sigma #G7021), 3.7 g/L sodium bicarbonate (Fisher Scientific #S233), 10% FBS (Gibco #16140-071), 6 mM L-glutamine (2 mM, Gibco #31600-091; 4 mM additional, Gibco #25030-081), penicillin (100 U/ μ L), and streptomycin (100 μ g/mL) (Gibco #15140122), in a 37°C incubator with 5% CO₂. Cells were subcultured at a 1:5 to 1:10 ratio every 2–3 d using Trypsin-EDTA (Gibco #25300-054). The HEK293FT cell line tested negative for mycoplasma with the MycoAlert Mycoplasma Detection Kit (Lonza #LT07-318).

B16F10 cell culture

B16F10 cells were cultured in DMEM (Gibco #31600-091) with 4.5 g/L glucose (1 g/L, Gibco #31600-091; 3.5 g/L additional, Sigma #G7021), 3.7 g/L sodium bicarbonate (Fisher Scientific #S233), 10% FBS (Gibco #16140-071), 6 mM L-glutamine (2 mM, Gibco #31600-091; 4 mM additional, Gibco #25030-081), penicillin (100 U/ μ L), and streptomycin (100 μ g/mL) (Gibco #15140122), in a 37°C incubator with 5% CO₂. Cells were subcultured at a 1:10 to 1:20 ratio every 2–3 d using Trypsin-EDTA (Gibco #25300-054).

HEK293FT Transfection based assays (calcium phosphate)

Transient transfection of HEK293FT cells was conducted using the calcium phosphate methodology. Cells were plated at a minimum density of 1.5×10^5 cells per well in a 24-well plate in 0.5 mL DMEM, supplemented as described above. For surface staining experiments, cells were plated at a minimum density of 3.0×10^5 cells per well in a 12-well plate in 1 mL DMEM. After at least 6 h, by which time the cells had adhered to the plate, the cells were transfected. For transfection, plasmids were mixed in H₂O, and 2 M CaCl₂ was added to a final concentration of 0.3 M CaCl₂. This mixture was added dropwise to an equal-volume solution of 2 \times HEPES-Buffered Saline (280 mM NaCl, 0.5 M HEPES, 1.5 mM Na₂HPO₄) and gently pipetted up and down four times. After 2.5–4 min, the solution was mixed vigorously by pipetting ten times. 100 μ L of this mixture was added dropwise to the plated cells in 24-well plates, 200 μ L was added to the plated cells in 12-well plates, or 400 μ L was added to the plated cells in 6-well plates, and the plates were gently swirled. The next morning, the medium was aspirated and replaced with fresh medium. In some

assays, fresh medium contained cobalt(ii) chloride, as described in the figures. Typically, at 36–48 h post-transfection and at least 24 h post-media change, cells were harvested.

HEK293FT Transfection based assays (Lipofectamine)

From exponentially growing HEK293LP cells, 1.0×10^5 cells were plated per well (1.0 mL medium) in 12-well format, and cells were cultured for 24 h to allow cells to attach and spread. When cells reached 50–75% confluence, plasmids were transfected by lipofection with Lipofectamine LTX with PLUS Reagent (ThermoFisher 15338100). Plasmids were mixed with 1.0 μ L of PLUS reagent in a 50 μ L total volume reaction, with the remainder of the volume being OptiMEM (ThermoFisher/Gibco 31985062). In a separate tube, 3.8 μ L of LTX reagent was mixed with 46.2 μ L of OptiMEM. The DNA/PLUS Reagent mix was added to the LTX mix. pipetted up and down four times, and then incubated at room temperature for 5 min. 100 μ L of this transfection mix was added drop-wise to each well of cells, which was mixed by gentle swirling. Cells were incubated in normoxia overnight and then cultured for 1-2 d in normoxia or hypoxia prior to microscopy. While the 12-well plate format was used for the oxidation/microscopy experiment, a 24-well plate format was used for the oxidation/harvest method experiment with half of the number of cells, volumes of media and reagents, and masses of DNA as listed above.

Experiment involving B16F10 landing pad cell line

Stable cell lines were plated in 0.750 mL of DMEM in triplicate in 24-well format at a density expected to generate 5% confluent wells and placed in a normoxic or hypoxic incubator. Cells were harvested 3 days later for flow cytometry analysis.

Experiments involving HEK293FT landing pad cell lines

For the 5 d time course assay, stable cell lines were plated in 1 mL of DMEM in triplicate in 24-well format with 2×10^4 cells/well. Some lines were plated in medium 150 μ M CoCl₂ or 1 μ g/mL doxycycline. Cells were cultured under normoxic or hypoxic conditions for 5 d, and each day (24 h), cells were harvested for flow cytometry analysis. In some experiments, cells were cultured in a different concentration of CoCl₂

or harvested at the time point indicated in the figure legend.

Harvesting cells for flow cytometry

Cells were harvested for flow cytometry using FACS buffer (PBS pH 7.4, 2–5 mM EDTA, 0.1% BSA) or using 0.05% Trypsin-EDTA (with or without Phenol Red) for 5 min followed by quenching with medium (with or without Phenol Red). The resulting cell solution was added to at least 2 volumes of FACS buffer. Cells were spun at 150×g for 5 min, supernatant was decanted, and fresh FACS buffer was added.

Flow cytometry assays

Flow cytometry was run on a BD LSR Fortessa Special Order Research Product (Robert H. Lurie Cancer Center Flow Cytometry Core). Lasers and filter sets used for data acquisition are listed in **Supplementary Table A4.2** (for experiments involving reporter expression). Samples were analyzed using FlowJo v10 software (FlowJo, LLC). Fluorescence data were compensated for spectral bleed-through. The HEK293FT and B16F10 cell populations were identified by SSC-A vs. FSC-A gating, and singlets were identified by FSC-A vs. FSC-H gating. To distinguish transfected from non-transfected cells, a control sample of cells was generated by transfecting cells with a mass of pcDNA (empty vector) equivalent to the mass of DNA used in other samples in the experiment. For the single-cell subpopulation of the pcDNA-only sample, a gate was made to identify cells that were positive for the constitutive fluorescent protein used as a transfection control in other samples, such that the gate included no more than 1% of the non-fluorescent cells. To distinguish cells expressing cargo from a landing pad, a sample of the corresponding parental, non-landing pad line was used.

Conversion of arbitrary units to standardized fluorescence units

To determine the conversion factor for MFI to MEFLs, Rainbow Calibration Particles (Spherotech, RCP-30-5) or UltraRainbow Calibration Particles (Spherotech URCP-100-2H) were run with each flow cytometry experiment. This reagent contains six (RCP) or nine (URCP) subpopulations of beads, each of a specific size and with a known number of various fluorophores. The total bead population was identified

by SSC vs. FSC gating, and the subpopulations were identified through two fluorescent channels. The MEFL values corresponding to each subpopulation were supplied by the manufacturer. A calibration curve was generated for the experimentally determined MFI vs. manufacturer supplied MEFLs, and a linear regression was performed with the constraint that 0 MFI equals 0 MEFLs. The slope from the regression was used as the conversion factor, and error was propagated. A similar process was done for fluorophores obtained in the PE-Texas Red (MEPTRs) and Pacific Blue (MEPB) channels.

HK293FT Landing pad integration and selection

From exponentially growing HEK293LP cells, 0.5×10^5 cells were plated per well (0.5 mL medium) in 24-well format, and cells were cultured for 24 h to allow cells to attach and spread. When cells reached 50–75% confluence, Bxb1 recombinase was co-transfected with the integration vector by lipofection with Lipofectamine LTX with PLUS Reagent (ThermoFisher 15338100). 300 ng of BxB1 expression vector was mixed with 300 ng of integration vector and 0.5 μ L of PLUS reagent in a 25 μ L total volume reaction, with the remainder of the volume being OptiMEM (ThermoFisher/Gibco 31985062). In a separate tube, 1.9 μ L of LTX reagent was mixed with 23.1 μ L of OptiMEM. The DNA/PLUS Reagent mix was added to the LTX mix. pipetted up and down four times, and then incubated at room temperature for 5 min. 50 μ L of this transfection mix was added drop-wise to each well of cells, which was mixed by gentle swirling. Cells were cultured until the well was ready to split (typically 3 d), without any media changes.

Cells were harvested from the 24-well plate when confluent by trypsinizing and transferring to a single well of a 6-well plate in 2 mL of medium, and then cells were cultured until they reached 50-70% confluence. Then, medium was aspirated and replaced with 2 mL of fresh media containing appropriate selection antibiotic 1 μ g/mL puromycin (Invivogen ant-pr) and/or 6 μ g/mL blasticidin (Alfa Aesar/ThermoFisher J61883). Medium was replaced daily with fresh medium containing antibiotics until cell death was no longer evident. Selection was first performed in puromycin for 7 d, then cells were expanded for 7 d without antibiotics. Cells were then cultured in both puromycin and blasticidin to maintain selective pressure until flow sorting. Cells were sorted as described for each line and in *Flow sorting*.

B16F10 Landing pad integration and selection

From exponentially growing HEK293LP cells, 0.8×10^5 cells were plated per well (0.5 mL medium) in 12-well format, and cells were cultured for 24 h to allow cells to attach and spread. Bxb1 recombinase was co-transfected with the integration vector by lipofection with Lipofectamine LTX with PLUS Reagent (ThermoFisher 15338100). 200 ng of BxB1 expression vector was mixed with 200 ng of integration vector and without PLUS reagent in a 50 μ L total volume reaction, with the remainder of the volume being OptiMEM (ThermoFisher/Gibco 31985062). In a separate tube, 5 μ L of LTX reagent was mixed with 45 μ L of OptiMEM. The DNA/PLUS Reagent mix was added to the LTX mix, pipetted up and down four times, and then incubated at room temperature for 5 min. 100 μ L of this transfection mix was added drop-wise to each well of cells, which was mixed by gentle swirling. Cells were cultured until the well was ready to split (typically 3 d), without any media changes.

Cells were harvested from the 12-well plate when confluent by trypsinizing and transferring to a single well of a 6-well plate in 2 mL of medium with appropriate selection antibiotic, and then cells were cultured until they reached 50-70% confluence, with frequent trypsinization to remove dead cells. Antibiotic concentrations were 1 μ g/mL puromycin (Invivogen ant-pr) and 15 μ g/mL blasticidin (Alfa Aesar/ThermoFisher J61883). Medium was replaced daily with fresh medium containing antibiotics until cell death was no longer evident. Cells were then cultured in both puromycin and blasticidin to maintain selective pressure until assay or flow sorting. Cells were sorted as described for each line generation and in **Flow cytometry-based cell sorting**.

B16F10 Landing pad line generation

From exponentially growing B16F10LP cells, 4×10^5 cells were plated per well (2 mL medium) in 6-well format, and cells were cultured for 24 h to allow cells to attach and spread. Cells were transfected with 160 ng each of pPD720, pPD783, and pPD864, 1520 ng of pPD005, in 100 μ L total volume (balance OptiMEM), mixed with 100 μ L of OptiMEM containing 10 μ L of Lipofectamine LTX (total volume of 200 μ L transfection mixture per well). Beginning 3 d after transfection, cells were selected with 1000 μ g/mL Hygromycin (20 μ L stock / mL of medium). EYFP+, single cells were sorted 18 d after transfection into a

96-well plate. Wells were visually verified to contain only 1 cell per well. Cell lines were expanded for 2-4 weeks under continuous antibiotic selection. Approximately 60 monoclonal lines were generated.

B16F10 landing pad line validation

Validation was performed by genomic PCR through the right Rosa26 homology arm. Lines were also assessed for their ability to maintain EYFP fluorescence for 6 weeks without antibiotic selection pressure and for the performance of an integrated circuit.

Flow cytometry-based cell sorting

Cells were harvested by trypsinizing, resuspended at approximately 10^7 cells per mL in pre-sort medium (DMEM with 10% FBS, 25 mM HEPES (Sigma H3375), and 100ug/mL gentamycin (Amresco 0304)), and held on ice until sorting was performed. Cells were sorted using one of several BD FACS Aria Special Order Research Products (Robert H. Lurie Cancer Center Flow Cytometry Core) with the optical configuration listed in **Supplementary Table A4.3**. Cells were collected for each line in post-sort medium (DMEM with 20% FBS, 25 mM HEPES, and 100 μ g/mL gentamycin), and cells were held on ice until they could be centrifuged at 150 x g for 5 min and resuspended in DMEM. Cells were plated and expanded until used in experiments. Gentamycin was included in the culture medium for one week after sorting. For monoclonal cell sorting, cells were sorted directly into 96-well plates and maintained in post-sort medium until adherent, at which point the medium was changed.

4.5 Results

Validating methods for studying hypoxia

We first validated several methodologies for the study of hypoxia in our lab. As oxygen must diffuse out of the media for the culture to become hypoxic, we first studied the rate at which this would happen, using a specialized plate with an electronic oxygen sensor at the bottom of the well (**Fig. 4.1**)^{273, 274}. We

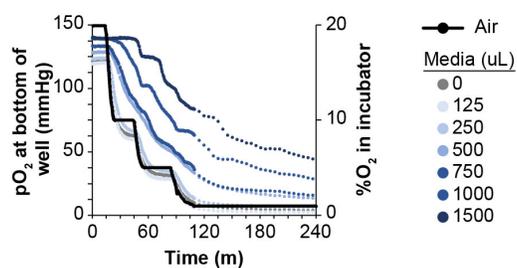


Fig 4.1 Oxygen diffusion through various volumes of media. Oxygen pressure was measured at the bottom of a 24-well plate containing various volumes of DMEM in each well, as the desired oxygen concentration in the incubator was adjusted stepwise down to 1%.

found that in a 24-well plate, volumes of 250 uL or less rapidly equilibrated, the oxygen pressure in volumes of media between 500 uL and 750 uL lagged behind the pressure in the air slightly, and volumes larger than this took hours longer to equilibrate. Therefore, in future assays, we used volumes that did not exceed the equivalent of 750 uL in a 24-well plate.

Another concern with studying hypoxia using fluorescent proteins as reporters is that fluorescent protein maturation requires an oxidation step for most fluorophores (the biliverdin-based iRFPs are an exception), so we evaluated whether this would be a limitation for several fluorophores (**Fig. 4.2**)^{275, 276}. Cells were transfected with fluorescent protein expressing plasmids, then cultured under hypoxic conditions, and then serially imaged in a microscope under normoxic conditions, at 37°C and 5% CO₂; these cells were compared to cells that were cultured continuously under normoxic conditions after transfection. We found that while mRFP670 and mRFP720 showed similar brightness after hypoxic and normoxic culture, other fluorescent proteins require various lengths of time to mature. mTagBFP2 and mNeonGreen oxidized rapidly in under an hour, EYFP, EBFP2, and DsRed-Express2 required about two hours to oxidize, and mRuby3 required several hours. Notably, the oxidized fluorophores appeared slightly dimmer than those cultured under normoxic conditions. As oxygen diffusion out of the media has a slight lag time, oxygen diffusion into the media likely also lags, meaning that in this assay, fluorophores such as mNeonGreen and mTagBFP2 were capable of oxidation at pressures of oxygen less than that of room air.

We next studied under what conditions oxidation could occur for the fastest maturing fluorophore of each color, as this would determine the preparation of samples in each assay. A process in which samples mature during the time between harvest and analysis is prone to bias if this time is different for each sample. Cells were co-transfected with fluorescent protein expressing plasmids and cultured under normoxia or hypoxia, then subject to various treatments prior to analysis (**Fig. 4.3a**). Compared to cells cultured under normoxia, cells cultured under hypoxia and immediately taken for flow cytometric analysis after harvest, either on ice or at room temperature, showed lower levels of fluorescence, and this effect was more pronounced for DsRed-Express2. If cells were allowed to oxidize for 2h in the FACS buffer after harvest, oxidation happened to varying degrees and was more complete at higher temperatures. In contrast, cells cultured under hypoxia and then allowed to oxidize under normoxic culture conditions for 2

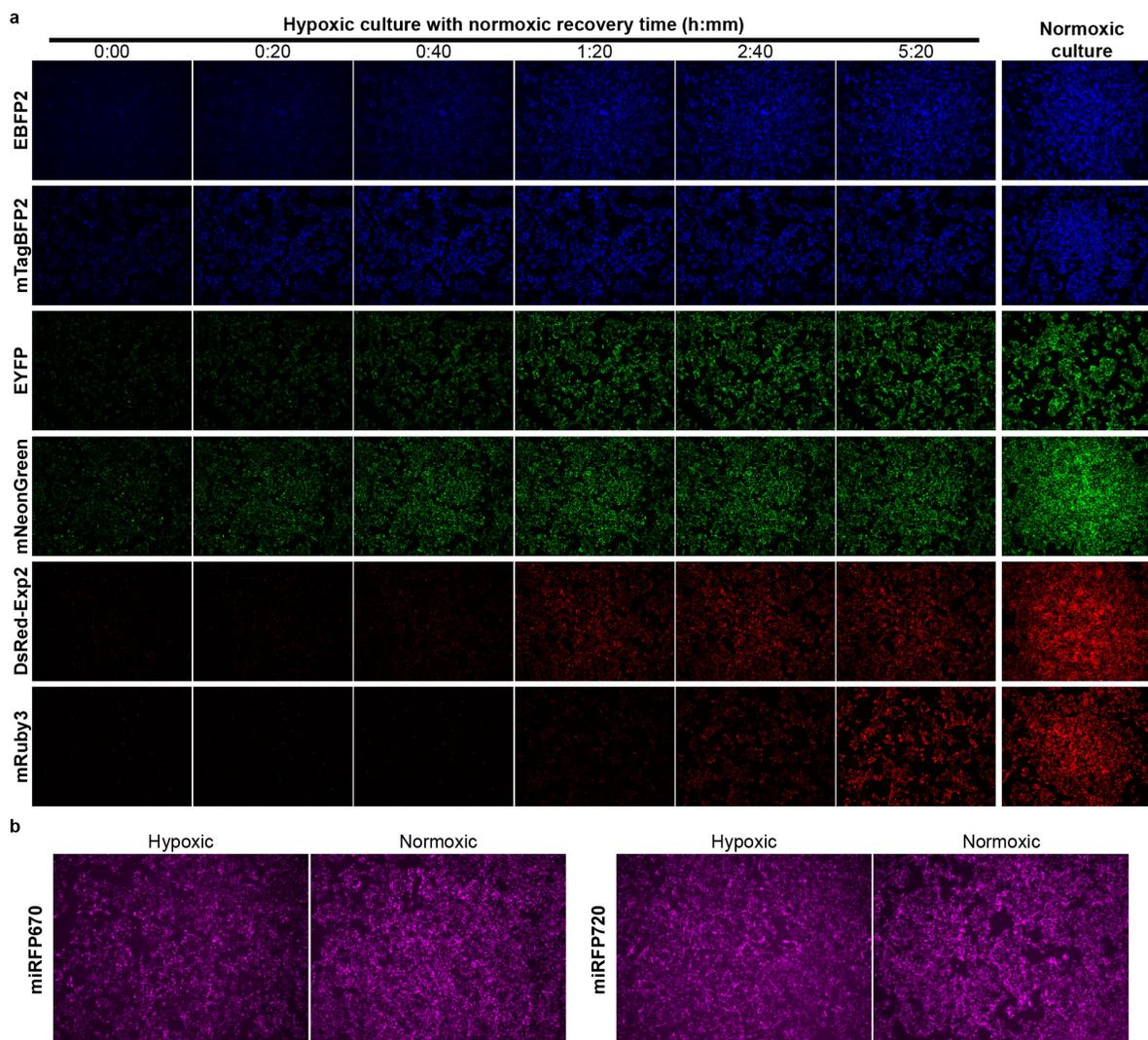


Fig 4.2 Oxidation rates of fluorescent proteins. a) Fluorescent time-lapse micrographs of cells expressing various fluorescent proteins. Microscopy was performed under normoxic conditions for cells that had been cultured under either hypoxic or normoxic conditions after transfection. b) Fluorescent micrographs of cells expressing iRFPs after hypoxic or normoxic culture after transfection.

h prior to harvest showed complete oxidation of mTagBFP2, nearly complete oxidation of mNeonGreen, and 80% complete oxidation of DsRed-Express2. Cells that were fixed with paraformaldehyde after harvest and then oxidized at various temperatures showed complete oxidation of mTagBFP2, near complete oxidation of mNeonGreen, and only minimal oxidation of DsRed-Express2. Given the timescales over which the oxidation of these fluorescent proteins occurs (**Fig. 4.2a**), it is likely that the oxidation that did happen occurred during the processing prior to complete fixation, and that the fixation inhibited the oxidation from progressing. Further evidence for this is seen in that the fixed samples do not show the same temperature-dependent oxidation of the DsRed-Express2 samples that the non-fixed samples do. Therefore, the protocol for all assays included a 2 h period of normoxic culture after the hypoxic culture was completed, in order to allow for maturation of the fluorescent proteins.

When then validated these findings with doxycycline-inducible expression of DsRed-Express2 from a single-copy, genomically integrated locus in a HEK293FT-LP line. Cells were cultured under hypoxia or normoxia for five days with or without doxycycline and sample was harvested each day for analysis (**Fig. 4.3b**). If our protocol was valid, we would expect to see only a minimal difference between these populations of cells. The DsRed-Express2 levels from the hypoxic and normoxic conditions were largely similar over the course of 5 days, with the slight difference observed possibly attributable to minimally incomplete oxidation or a small difference in protein expression levels in hypoxic conditions. Thus, our oxidation and harvest protocols are valid and any differences in the expression of the reporter protein we observe in future assays can be mostly attributed to differences in activity of its promoter, rather than an artifact.

Evaluating the performance of genomically integrated HBS in a landing pad

We next integrated an HBS construct into the genome of the HEK293FT-LP line (**Fig. 4.4a**). This HBS was built on the YB_TATA minimal promoter, which has previously been shown to confer low levels of background signaling, downstream of an optimized set of HREs. This construct constitutively expressed EBFP2, a gene for blasticidin resistance, and a gene for puromycin resistance. The HBS controlled expression of a DsRed-Express2 fluorescent protein. After integration, cells were selected with antibiotics for two weeks and then expanded prior to the experiment. To test the activity of the HBS, cobalt(II) chloride,

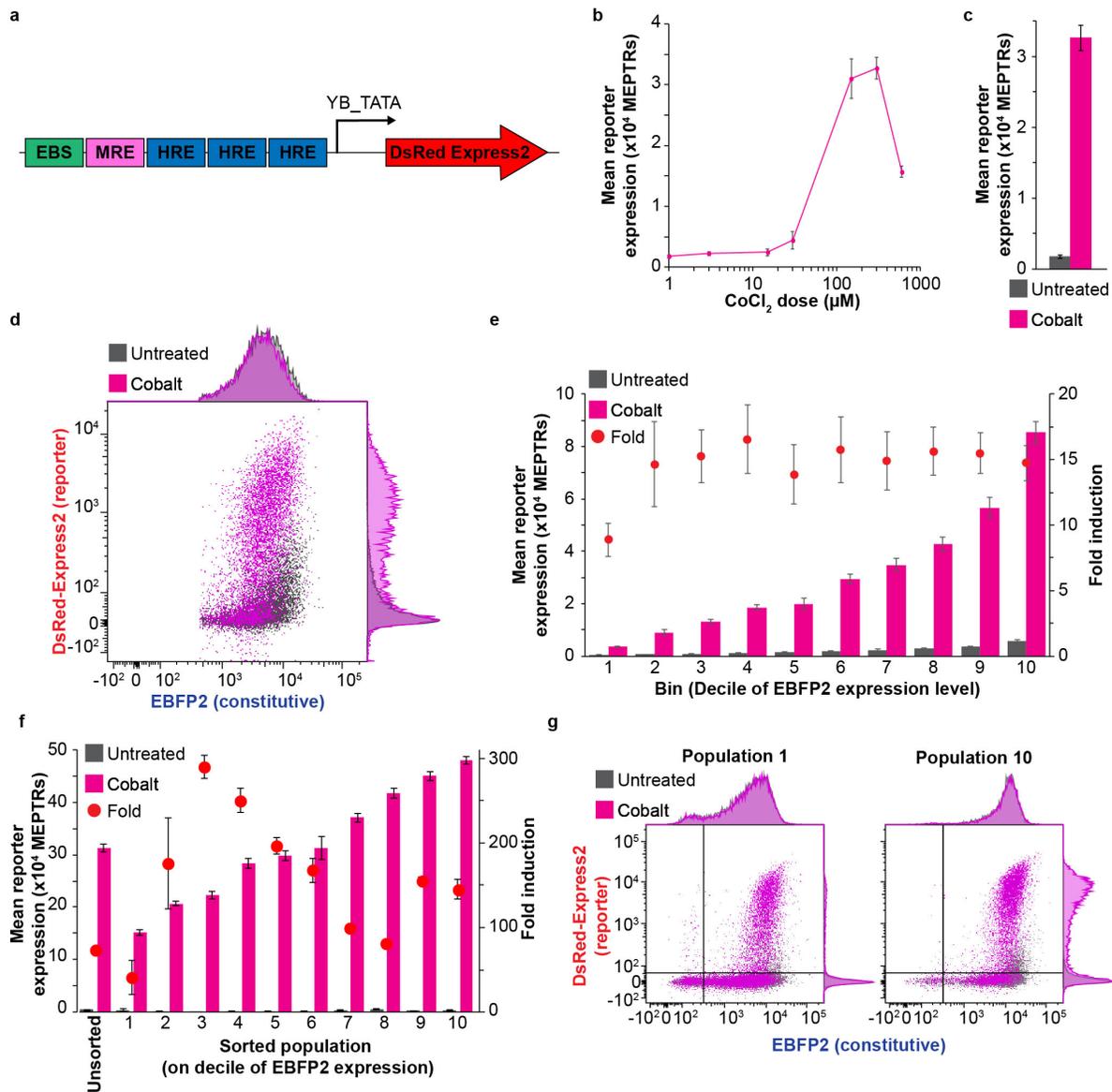


Fig 4.4 Evaluating the performance of a genomically integrated HBS. a) Schematic depicting the HBS. b) Activation of the HBS at various concentrations of CoCl₂, a hypoxia mimetic. c) Activation of the HBS with a hypoxia mimetic d) Flow cytometry analysis of the HBS activated with a hypoxia mimetic. e) Subpopulation analysis of cells expressing various levels of EBFP2 from the experiment in c and d. f,g) Analysis of cell lines sorted on EBFP2 expression level, with representative flow cytometry plots.

a hypoxia mimetic, was used (**Fig 4.4b**)²⁷⁷. Cobalt induced the hypoxia biosensor, leading to a 19-fold increase in reporter expression (**Fig 4.4c**). However, this response was homogenous, and many cells did not show any reporter expression (**Fig 4.4d**). The expression level of the DsRed-Express2 reporter and the expression level of the EBFP2 constitutively expressed gene trended the same way—cells with higher constitutive protein expression levels were more likely to express the reporter. An analysis of this data, in which cells were divided into ten populations based on their EBFP2 expression showed the same trend—cells with higher EBFP2 showed higher reporter expression, without change in fold induction (**Fig 4.4e**). We therefore postulated that sorting cells based on their EBFP2 levels could help to make the population more homogenous. Even though landing pad cells are genetically identical, perhaps there were some heritable, epigenetic effects going on at this locus.

We sorted the cell line based on EBFP2 expression, cultured the cells at a low density for two weeks, and then cultured them with or without cobalt. We saw the similar trends as expected—the reporter expression increased with the expression level of the EBFP2 that the cells had at the time of sorting (**Fig 4.4f**). Surprisingly, however, this was not driven by an increase in the mode of the reporter-expressing population but rather an increase in the fraction of cells that were turning on; a similar trend was noticeable for the EBFP2—cells with higher EBFP2 at the time of sorting were less likely to be silenced (**Fig 4.4g**). Therefore, sorting the integrated, landing pad cells for those most highly expressing the cargo is a strategy for improving the performance of the landing pad system, likely by isolating the cells less likely to undergo genetic silencing, thereby generating a more homogenous population.

Evaluating the role of minimal promoters

As the choice of minimal promoter is a key determinant of DNA-based biosensor performance and some choices can enhance performance, by decreasing the background more than the induced signaling or vice-versa⁸⁷, we next evaluated this for a several choices of minimal promoters (**Fig. 4.5a**). We did so across two cell lines, HEK293-FT and B16F10 (**Fig. 4.5b**). Though the SV40_min did not lead to a useful HBS in the HEK293-FT cells, likely due to the presence of the SV40 large T antigen in this line (the SV40_min and the SV40 origin of replication have sequence homology), the YB_TATA showed a very low

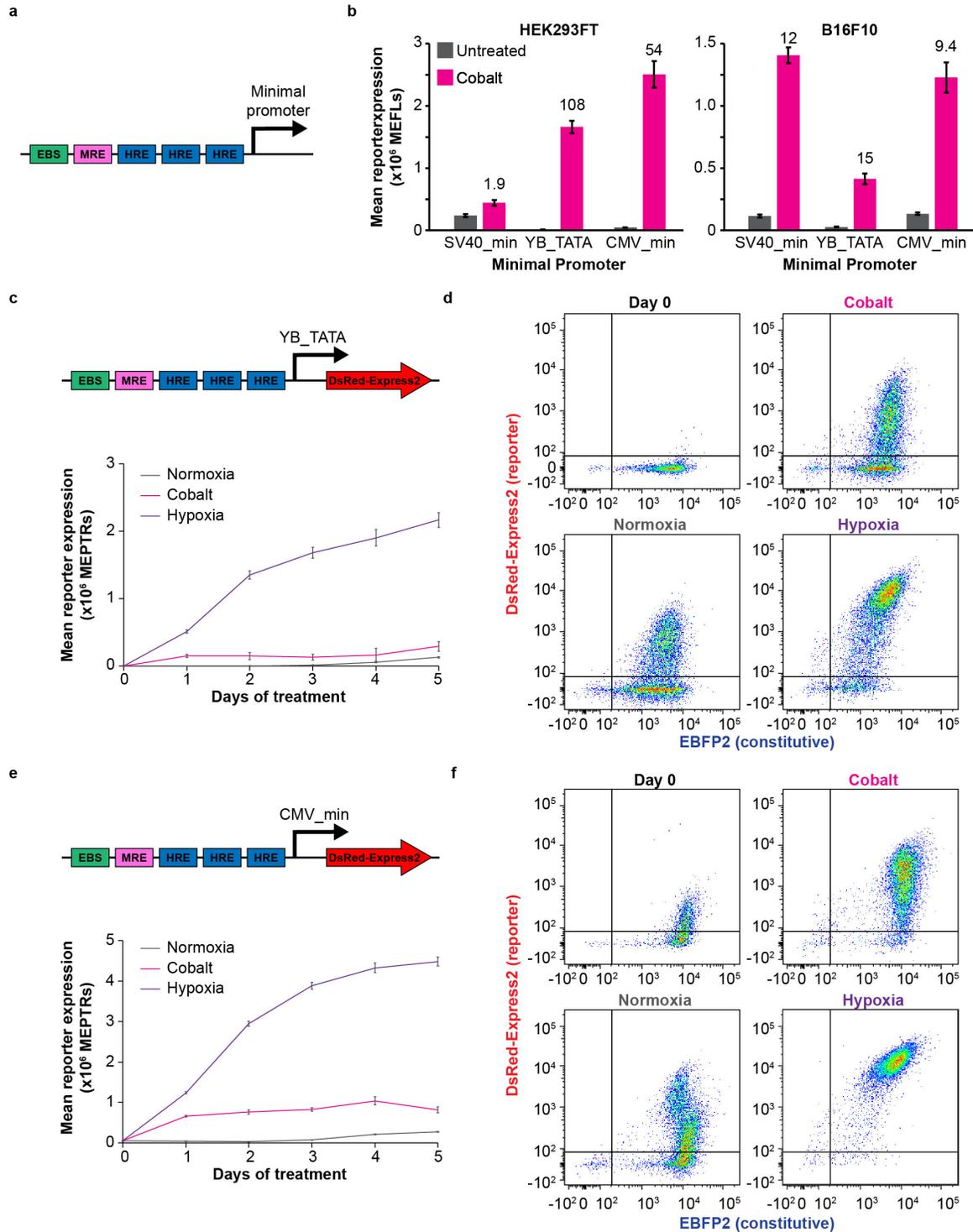


Fig 4.5 Evaluating the choice of minimal promoter for genomically integrated HBS lines. a) Schematic depicting minimal promoter choice. b) Evaluation of minimal promoter choice for the HBS in two cell lines in transient transfection. c-f) Time-course assay of HEK293FT-LP lines with integrated HBS with different minimal promoters. Flow cytometry plots in d and f are representative samples.

background signal leading to a fold induction of 108 and the CMV_min slightly increased the induced and background signal, leading to decreased fold induction of 54. In the B16F10 line, the YB_TATA minimal promoter showed much lower background signal and lower induced signal than did the SV40_min or CMV_min promoters, leading to a fold induction of 15. These trends agree with those previously described for these minimal promoters, and in both cases, the choice of the minimal promoter improved the performance of the biosensor above that of the SV40_min with which this HRE configuration was originally characterized.

We next integrated the HBS constructs into the genome of HEK293FT-LP cells; the SV40_min construct could not be integrated, despite repeated attempts, likely due to the presence of the SV40 large T antigen. After 2 weeks of antibiotic selection, the resulting lines were sorted for the top 10% of EBFP2 expressing cells and expanded prior to experimentation. To compare the performance of the promoters, cells were cultured for 5 days in hypoxia, in normoxia, or in normoxia with cobalt(II) chloride and some samples were harvested each day for flow cytometry. For both promoters, hypoxia induced a much higher signal from the HBS than cobalt, and some gene expression was seen after days of normoxic culture that was not present at the outset (**Fig. 4.5c,e**). This latter effect is likely due to cell overgrowth and resultant hypoxia rather than leaky sensor performance, as cells cultured at high densities under normoxic conditions show HBS activation. The biosensor showed the greatest increase in activation over the first 3 days of hypoxia, with diminishing marginal improvement after that point. Flow cytometry plots showed that the CMV_min promoter had a much more homogenous response to hypoxia than the YB_TATA, but at a cost of more leaky gene expression (**Fig 4.5d,f**). Further, for both constructs, the stronger response to hypoxia than to cobalt was driven by both an increase in the fraction of cells expressing the reporter and the level of reporter expression of this population.

To determine how sensitive the HBS is to oxygen levels, we also cultured these lines at 5% O₂, a concentration that is often thought of as physoxia (**Fig. 4.6**). Surprisingly, we saw some activation of the biosensor at these concentrations over time, though at lower levels than in previous experiments with 1% O₂ (**Fig. 4.6a,d**). Flow cytometry showed that this was due to a low frequency of reporter-expressing cells (**Fig. 4.6b,e**). Microscopic analysis of these cultures revealed that most of this signal was coming from the

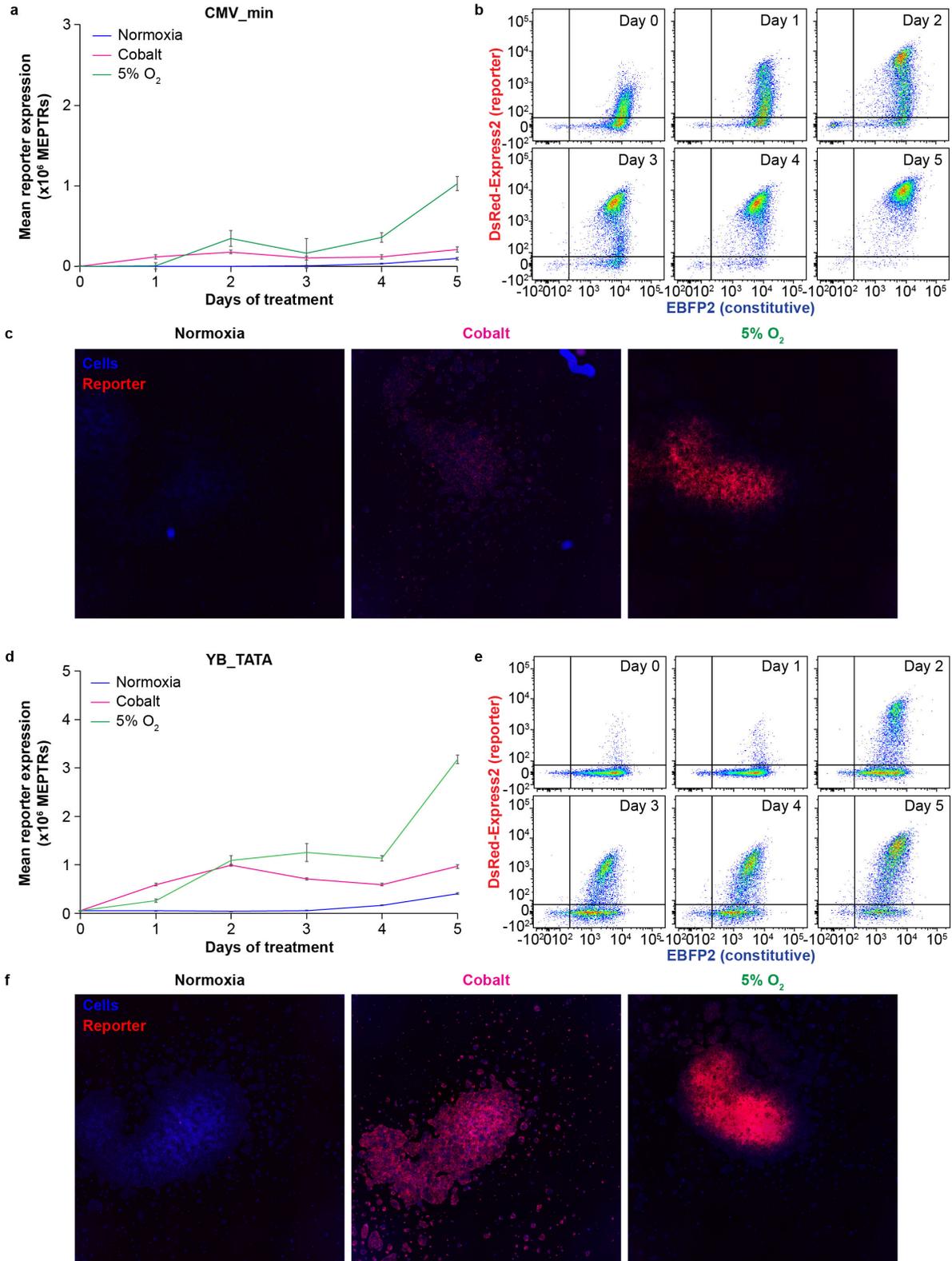


Fig 4.6 Evaluation of genomically integrated HBS under physoxic conditions. Time course analysis (a,d) and representative flow cytometry plots (b,e) of HEK293FT-LP lines with integrated HBS with different minimal promoters. c,f) Representative fluorescent micrographs of cells cultured under the indicated conditions shows localized activation of the HBS under physoxic conditions.

center of a clump of cells in the middle of the plate, as opposed to diffuse signaling observed with cobalt treatment (**Fig. 4.6c,f**). Thus, this probably represents a response to true hypoxia in overgrown cultures rather than a response to 5% O₂, and we would not expect the HBS to turn on at the 5% O₂ concentrations present throughout the human body.

In these experiments, while the YB_TATA-based construct led to a high fold induction, the level of induced signaling was relatively low compared to the CMV_min construct, indicating that we might be able to increase the level of hypoxia-signaling from the YB_TATA-based construct. To test this, we employed a mutant version of HIF1 α , that has previously been shown to induce gene expression in the presence of oxygen, as the proline residues that are normally oxidized have been mutated to alanine residues²⁶⁵. We placed this gene under the control of a doxycycline-responsive promoter and integrated a construct including this transcription unit, the HBS, and a constitutive EBFP2 gene into HEK293FT-LP cells (**Fig. 4.7a**). Following selection, flow sorting, and recovery, we cultured this line with cobalt, hypoxia, or doxycycline. Culture with doxycycline, leading to expression of the stable HIF1 α , led to higher levels of induced reporter gene expression than culture with cobalt or in the presence of hypoxia (**Fig. 4.7b**) and the resulting population was more homogenous (**Fig. 4.7c**). This indicates that in HEK293FT cells, the supply of HIF1 α limits HBS performance. Given this finding, we designed several feedback circuits that would produce more wild type HIF1 α or HIF2 α upon activation of the biosensor (**Fig. 4.7d**). Here, the use of positive feedback is preferable to constitutive expression of these components, as this might overwhelm the cell's ability to hydroxylate and degrade these components in the presence of oxygen, leading to nonspecific HBS activation. It is also preferable to positive feedback with the stable mutant components, as leaky expression of these from the HBS could lead to a self-propagating feedback loop, also leading to nonspecific activation.

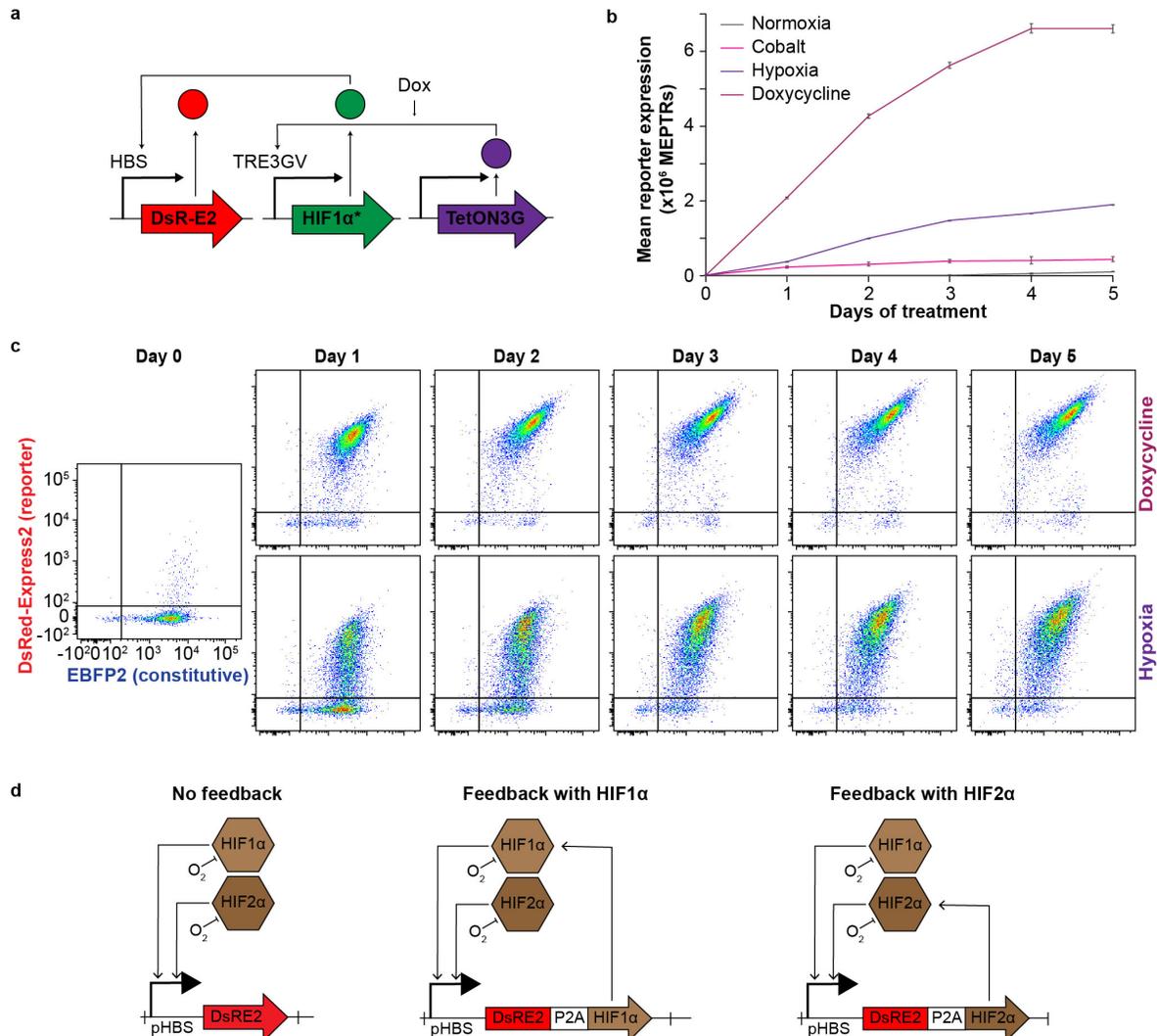


Fig 4.7 HBS activation with stable HIF1 α . a) Schematic depicting the doxycycline-induced production of stable HIF1 α and its action on the HBS. b,c) Analysis of HEK293FT-LP lines with doxycycline-inducible stable HIF1 α and HBS over a 5 d time course and representative flow cytometry plots. d) Schematic of possible positive feedback circuits that could be generated with HIFs.

Evaluating biosensor performance in a melanoma cell line

Given that the ultimate goal of this project is a system for the *in vivo* prototyping of biosensors, we decided to evaluate these feedback circuits in the B16F10 cell line, which is derived from a murine melanoma and readily generates tumors when subcutaneously implanted in immunocompetent mice. First, we generated a B16F10-LP line by integration of a landing pad into the Rosa26 safe harbor locus using Cas9-mediated integration, with simultaneous inhibition of NHEJ repair to increase efficiency (**Fig. 4.8a,b**). Fluorescent cells (those with a landing pad in the genome) were sorted into wells, with one cell per well, and monoclonal populations were generated. These were validated by flow cytometry, genomic PCR, and integration of a test circuit. A plasmid dosing regimen for integrating circuits into the B16F10-LP line was also optimized.

We then integrated YB_TATA-based HBS constructs, without and with positive feedback with murine HIF1 α and murine HIF2 α into the B16F10-LP and selected them with antibiotics for two weeks. After this, cells were cultured in normoxia or hypoxia for three days and analyzed via flow cytometry. In this line, the HBS showed some leaky gene expression and low induced signaling. Contrary to our expectation, neither positive feedback circuit increased gene expression—both actually yielded lower levels of gene expression than the HBS without positive feedback (**Fig 4.8c**).

Although we demonstrated in the HEK293-FT-LP line that the HIF1 α was a limiting reagent, the addition of HIF1 α or HIF2 α mediated positive feedback loops in the B16F10-LP line actually led to decreased levels of gene expression compared to the HBS without feedback. We have hypothesized several possible reasons for this. It is possible that this may be an artifact of the P2A mediated strategy for expressing the feedback components. DsRed-Express2 forms tetramers in cells; it is possible that any noncleaved P2A peptide may result in a HIF protein attached to four molecules of the reporter and potential degradation of the reporter during the two hours of normoxic maturation at the end of the experiment. Alternatively, it is possible that in the B16F10 cell line, HIF1 α and HIF2 α are not limiting reagents for the HBS. Finally, it is possible that in this cell line the exogenous HIF1 α and HIF2 α do not have enough HIF1 β to be active and are sequestering transcriptional machinery away from the HBS. Transfection of the stable human HIF1 α mutant into the B16F10 line with the HBS leads to less induction than transfection of these

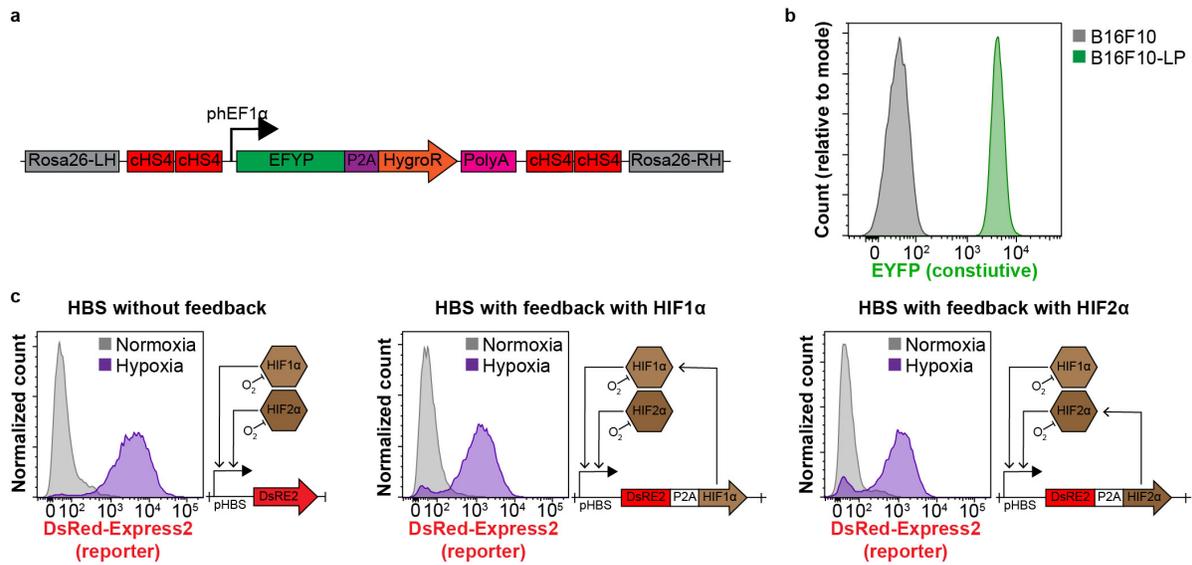


Fig 4.8 Evaluation of HBS feedback with HIFs in a B16F10-LP line. a) Schematic depicting a LP integrated into the Rosa26 safe harbor locus of the B16F10 cell line. b) Flow cytometry histograms of EYFP fluorescence from the B16F10-LP line. c) Schematics and preliminary evaluation of HIF-mediated feedback circuits in the B16F10-LP line.

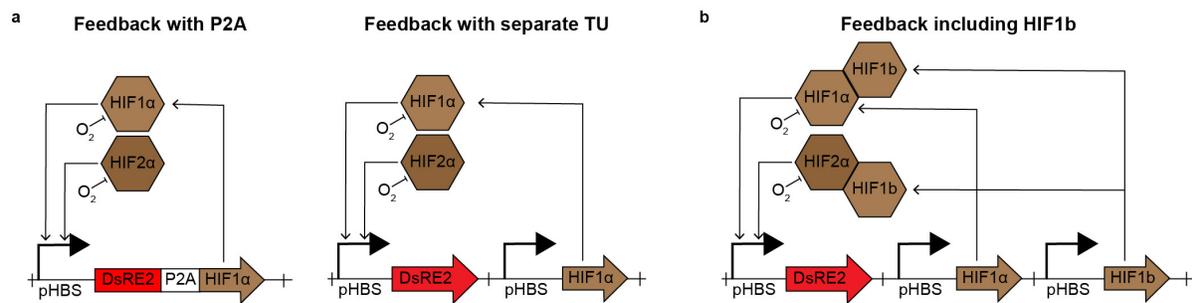


Fig 4.9 Possible issues with initial feedback circuits and potential solutions. a) Schematic depicting the differences between a P2A-based and separate-TU based strategy for feedback. b) Schematic depicting a new feedback circuit with additional HIF1 β .

cells with the HBS and culturing them with cobalt; it is unclear whether this experiment represents the toxicity HIF proteins in B16F10 cells, squelching, or the effects of a human HIF1 α in a murine cell (the two proteins are 70% identical²⁷⁸). For each of these issues, we have designed and constructed candidate circuits to test and potentially circumvent these issues (**Fig 4.9**).

Designing genetic circuits for modifying HBS output with COMET

Regardless of the outcome of the proposed studies, there may be some value in designing genetic circuits to modulate the response of the HBS that do not rely on the endogenous HIF proteins. Dysregulation of the HIF response is a vital step in the formation in many cancers and each is likely dysregulated differently; genetic circuits that rely on these endogenous components are therefore unlikely to be robust to the cell type in which they operate. To this end, we have designed feedback circuits that use COMET transcription factors (**Fig. 4.10**), with multiple different designs likely to lead to amplification of or positive feedback for the HBS, leading to increased hypoxia-induced signaling.

One concern with these circuits, as we discussed for the circuits that rely on the HIF proteins, is that amplification or positive feedback may amplify leaky gene expression or, in the case of positive feedback, lead to a circuit that is always induced. One way to mitigate this is to develop COMET TFs that are sensitive to the presence of oxygen (**Fig. 4.11a**). We tested whether appending a small amino acid motif from HIF1 α that has previously been used to confer oxygen dependent degradation of a fluorescent dye molecule onto a COMET ZFa would achieve this goal^{279, 280}. We tested several locations and copy number repeats for these tags in HEK293FT cells and found that while all conferred oxygen dependent gene expression, the 2x-C-terminal repeat did so the best (**Fig. 4.11b**). These new oxygen-unstable transcription factors also displayed conditional expression when evaluated in B16F10 cells (**Fig. 4.11c**). Interestingly, these two cell types showed different trends in response to hypoxia—while the transiently transfected HEK293FT cells showed slightly lower constitutive gene expression and ZFa-induced gene expression upon hypoxic culture, the B16F10 cells showed increased expression levels after hypoxia culture (**Fig. 4.11d**). One possible explanation is that the B16F10 cells do not proliferate as rapidly under hypoxic conditions and are therefore accumulating protein that is not diluted out upon cell division. This is

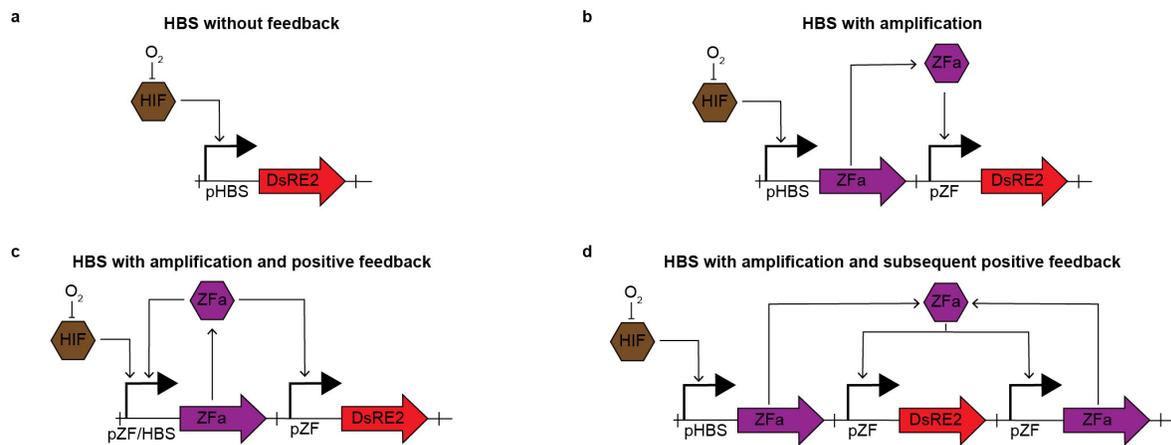


Fig 4.10 Genetic circuits for HBS modification with COMET. Schematics depicting a) the HBS without modification b) the HBS with amplification c) the HBS with amplification and positive feedback d) the HBS with amplification and subsequent positive feedback

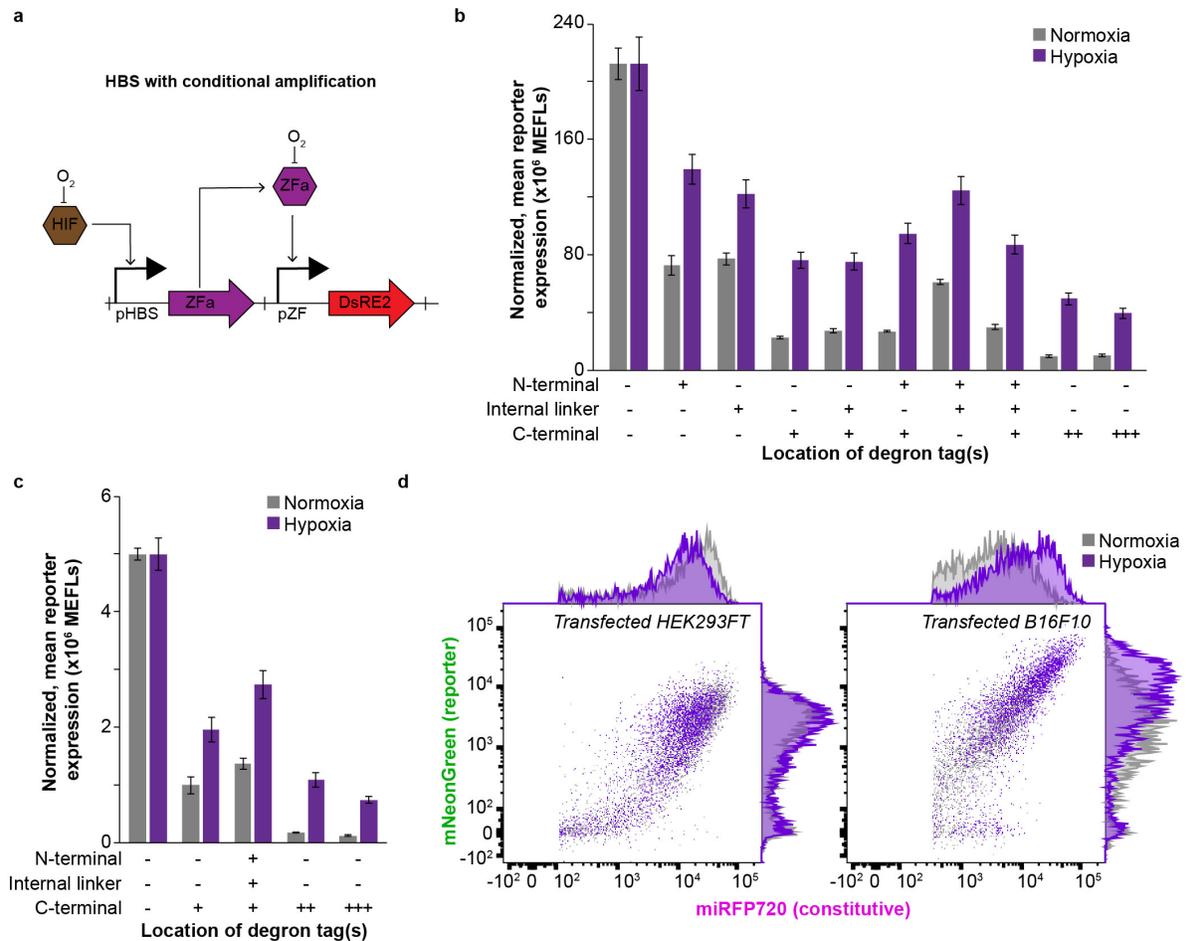


Fig 4.11 Conferring oxygen sensitivity to COMET TFs. a) Schematic depicting how oxygen sensitivity could be added to a COMET TF to address concerns regarding the amplification of leaky gene expression. b) Evaluating COMET TFs with oxygen degradation motifs in various locations and copy numbers in HEK293FT cells c) Evaluating the top performing COMET TFs in B16F10 cells d) Representative flow cytometry plots showing changes in constitutive (miRFP720) and COMET TF-induced reporter (mNeonGreen) gene expression in two cell lines.

possibly an artifact of transfection, as we did not observe this in the B16F10-LP line. The use of oxygen sensitive COMET TFs also would allow for sensing hypoxia while avoiding sensing HIF levels, potentially allowing avoiding the differential dysregulation between cell types, and producing a more robust biosensor with broad applications (**Fig. 4.12**). These circuits have been built and are awaiting testing. In summary, COMET TFs are amenable to oxygen-mediated degradation, and this property may confer beneficial effects when used to modulate the signal from a DNA-based HBS or when an oxygen-sensitive COMET TF is used as a stand-alone HBS.

4.6 Discussion

In this study, we first evaluated several methodologies for studying hypoxia. As we demonstrated, it is important that protocols new to a laboratory are developed and validated prior to use, as we uncovered several factors potentially confounding the study of hypoxia, as have been noted by others^{273, 274}. First, the time for media to acclimate to hypoxic conditions must be considered and this balanced with the ability to have enough media in the well for the cells to survive through the experiment, or methods must be developed to allow for media changes with de-oxygenated media under hypoxic conditions, in special cabinets designed for this work. Second, following culture under hypoxic conditions, the samples must be treated in such a way that the fluorescent proteins have sufficient time to mature in all samples. Failure to do so, thus allowing differential maturation times on a per sample basis, will lead to biased results. Although protein degradation, a first-order process, may occur during this time in a confounding manner, we have not observed any evidence of degradation on this time scale, and fluorescent proteins are generally regarded to be quite stable. Third, samples must be grown evenly distributed over the plate—clumps of cells become hypoxic at their core, causing reporter protein expression in response to true hypoxia rather, confounding interpretation of the response of the biosensor to the set oxygen concentration. This experiment also highlights the value of single cell measurements (such as by flow cytometry or microscopy) over those that take bulk populations values (*i.e.*, luciferase or SEAP plate reader assays), in which such trends, artifact or not, would be missed.

As other groups have noted⁸⁷, the minimal promoter is an important choice when designing

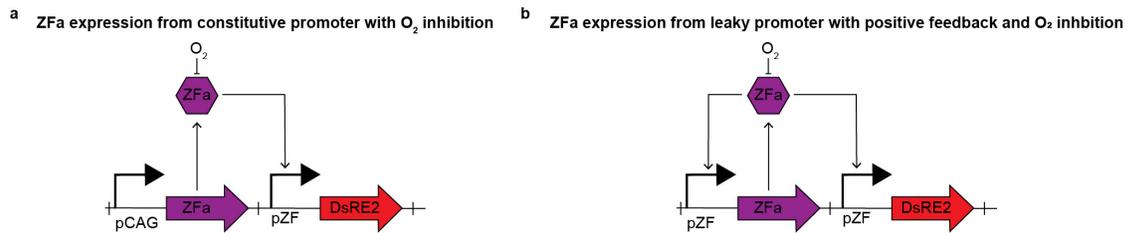


Fig 4.12 Hypoxia-sensing genetic circuits with COMET instead of HIFs. a) Schematic depicting an oxygen sensitive COMET TF that is constitutively produced b) Schematic depicting an oxygen sensitive COMET TF that is produced by expression from its cognate promoter, with positive feedback.

biosensors. Within this study, the effects of minimal promoter choice qualitatively held between our two cell lines, the HEK293FT and B16F10, with one exception that is readily explained. The SV40_min showed high background and low hypoxia induced signal in the HEK293FT line, and HBS constructs containing this minimal promoter could not be integrated in the LP—this is likely due to HEK293FT cells expressing the SV40 large T antigen and this protein acting on the minimal promoter, which bears high sequence similarity to the SV40 origin of replication. Though the SV40 large T antigen was synthetically introduced into the HEK293 line to make the HEK293T line, this result highlights how endogenous components can interfere with the performance of biosensors between cell types. Excitingly, however, the background signaling with the YB_TATA minimal promoter was very low, which allowed for very high fold inductions, though at the expected cost of a low hypoxia-induced signal. A key finding from this study was that the low levels of HIF1 α in HEK293FT limit performance of the HBS. Although this suggested that adding feedback with either HIF1 α or HIF2 α would increase hypoxia-induced gene expression, our initial attempts at this were unsuccessful. We are currently investigating whether this was due to an artifact in experimental design, due to the experimental finding underlying this strategy not applying to the B16F10 cell line, due to the need for co-expression of additional HIF1 β , or due to the toxicity of increasing the expression levels of the HIF proteins is currently under investigation.

Further, this study presented an opportunity to use our newly developed COMET toolkit for engineering genetic circuits in mammalian cells to modulate the signal from the HBS. We proposed and constructed circuits for amplifying the HBS response as well as for implementing positive feedback without relying on supplemental endogenous HIF proteins. As concerns about noise amplification or leaky gene expression leading to constitutive activation are common with such systems, we implemented an additional level of control by designing and testing COMET TFs degrade in the presence of oxygen to mitigate these concerns. These new TFs could additionally enable biosensors that no longer rely on HIFs, though they would still rely on the prolyl hydroxylases that mark the HIFs and oxygen sensitive COMET TFs for degradation. We have constructed plasmids to test these HBS-modulating and HIF-independent circuits but have not yet had the ability to test them, due to the COVID-19 pandemic. In conjunction with this first round of circuit building, we have begun efforts into modeling the intracellular hypoxia response and the

behavior of the HBS, by using data from this work. These will be combined with our models for COMET TFs in order to describe and predict how genetic circuits can modify the HBS signaling *in silico*. This modeling effort has several goals, including describing how our system performs, evaluating proposed circuit designs, yielding insights into which design choices will be crucial for such circuits, and predicting circuits that would be robust to factors that vary among cell types, yet still allow for a sensitive and specific response. I discuss these efforts further in **Chapter 6**.

4.7 Acknowledgements

This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH) under award number 1R01EB026510, the National Cancer Institute (NCI) of the NIH under award number F30CA203325, the National Institute of General Medical Sciences (NIGMS) of the NIH under award number T32GM008152 (to Hossein Ardehali), the Northwestern University Flow Cytometry Core Facility supported by Cancer Center Support Grant (NCI 5P30CA060553), and the NUSeq Core of the Northwestern Center for Genetic Medicine. This material is based upon work supported by the National Science Foundation under Grant No. 1745753. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

**Chapter 5. Tumor microenvironment induced natural killer (MINK) cells for cancer
immunotherapy**

5.1 Preface

I performed the majority of the work in this study, made all of the figures, and wrote this chapter, except as noted. Several times, Simrita Deol and Hailey Edelstein helped with the transductions, cell culture, and flow prep. The grant from AbbVie funding this study was a collaborative writing effort between myself, Joshua Leonard, Hailey Edelstein, and Bin Zhang. Some of the text in the introduction was sourced from this grant and edited for this context. The studies described here will be published as part of a larger study on the MINK concept that includes the *in vivo* evaluation of the concept (**Chapter 6**). This future work will be carried out by Simrita Deol in collaboration with members of the Bin Zhang lab (NU).

5.2 Abstract

Natural killer (NK) cells are an exciting frontier in immunotherapy. As chimeric antigen receptor (CAR) T cells have demonstrated, engineered cell-based therapies can be generally safe and have high clinical efficacy; however, T cells have had difficulty performing anti-tumor effector functions in immunosuppressive tumor microenvironments. In clinical trials, NK cells home to tumors and are safe, but generally require further engineering for efficacy. To this end, we engineered a novel functionality into NK cells—detecting signatures of the tumor microenvironment and carrying out a therapeutic effector function, such as producing cytokines to potentiate the anti-tumor immune response. We term these cells microenvironment induced natural killer cells (MINK). In this study, we first adapted methods for working with these cells to our laboratory, including establishing a lentiviral transduction protocol and developing a flow-cytometry based assay for evaluating NK cell effector function. We investigated the use of several fluorescent proteins, luciferases, and bioluminescence resonance energy transfer (BRET) systems for their utility in studying hypoxia *in vivo*. We then compared several strategies for conferring IL-2 independence on the NK-92s. Notably, we then demonstrated that NK cells can be engineered to sense hypoxia and produce a gene of interest, meanwhile evaluating several of the design choices for the hypoxia biosensor. We also report on strategies for engineering stable, biosensing NK cell lines and assaying their function.

5.3 Introduction

Engineered cell-based therapies are rapidly emerging as clinically successful technologies for treating cancer. The key driver of this new therapeutic class is autologous T-cells engineered to express chimeric antigen receptors (CARs) that confer recognition of a specific tumor antigen that is expressed on the surface of cells³⁹. While this approach has demonstrated great promise for treating hematologic malignancies, namely B-cell leukemias and lymphomas, cancers, the application of CAR T cells to treat solid cancers has been substantially more challenging. Barriers include an immunosuppressive tumor microenvironment (TME) and lack of targetable surface antigens³⁸. Though systemic delivery of immune-potentiating factors can alter the TME and induce tumor regression, it risks life-threatening toxicities¹⁴⁷, and engineering cell therapies to constitutively produce immune-potentiating factors, as has been done to armor CAR T cells against the immunosuppressive TME, may carry similar risks. Direct intratumoral injection of immunostimulatory factors including cytokines can promote immune control of cancers in some cancers¹⁵⁰, but generally this is not a feasible strategy, as, especially for metastatic cancer, the direct injection of all tumors is difficult, if not impossible, for many patients.

When infused, NK cells naturally infiltrate the spleen, liver, lymph nodes, bone marrow, and peripheral blood^{159, 281}. *Ex vivo* activated NK cells confer some innate antitumor immune control, but cancer cells evolve to evade this attack. Clinical evaluations of NK cell adoptive transfer show no obvious adverse side effects, however these infusions also show limited clinical benefit^{161, 162, 163}. NK cells may be a better platform for engineering cell therapies that provide transient immune potentiation than are T-cells, since NK cells circulate for only days in the human body¹⁶³, mitigating concerns about chronic immune stimulation and immune rejection of the engineered cells. Additionally, allogeneic NK cells, do not induce graft-versus-host disease, as T-cells do¹⁶⁴. Thus, NK cell therapies may be produced as off-the-shelf products derived from a few donors, or even stable cells lines (such as the NK-92 cell line²⁸²), providing substantial manufacturing and administration advantages over autologous products. For instance, the cost of an engineered NK-92 cell therapy may currently be less than \$20,000 compared to \$250,000 or more for a CAR T cell therapy¹⁶⁵.

Several strategies exist to confer NK cells with additional functions to improve their antitumor

efficacy. Target-activated NK cells (taNKs) are engineered to express CAR receptors; these cells exhibit tumor localization and potent anti-tumor activity^{157, 158, 283}. Although as with CAR T cells, this approach requires identification of tumor antigens and is subject to antigenic escape by tumor cells. High affinity NK cells (haNKs) are engineered to express a high-affinity version of the CD16 receptor, allowing them to recognize the Fc chain of antibodies, such as those targeted against tumor antigens, and engage in antibody-dependent cell-mediated cytotoxicity (ADCC)²⁸⁴. Though the NK-92 cell line is ADCC deficient, haNKs engage in ADCC to augment the therapeutic action of these antibodies.

We propose a new strategy to augment the anti-tumor of NK cells in the tumor—microenvironment induced natural killer cells (MINK). These cells will recognize features of the tumor environment, which, as mentioned in **Chapter 1**, may be found in a broader range of cancers, as opposed to tumor antigens which are cancer specific and not expressed in all patients. In response to these features, MINK will produce immune-potentiating cytokines. Ultimately, MINK cells could comprise a standalone product or complement other therapeutic modalities (including CAR T cells, antibodies, drugs, or radiotherapy). In this work, we first evaluate strategies for engineering and assaying the NK-92 cell line, including investigations of strategies for IL-2 independence and *in vivo* reporters that allow visualization of the cell's state through multiple layers of biologic tissue. We then evaluate whether NK-92 cells can be engineered to respond to hypoxia and strategies for doing so and evaluating the response.

5.4 Materials and methods

Plasmid construction

The Firefly Luciferase gene and the IL-2-IL-2R β gene was codon optimized and synthesized by ThermoFisher as Gene Strings. IL-2 and IL-2-KDEL were PCR amplified from this and inserted into appropriate vectors with restriction enzyme cloning. The LumiScarlet (#126623) and Antares2 (#100027) genes were obtained from Addgene. The AkaLuc gene was codon optimized and synthesized by ThermoFisher. The mNeonGreen, miRFP720, PuroR, and BlastR genes were obtained from plasmids described in **Chapter 4**. All plasmids were cloned into a version of pcDNA3.1(+), modified as previously described¹ or the lentiviral vector pGIPZ.

NK-92 culture

NK-92 cells were obtained from ATCC (CRL-2407). Cells were cultured in MEM α (Gibco 12000-022) with supplementation: 0.2 mM Myo-inositol (Sigma I-7508, dissolved in water), 0.1 mM 2-mercaptoethanol (Gibco 21985-023), 0.02 mM Folic acid (Sigma F-8758, stock at 4mM dissolved in 1M NaOH), 1.5 g/L sodium bicarbonate, 12.5% non-heat-inactivated bovine serum (Gibco 16000044), 12.5% horse serum (Gibco 16050122), 100 IU/mL human recombinant IL-2 (Peprotech #200-02, dissolved in acetic acid/0.2% BSA), 1% Penicillin-Streptomycin (Gibco 15140122), L-glutamine (Gibco 25030081). Cells were seeded in a low adherence flask (T-75: Corning 431464U) between 1×10^5 and 4×10^5 cells/mL and subcultured every 2-4 d, with a maximal density of 1×10^6 cells/mL.

NK-92 transduction

Several methods for NK-92 transduction were evaluated. In strategy 1, 1×10^5 cells in 100 μ L of media were mixed with 2 mL of viral supernatant in a 12-well plate, polybrene was added to final concentration of 8 μ g /mL, and the cells were incubated at 37°C for 5 h, after which time the media was changed. In strategy 2, 10^5 cells in 100 μ L of media were mixed with 2 mL of viral supernatant in a 12-well plate, polybrene was added to a final concentration of 8 μ g/mL and the cells centrifuged at room temperature at 360 g for 90 min, then incubated at 37°C overnight, after which time the media was changed. In strategy 3, 1.8 mL of viral supernatant was mixed with 18 μ L of Lipofectamine Plus Reagent, incubated at room temperature for 15 m, and then 7.6 μ L of Lipofectamine LTX was added, followed by a second incubation at room temperature (approximately 20°C to 24°C) for 15 min. This mixture was added to 500 μ L of media containing 5×10^5 cells in a 12-well plate, and this was centrifuged at room temperature at 500 g for 30 minutes, incubated at 37°C for 3 h, and spun again at 500 g for 30 min at room temperature, after which point 6 mL of media were added and the cells allow to incubate until analysis. In strategy 4, 10^6 cells were mixed with 4 mL of viral supernatant and polybrene at a final concentration of 4 μ g/mL, centrifuged at 30C at 2500 g for 90 minutes, incubated at 37°C overnight, at which point the media was changed. For strategy 4, an increasing number of cells was tested, at a concentration of 10^6 cells/mL with the following volumes of cells and viral supernatant 1/4, 2.5/10, 5/20, and 10/20, with polybrene added to a final

concentration of 4 µg/mL. Cells were centrifuged at 30°C at 2500g for 90 min in a 50 mL conical tube in a fixed-angle rotor and then incubated at 37°C overnight, at which point the medium was changed. For all future transductions, strategy 4 with a starting cell number of approximately 5×10^6 NK-92s was used.

K-562 culture

K-562 were obtained from ATCC (CCL-243). Cells were cultured in Iscove's Modified Dulbecco's Medium (Gibco 12200036), supplemented with sodium bicarbonate (3 g/L Fisher Scientific #S233), 10% heat-inactivated fetal bovine serum (Gibco #16140-071), and Penicillin-Streptomycin (100 U/µL and 100 µg/mL, Gibco #15140122). Cells were maintained between 1×10^5 and 1×10^6 cells/mL by subculturing at a 1:10 or 1:20 ratio every 2 or every 3 d, respectively.

K-562 transduction

10^6 K-562s in 7 mL of IMDM were mixed with 1 mL of lentivirus containing supernatant or with 100 µL of lentivirus containing supernatant and 0.9 mL of IMDM, without polybrene, in a 50 mL conical tube. Cells were spun at 900 x g for 2 h at room temperature in a fixed angle rotor. This solution was added to 12 mL of IMDM and cells were cultured for 3 d at 37°C. At this point, lines were analyzed by flow cytometry and the line with the highest transduction efficiency that was < 30% (to ensure an MOI < 1) was selected with puromycin (1 µg/mL) or blasticidin (10 µg/mL) for approximately 1 week.

HEK293FT cell culture

The HEK293FT cell line was purchased from Thermo Fisher/Life Technologies (RRID: CVCL_6911 [https://web.expasy.org/cellosaurus/CVCL_6911]). Cells were cultured in DMEM (Gibco #31600-091) with 4.5 g/L glucose (1 g/L, Gibco #31600-091; 3.5 g/L additional, Sigma #G7021), 3.7 g/L sodium bicarbonate (Fisher Scientific #S233), 10% FBS (Gibco #16140-071), 6 mM L-glutamine (2 mM, Gibco #31600-091; 4 mM additional, Gibco #25030-081), penicillin (100 U/µL), and streptomycin (100 µg/mL) (Gibco #15140122), in a 37°C incubator with 5% CO₂. Cells were subcultured at a 1:5 to 1:10 ratio every 2–3 d using Trypsin-EDTA (Gibco #25300-054). The HEK293FT cell line tested negative for mycoplasma with the

MycoAlert Mycoplasma Detection Kit (Lonza #LT07-318).

Assays with transiently transfected HEK293FTs

From exponentially growing HEK293LP cells, 0.5×10^5 cells were plated per well (0.5 mL medium) in 24-well format, and cells were cultured for 24 h to allow cells to attach and spread. When cells reached 50–75% confluence, plasmids were transfected by lipofection with Lipofectamine LTX with PLUS Reagent (ThermoFisher 15338100). Plasmid was mixed with 0.5 μ L of PLUS reagent in a 25 μ L total volume reaction, with the remainder of the volume being OptiMEM (ThermoFisher/Gibco 31985062). In a separate tube, 1.9 μ L of LTX reagent was mixed with 23.1 μ L of OptiMEM. The DNA/PLUS Reagent mix was added to the LTX mix. pipetted up and down four times, and then incubated at room temperature for 5 min. 50 μ L of this transfection mix was added drop-wise to each well of cells, which was mixed by gentle swirling. Cells were incubated in normoxia overnight and then cultured for 1-2 d in normoxia or hypoxia prior to microscopy.

Cells were harvested for flow cytometry using FACS buffer (PBS pH 7.4, 2–5 mM EDTA, 0.1% BSA) or using 0.05% Trypsin-EDTA (with or without Phenol Red) for 5 min followed by quenching with medium (with or without Phenol Red). The resulting cell solution was added to at least 2 volumes of FACS buffer. Cells were spun at $150 \times g$ for 5 min, supernatant was decanted, and fresh FACS buffer was added.

Lentiviral production

5×10^6 exponentially growing HEK293FTs were plated on a 10 cm dish in 10 mL of media. 6-12 h later the cells were transfected using the calcium phosphate methodology. In a total volume of 850 μ L, 10 μ g transfer vector, 8 μ g psPAX.2, 3 μ g pMD2G, and 1 μ g vector including a constitutive fluorescent protein as a transfection control were mixed. Then 150 μ L of 2M CaCl₂ was added and the solution was mixed. This mixture was added dropwise to an equivolume amount of HEPES Buffered Saline (HEBS), pipetted up and down gently four times, and allowed to rest for four minutes before being vigorously pipetted eight times and applied dropwise to the dish of cells. The next morning, the media was aspirated and replaced with fresh DMEM. 28 to 36 h after this media change, the lentivirus containing supernatant was harvested by aspiration, centrifugation at $500 \times g$ for 2 min to remove cells, and filtration through a 0.45 μ M filter to

remove debris.

Cell sorting

NK-92 cells were suspended at 10^6 - 10^7 cells per mL in sorting medium (MEM α supplemented as described in *NK-92 culture*, with the addition of HEPES (25 mM, Sigma H3375) and gentamycin (100ug/mL, Amresco 0304)), and held on ice until sorting was performed. Cells were sorted using one of several BD FACS Aria Special Order Research Products (Robert H. Lurie Cancer Center Flow Cytometry Core) with the optical configuration listed in **Supplementary Table A5.1**. Cells were collected for each line in sorting medium and held on ice until they could be centrifuged at 150 x g for 5 min and resuspended in DMEM. Cells were plated and expanded until used in experiments. Gentamycin was included in the culture medium for one week after sorting.

Flow cytometry

Flow cytometry was run on a BD LSR Fortessa Special Order Research Product (Robert H. Lurie Cancer Center Flow Cytometry Core). Lasers and filter sets used for data acquisition are listed in **Supplementary Table A5.2** (for experiments involving reporter expression). Samples were analyzed using FlowJo v10 software (FlowJo, LLC). Fluorescence data were compensated for spectral bleed-through. The NK-92, K-562, and HEK293-FT cell populations were identified by SSC-A vs. FSC-A gating, and singlets were identified by FSC-A vs. FSC-H gating. In some experiments, to distinguish transfected from non-transfected cells, a control sample of cells was generated by transfecting cells with a mass of pcDNA (empty vector) equivalent to the mass of DNA used in other samples in the experiment. For the single-cell subpopulation of the pcDNA-only sample, a gate was made to identify cells that were positive for the constitutive fluorescent protein used as a transfection control in other samples, such that the gate included no more than 1% of the non-fluorescent cells. To distinguish transduced cells a sample of the corresponding parental line was used. In some experiments, propidium iodide (PI), 7-AAD, or DAPI was included to assist in determination of cell viability and/or to exclude non-viable cells from analysis.

Killing assay

In each well of a 96-well plate, NK-92 cells were seeded, starting with an initial density of 5×10^5 cells/well in 100 μL of supplemented MEM α and serially diluted 2-fold down to the desired lower limit of cells to be assayed. To each well, 10^4 K-562s were added. In some assays, these K-562 were previously stained with Cell Trace Violet (CTV) or Cell Trace Far Red (CTFR). In some assays, the NK-92s expressed an EYFP or mNeonGreen transgene. For certain conditions, IL-2 was included in the medium, as indicated for each experiment. Plates were incubated at 37°C for 4 h. Then, 50 μL of a solution (0.25 μL of 1 mg/mL PI, 40 μL of fluorescent counting beads (Sigma), and 9.75 μL PBS) was added to each well, each well was mixed, and the plate was incubated in the dark at room temperature for 20 min. The contents of each well were triturated and transferred to a 96-well U bottom plate. The samples were spun for 5 min at 400 x g and resuspended in FACS buffer after decanting. Plates were run on the high throughput sampler for the flow cytometer, and 10 μL of volume was acquired for each well. Samples were analyzed to determine the number of viable NK-92s and K-562s in each well based on their stain and PI exclusion. These numbers were normalized to the number of beads acquired in each well to determine the absolute number of cells per well.

IL-2 independent assay

10^5 total NK-92s were seeded per mL in 5 mL in a T-25 flask, in supplemented MEM α without antibiotics or IL-2. Approximately 5% of these were from the transduced and selected IL-2 independent NK-92 lines and 95% were parental. This starting ratio was measured at the outset for each sample. After 4 d of culture, cells were analyzed by flow cytometry and passaged in fresh media at a maximum density of 10^5 total NK-92s were seeded per mL in 5 mL. After another 5 d, cells were analyzed by flow cytometry.

Luciferase substrates

The following were used for Firefly Luciferase Imaging: Dual-Glow Buffer and Reagent (Promega), D-Luciferin Potassium Salt (Sigma 50227) suspended in PBS at 10 mM, and AkaLumine-HCl (aka TokeOni, Sigma 808350) suspended in PBS at 10 mM. The Antares2 substrate (Diphenylterazine (DTZ) #HY-

111382) and the LumiScarlet substrate (8pyDTZ #HY135368) were obtained from MedChem Express and suspended in ethanol that had been pH adjusted to pH 2 with HCl at concentrations of 1 mM. Dilutions of all substrates were made in Dulbecco's Phosphate-Buffered Saline (DPBS) without calcium or magnesium (Gibco #14190144) for assays. Powdered substrates were stored, protected from light, at -20°C except for DTZ which was stored at 4°C. All resuspended substrates were stored at -80°C.

IVIS imaging and plate reader assays

Opaque plates with white wells were used for imaging (Costar). Unless otherwise noted, for initial experiments, 10^4 K-562s or 10^5 NK-92s were seeded per well. In some cases, less K-562s were seeded per well to achieve a total flux equivalent to that of the NK-92s. For analysis of samples after hypoxic and normoxic culture, 100 μ L of the culture was used rather than a fixed number of cells.

Imaging of plates was performed on an IVIS Spectrum (Perkin Elmer) at the Northwestern University Center for Advanced Molecular Imaging. 100 μ L of cells and 100 μ L of substrate were mixed immediately prior to imaging. In some cases, plates were incubated in the dark in the IVIS Spectrum chamber to dark-adapt the plates and decrease non-specific background, particularly for the imaging of Firefly Luciferase, where this adaptation continued for several minutes prior to imaging. For some experiments, ham (Mariano's) was placed on top of the plate. Acquisition settings were adjusted for each experiment. Images were processed using the Living Image Software (Perkin Elmer). The scale was adjusted to remove the background signal from non-luciferase expressing NK-92s.

Analysis of plates was performed on a Synergy H1 plate reader (Biotek) with 100 μ L of cells and 100 μ L of substrate, which were mixed immediately prior to analysis, except for Firefly Luciferase analysis, where D-Luciferin was incubated with the cells for 5 min prior to analysis. Gain was adjusted for each experiment.

5.5 Results

Comparing and optimizing methods for NK-92 transduction

Gene introduction into NK-92s is typically done through electroporation or lentiviral transduction.

The literature reports several strategies for transducing these cells, typically involving high titers of lentivirus and spinoculation, a process in which the cells are centrifuged in media containing the virus. We evaluated four methodologies based on the literature (**Methods**). Of these, strategy 1 and strategy 4 produced approximately 20% transduction efficiency, while the others resulted in around 7% efficiency (**Fig. 5.1a**). Both involved centrifugation, strategy 1 at low speeds and strategy 4 at high speeds. We proceeded with strategy 4, as this method started with 5 times as many cells, a critical advantage, as NK-92 cells have a long doubling time of approximately 48 h.

Next, we investigated scaling this method up even further. The use of 2.5 or 5 times more cells did not substantially decrease transduction efficiency but using 10 times more cells dropped the efficiency to 10% (**Fig 5.1b**). Therefore, all transductions after this point were carried out using Strategy 4 with a starting cell number of approximately 5×10^6 cells. This typically produced a dense culture of cells in a full T-75 flask three days after transduction, making the cells ready to undergo antibiotic or FACS selection.

Adaptation of an assay to measure effector function

The primary effector function of NK cells is to kill cells, particularly those that are virally infected or demonstrating signs of other pathology. The K-562 human lymphoma line is the typical target line for assaying this function. Reported assays use release of a radioactive isotope, plate reader fluorescence, or microcytometry. In order to develop a high-throughput assay with readily available equipment in our lab, we adapted these protocols to be amenable to flow cytometry. In such an assay, a set amount of K-562 tumor cell are seeded in a 96-well plate and cultured with varying numbers of NK-92s for 4 h. Each cell population can be identified by flow cytometry as is either expresses a fluorescent protein or has been dyed or stained prior to the assay. After the co-culture period, a set number of fluorescent microbeads (to enable absolute cell counts per well from a sample of the total volume) and propidium iodide (to discriminate live versus dead cells) were added. Samples were analyzed by identifying the number of live K-562s remaining in each well (**Fig. 5.1c**) and plotting this against the number of live NK-92s per well (**Fig. 5.1d**). This sample protocol can be used for microcytometry, where the propidium iodide is added prior to the assay (**Fig 5.1e**).

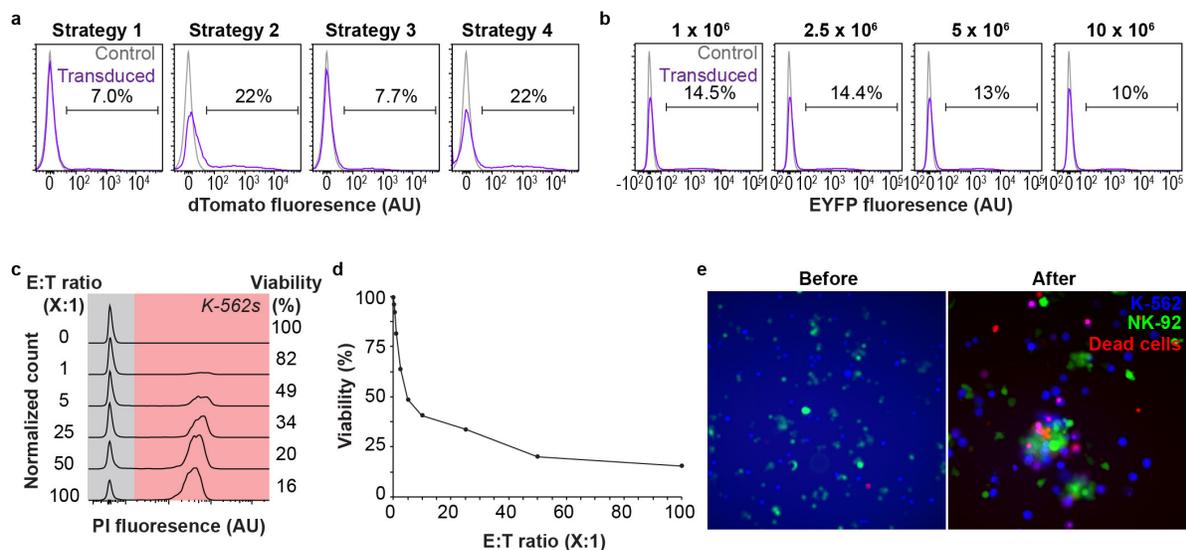


Fig 5.1 Development and validation of methodologies for NK-92 cell laboratory work. a) Comparing the efficiency of four strategies (**Methods**) for lentiviral transduction of NK-92 cells. b) Scaling up the starting number of cells in Strategy 4 and the effects on transduction efficiency. c-e) an assay to evaluate the NK-92 effector function. c) the survival of K-562s can be determined by the exclusion of propidium iodide (PI) in viable cells. d) plotting K-562 survival against the NK-92 effector to K-562 target (E:T) ratio. e) evaluation of killing activity by microscopy, as the number of dead K-562s (purple) increases during the four hour incubation.

Comparison of reporter systems

The *in vivo* imaging of fluorescent proteins requires both excitation and emission light to penetrate through several layers of tissue in the animal. As most light in the visible spectrum is absorbed by various molecules in the tissue, particularly myoglobin and hemoglobin, this presents a major obstacle to using fluorescent proteins as *in vivo* reporters. One solution to this is to use infrared fluorescent proteins (iRFPs), as infrared light does penetrate through tissue better than visible light. We first evaluated two iRFPs with different spectra, miRFP670 and miRFP720. K-562s were transduced with a lentiviral construct encoding the FP driven by the EF1 α promoter, an internal ribosome entry site (IRES) and the PuroR gene, which confers resistance to puromycin, at a multiplicity of infection (MOI) < 1 and selected with puromycin. These constructs were then analyzed via flow cytometry (**Fig 5.2a**). We found that the miRFP720 had several advantageous properties compared to the miRFP670: it had a very high signal-to-noise ratio in the Alexa750 channel (excited by a far-red laser), it maintained good levels of fluorescence in the Alexa700 channel (excited by a red laser) and therefore was compatible with our FACS instrument, and it only bled minimally into the APC channel. miRFP720 was thus selected to be our constitutively expressed fluorescent protein for cell engineering.

To investigate options for a second output, we turned to a strategy known as bioluminescence resonance energy transfer (BRET), wherein a tethered luciferase provides the excitation energy for a fluorescent protein (**Fig. 5.2b**). Luciferase substrates diffuse throughout mice, enabling deep-tissue excitation and subsequent emission of light, and the emitted red light penetrates through tissue well enough. While previous generations of BRET proteins relied on the very bright NanoLuc, this protein emits blue light, which only poorly excites proteins capable of emitting red-shifted light. Several generations of engineering efforts have produced two leading BRET proteins: Antares2²⁸⁵ (a teLuc flanked by two cyOFP1 fluorescent proteins; its substrate is diphenylterizine (DTZ)) and LumiScarlet²⁸⁶ (a LumiLuc fused to a mScarlet-I; its substrate is 8pyDTZ). We compared these two constructs, for which the substrates have only recently become commercially available. Ideally, the better construct would produce red-shifted light, penetrate through tissue, be brighter, and rapidly oxygenize after exposure to oxygen.

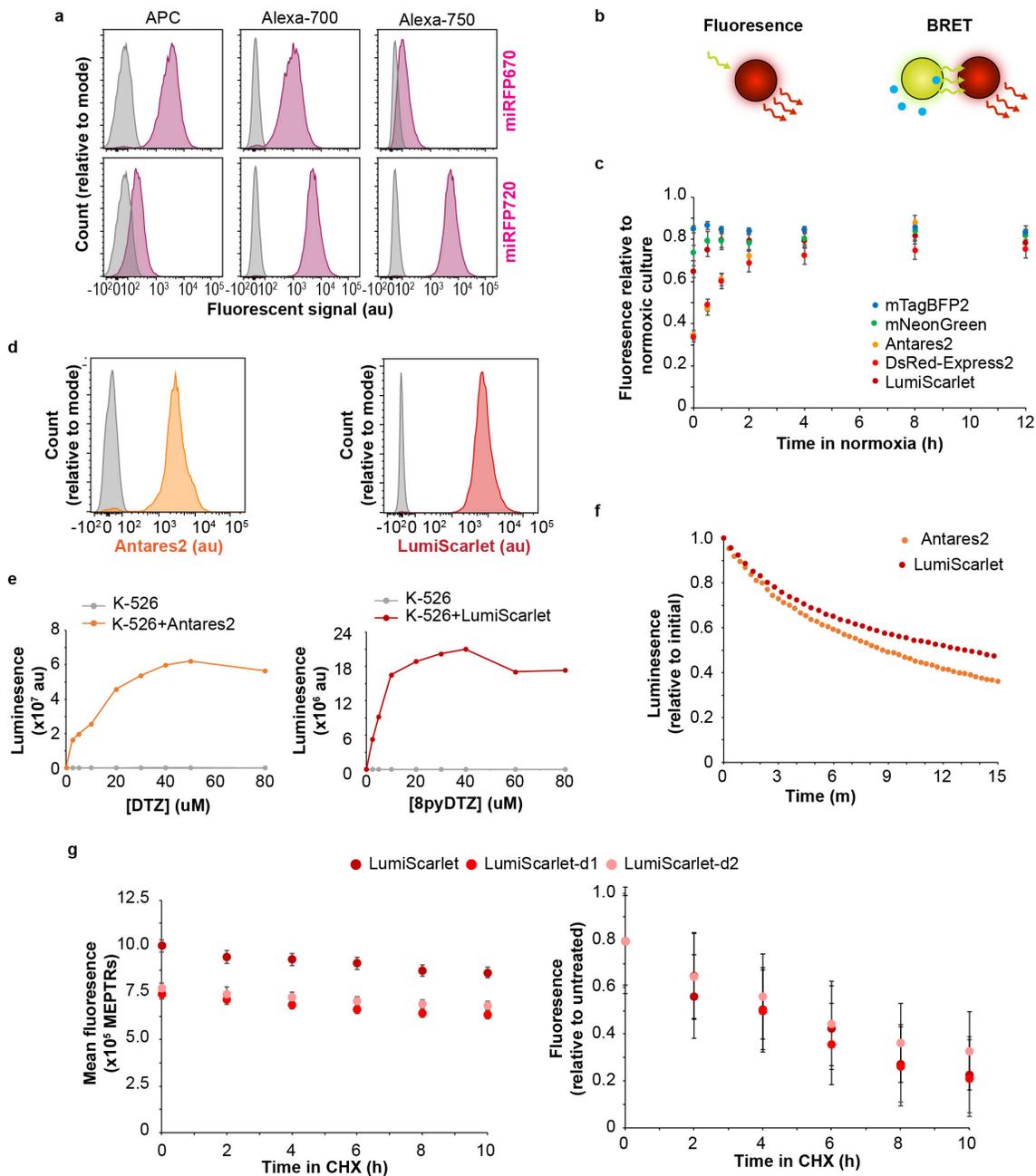


Fig 5.2 Evaluating infrared and BRET reporter systems. a) Evaluation of miRFP670 and miRFP720 with several laser and filter combinations. b) Schematic depicting BRET. c) Oxidation rate of commonly used fluorophores d) Comparison of Antares2 and LumiScarlet by flow cytometry in a stable K-562 line. e) Dose-response relationships for BRET proteins and cognate substrates. f) Kinetics of BRET enzymatic reaction after substrate addition. g) Time course of LumiScarlet degradation after treatment with the translation-inhibiting drug cyclohexamide (CH).

Having previously shown that fluorescent proteins mature at different rates following exposure to oxygen, we first tested how long the fluorophores in Antares2 (CyOFP1) and LumiScarlet (mScarlet-I) would take to mature. These two fluorophores were compared to other commonly used reporters in our lab, mNeonGreen, mTagBFP2, and DsRed-Express2. Cells were transfected with a construct encoding each protein, allowed to transfect overnight under normoxic condition, and then cultured under hypoxic conditions for another two days. At this point, samples were moved back into a normoxic incubator at different times to allow for oxidation before being analyzed by flow cytometry (**Fig. 5.2c**). As expected, mTagBFP2 and mNeonGreen oxidized the fastest, with DsRed-Express2 oxidizing more slowly than either. LumiScarlet oxidized at a similar rate to mNeonGreen, while Antares2 oxidized more slowly like DsRed-Express2.

To study the brightness and kinetics of these two reporters in the genomic context, K-562 cells were transduced with lentiviral constructs encoding LumiScarlet or Antares2 upstream of an internal ribosome entry site (IRES) and a mRFP720-P2A-PuroR gene at an MOI < 1 and selected with puromycin. These cell lines were then used to validate the properties of the BRET reporters by plate reader, IVIS imaging, and flow cytometry. By flow cytometry, Antares2 had a signal-to-noise ratio (SNR) of 137 while LumiScarlet had a SNR of 1,320, primarily due to cells having lower autofluorescence in the channel he LumiScarlet is detected than in the channel Antares2 is detected in (**Fig. 5.2d**). Various concentrations of substrate-containing buffer were added to these cell lines, which were analyzed in a plate reader. From this, we determined concentrations for each that produced near maximal signal output: for DTZ, this was 30 μ M, and for 8pyDTZ, this was 15 μ M (**Fig. 5.2e**). Kinetics for both reporters were slow flash (**Fig. 5.2f**). As we are interested in whether a cell is currently experiencing hypoxia, rather than just if a cell has ever experienced hypoxia, we investigated methods for shortening the half-life of the LumiScarlet reporter. LumiScarlet showed a very long protein half-life, and this could not be decreased by addition of degradation tags (**Fig 5.2g**).

Cells were then serially diluted and the above-determined concentrations of substrate were added; these were imaged on the IVIS Spectrum, which is used for plate and *in vivo* imaging (**Fig 5.3a**). The LumiScarlet produced more signal than the Antares2, and with an exposure time of 10 s, a well containing as few as 20 LumiScarlet transduced K-562s was distinguishable from a well containing only buffer and

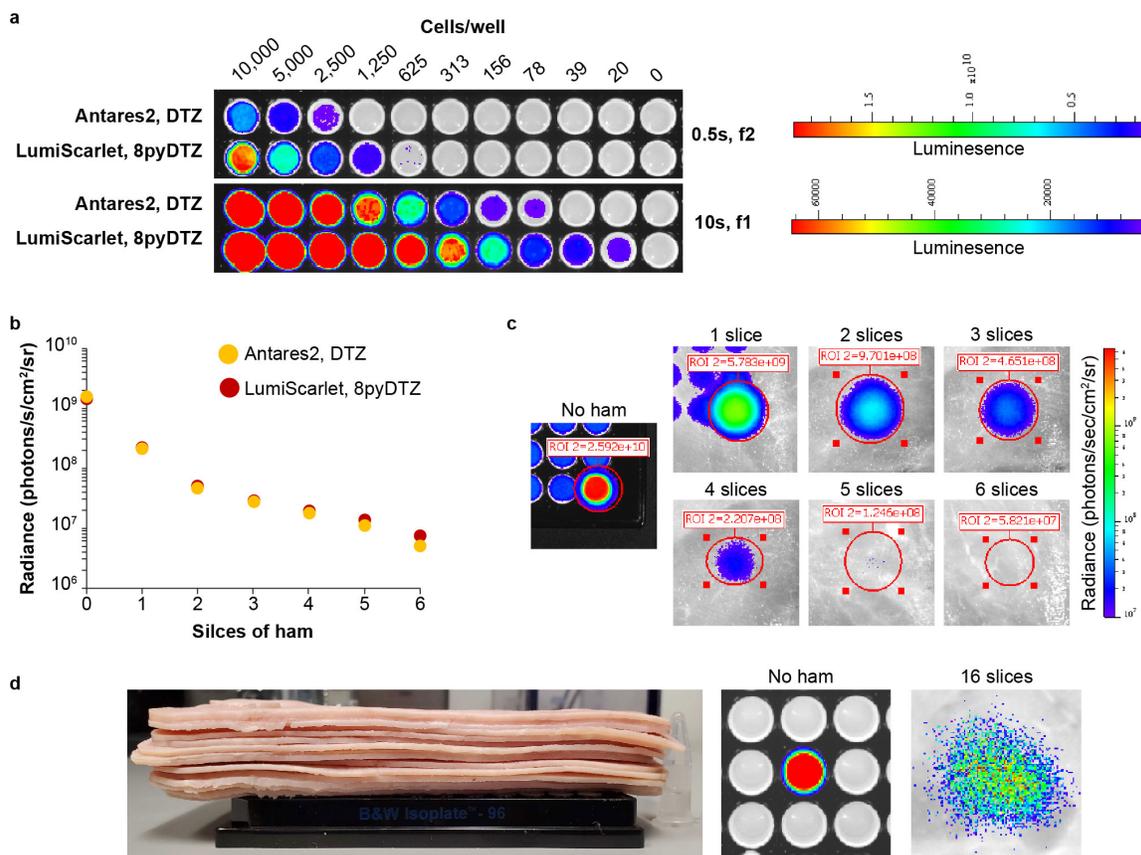


Fig 5.3 Imaging BRET with the IVIS spectrum and under biologic tissue. a) Serial dilution of stable K-562 BRET-expressing cell lines; image acquired on IVIS Spectrum several minutes after substrate addition, under two acquisition different conditions. b) Comparison of the penetrance of light emitted by the two BRET fluorophores under an increasing number of ham slices. c) Images from experiment in **b** for LumiScarlet show progressive loss of signal as slices are increased d) Imaging of LumiScarlet under 16 slices of ham with increased exposure time and camera sensitivity.

substrate (**Fig. 5.3a**). To mimic penetration of light through, we added slices of ham on top of the plates to determine how much light would penetrate (**Fig. 5.3b-d**). In these assays, the ham mimics the tissue of an animal, containing substances responsible for the absorption and scattering light in humans other animals, including water and hemoglobin (as muscle tissue, ham contains more myoglobin than hemoglobin, though the two proteins are similar²⁸⁷)²⁸⁸; roast beef has similarly been used by other groups for this purpose²⁸⁹. For this assay, less Antares2 expressing cells were seeded than LumiScarlet expressing cell, in order to make the starting signal from each reporter comparable. The first slice of ham (a few mm thick) induced an approximately 10-fold reduction in transmitted light, while each additional piece blocked slightly less (**Fig. 5.3b,c**). Both reporters were still detectable under 6 slices of ham, given sufficient exposure time. At a 30 s exposure time, the LumiScarlet was detectable under as many as 16 slices (**Fig. 5.3d**). The results of these investigations, combine with the reported increased water solubility (and thus better *in vivo* dosing) of LumiScarlet, makes it the better BRET system for our studies of hypoxia.

Evaluating methods for NK-92 IL-2 independence

The NK-92 cell line is dependent on exogenous IL-2 for growth and survival. Several methods have been reported in the literature for making this cell line independent of exogenous IL-2, a property that in some implementations confers enhance killing activity and *in vivo* persistence, but it is currently unclear how these strategies compare to one another. Therefore, we compared these three methodologies head to head (**Fig. 5.4a**). In the first strategy, the cell expresses an IL-2 transgene^{290, 291}. In the second, the expressed IL-2 transgene is tagged with a KDEL sequence to localize it to the endoplasmic reticulum, preventing its secretion²⁹². In the final strategy, the IL-2 is tethered to the IL2 receptor (IL-2R) β chain by a linker which comprises part of the IL-2R α chain and a *myc*-tag²⁹³. Each gene was codon optimized and placed under the constitutive EF1 α promoter in a lentiviral transfer vector, followed by an IRES and mNeonGreen-P2A-PuroR gene, which allowed for antibiotic selection and fluorescent identification of the transduced cells. A control construct contained no IL-2 gene. NK-92s were transduced at an MOI < 1 and

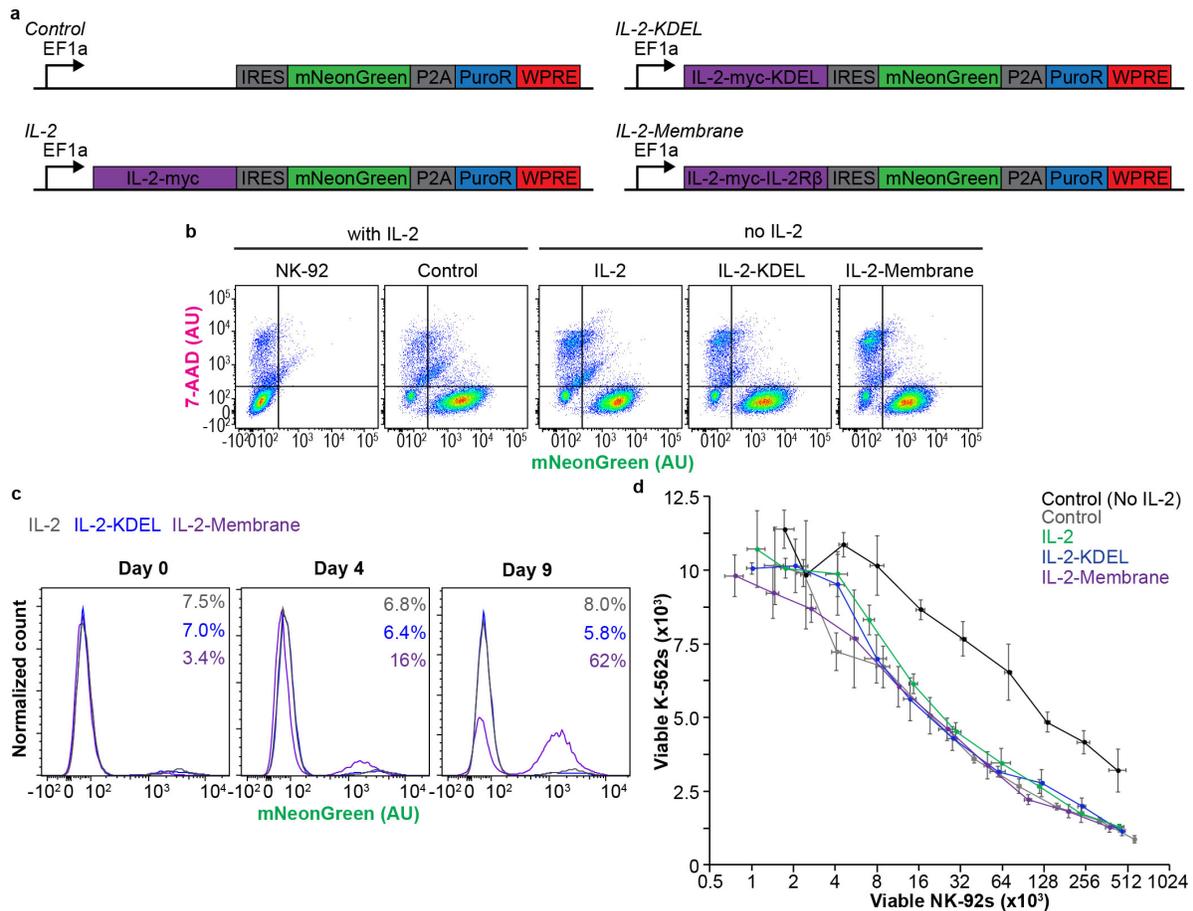


Fig 5.4 Comparing strategies for conferring exogenous IL-2 independence to the NK-92 cell line.

a) Lentiviral constructs for conferring IL-2 independence in NK-92 cells. b) Flow cytometry histograms of transduced and selected NK-92 cell lines following two weeks in culture with and without IL-2 as indicated. All single events are shown. 7-AAD is an infrared viability stain. c) Co-culture assay of IL-2 independent NK-92 lines with non-transduced IL-2 dependent NK-92 cells. d) Killing assay of IL-2 independent and dependent NK-92 cell lines.

selected with puromycin. The cells were then cultured without exogenous IL-2 to determine whether they could indeed grow independently. While the parental NK-92 and NK-92 line transduced with the control construct only survived when cultured with IL-2, all lines transduced with one of the three constructs conferring IL-2 independence survived and proliferated (**Fig. 5.4b**). Notably, this did not hold true for the murine LNK line, in which only the membrane-tethered IL-2 construct conferred exogenous IL-2 independent survival.

One concern with constitutive production of cytokines is that they could leak out of the cell and exert effects on surrounding cells, possibly leading to off target effects and toxicity when used as part of a therapy. To investigate to what extent this phenomenon was occurring with the IL-2 constructs, the purified lines were spiked into cultures of non-transduced NK-92s. These cultures were grown for 10 days in the absence of exogenous IL-2 and evaluated by flow cytometry. In this assay, transduced lines will only expand if they have a growth advantage, which only occurs if they are IL-2 independent and not supplying the cells around them with IL-2. The IL-2 and ER-localized IL-2 transduced lines did not selectively expand, as they supported the growth of the NK-92s in culture with them (**Fig. 5.4c**). The membrane-tethered IL-2 construct selectively expanded, indicating that this line has minimal to no trans-acting properties. While it has been previously reported that the ER localized construct was not secreted by the NK-92s, our line likely did secrete IL-2. It is possible that the codon optimization of this gene led to such high expression levels that the cellular pathways responsible for ER-retention of KDEL-tagged proteins were overwhelmed.

To determine whether the constructs affected the killing ability of the NK-92s, we performed a killing assay, as previously developed in **Fig. 5.1**. When the lines were cultured without IL-2, all IL-2 independent lines performed better than the control line, killing more K-562s per NK-92 (**Fig. 5.4d**). The IL-2 independent lines performed similarly to the control line only when the latter was cultured with IL-2. While the membrane-tethered line might have performed better than the control line at low numbers of NK-92s, this will need to be evaluated in an assay powered to detect these differences. Overall, these properties make the membrane-tethered IL-2 construct an excellent option for conferring exogenous IL-2 independence in the NK-92 cell line.

Design and evaluation of a luminescent, IL-2 independent line

We decided to confer the advantageous property of exogenous IL-2 independence a manner that did not use a fluorescent protein or antibiotic resistance gene for selection, thereby leaving more fluorescence channels open for future assays and more antibiotic selection markers available for future engineering efforts. A lentiviral construct was designed that expressed the membrane-tethered IL-2 under the EF1 α promoter, followed by an IRES and the Firefly Luciferase gene (**Fig. 5.5a**). This would allow visualization of cells transduced with this vector *in vivo*. NK-92 cells were transduced with this construct and cultured for several weeks without IL-2 to select for transduced cells.

After two weeks of selection, the line was proliferating without IL-2, but cell staining with an anti-*myc* antibody revealed no detectable surface expression of the receptor and only some intracellular staining (**Fig. 5.5b**). Western blot analysis was obscured by non-specific staining of the NK-92 lysates, whether using an anti-*myc* or anti-IL-2 antibody. The line, however, produced a significant luminescent signal when the Firefly Luciferase substrate D-luciferin was added to intact cells in media, with the signal from as few as 100 cells detectable on a plate reader (**Fig. 5.5c-d**). This signal was stronger with the kit from Promega than with generic D-luciferin in DPBS (**Fig. 5.5c**). The line was also detectable on the IVIS imager, though with a low signal to noise ratio (**Fig. 5.5e**).

As the light emitted by the catalysis of D-Luciferin substrate by Firefly Luciferase is yellow, it does not penetrate well through tissues. To address this issue, a novel substrate for this enzyme, AkaLumine, that emits more red-shifted light has been developed²⁸⁹. A version of this substrate, AkaLumine-HCl, has also been developed that is additionally highly soluble in aqueous solutions, allowing delivery of much higher doses *in vivo*, a major limitation for D-luciferin, which is poorly soluble in water and thus *in vivo* studies with this substrate are dose limited²⁹⁴. However, signal with the red-shifted substrate AkaLumine was much lower than the signal with D-luciferin, when measured by plate reader (**Fig. 5.5c**). To mitigate this, a new enzyme was evolved to be a better enzyme for this substrate; AkaLuc catalyzes AkaLumine-HCl with high efficiency, enabling the tracking of single cells *in vivo*²⁹⁴. A new NK-92 IL-2 independent line was generated, using a lentiviral vector similar to that in **Fig. 5.5a** with AkaLuc rather than Firefly Luciferase (**Fig. 5.5f**). This line was 5.2 times brighter than the line with Firefly Luciferase (**Fig 5.5g**).

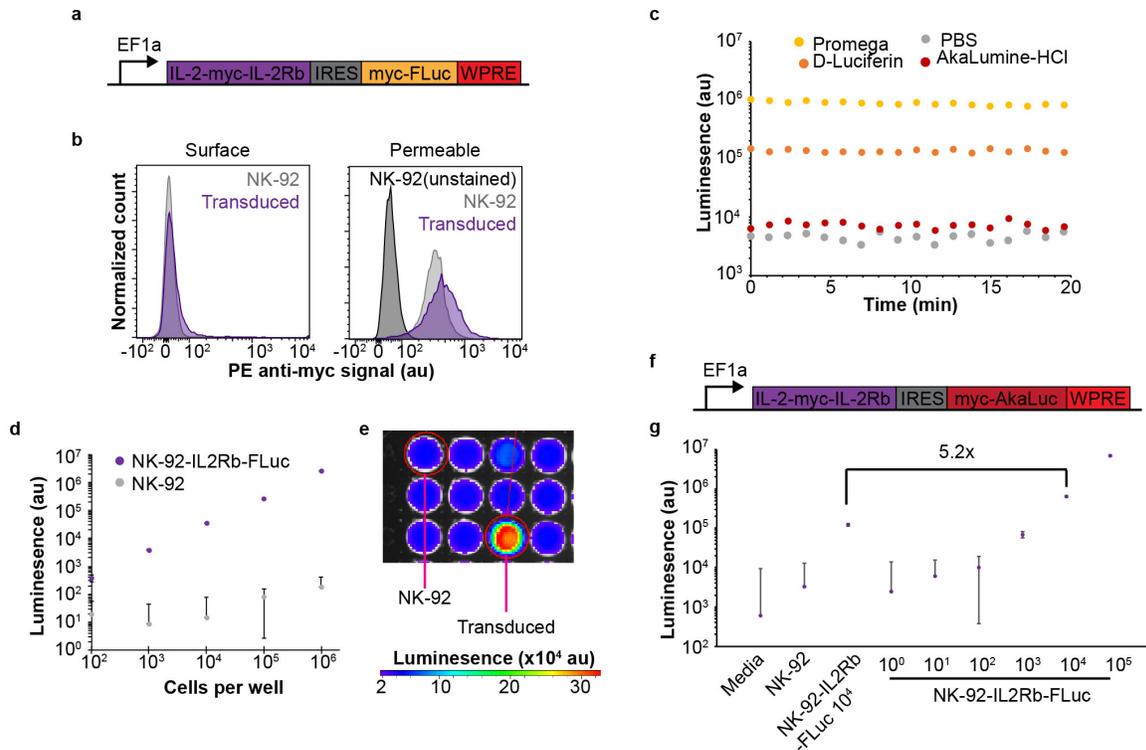


Fig 5.5 Characterization of an IL-2 independent NK-92 line with luciferase expression. a) Lentiviral construct for conferring IL-2 independence and firefly luciferase gene expression. b) Staining of the transduced, IL-2 independent NK-92 line for the myc antigen on the IL-2 and firefly luciferase genes. c) Plate reader analysis of transduced, IL-2 independent NK-92s expressing a firefly luciferase. d) Testing four substrates for the firefly luciferase and measuring signal kinetics. e) Dose-response of NK-92 cells expressing firefly luciferase or not, with Promega substrate. f) IVIS imaging of NK-92s that do or do not express firefly luciferase, using D-Luciferin as substrate. g) Plate reader analysis of transduced, IL-2 independent NK-92s expressing various luciferases with D-Luciferin and AkaLumine substrates.

Investigating an HBS in NK-92s

IL-2-independent, Firefly Luciferase expressing NK-92 cells were transduced with a lentiviral construct encoding a hypoxia biosensor (HBS) construct that we have previously shown to have low background and an excellent fold induction when genomically integrated (**Figs. 4.4,4.5**). The HBS expresses LumiScarlet and is upstream of an EF1 α miRFP720-P2A-PuroR gene, allowing for selection and identification of cells (**Fig. 5.6a**). After selection with antibiotics or sorting by flow cytometry, the line was cultured in hypoxic or normoxic conditions for 3-4 d and assayed for reporter expression by flow cytometry. Hypoxia resulted in a 20-30-fold increase in expression of the LumiScarlet reporter (**Fig. 5.6b**). This was higher for the sorted line than the antibiotic selected line, possibly because the drug selected line was crowded by residual cells and debris which could impair oxygen diffusion to the live cells, while the sorted line was cultured at a lower density. However, the response was heterogenous, with some cells turning on more than others and some not turning on at all; further, output from the HBS appeared correlated with the expression of the constitutive miRFP720, which would be determined by integration locus. Sorting on miRFP720 expression yielded 8 lines, each with distinct levels of miRFP720 expression, indicating that the sorting strategy was successful (**Fig. 5.6c**). In these lines, the magnitude of the OFF and ON states of the HBS increased with miRFP720 expression level and generally proportionately with each other, leading to higher signal but no substantial differences in fold induction (**Fig. 5.6b-c**).

We then compared the NK-92 HBS line with the line generated by sorting for the brightest octile of miRFP720 expressing cells, culturing them for 3-4 d in normoxia or hypoxia, and imaging the cells using the IVIS Spectrum. The signal from the constitutively expressed D-luciferin decreased with hypoxia, while the hypoxia-inducible LumiScarlet increased for both of the lines (**Fig 5.6e**). On this instrument the background was not significantly different between the two lines, but the ON state of the high-miRFP720 line was higher than that of the NK-92 HBS line, leading to an improved fold induction of 7.2 versus 4.3 (**Fig. 5.6f**).

We also examined several other hypoxia biosensor constructs with other minimal promoters (CMV_min and SV40_min) that have demonstrated efficacy in other contexts (**Fig. 5.7a**). The HBS as originally reported used the SV40_min, and the CMV_min generally confers both higher OFF and ON states

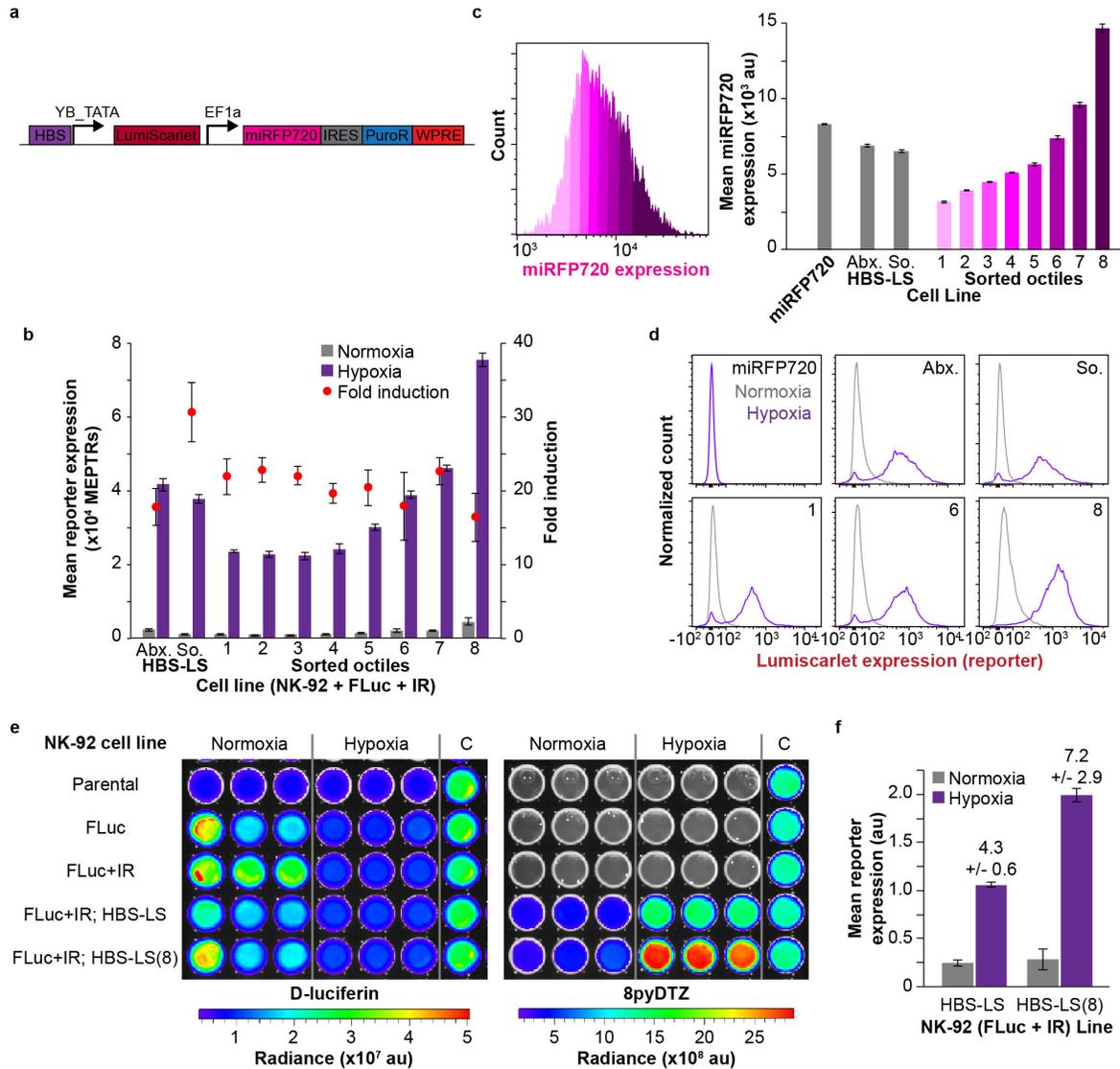


Fig 5.6 Investigating HBS transduced NK-92 cell lines. a) Schematic of HBS lentiviral construct b) Functional assay of NK-92 lines culture in hypoxic or normoxic conditions for 4 days c) Sorting strategy for NK-92 cell lines and evaluation of results by flow cytometry d) Representative flow cytometry histograms from **b** e) Evaluation of samples from experiment in **c** imaged on IVIS Spectrum f) Quantification of LumiScarlet reporter expression in NK-92 cell lines with HBS from **e**.

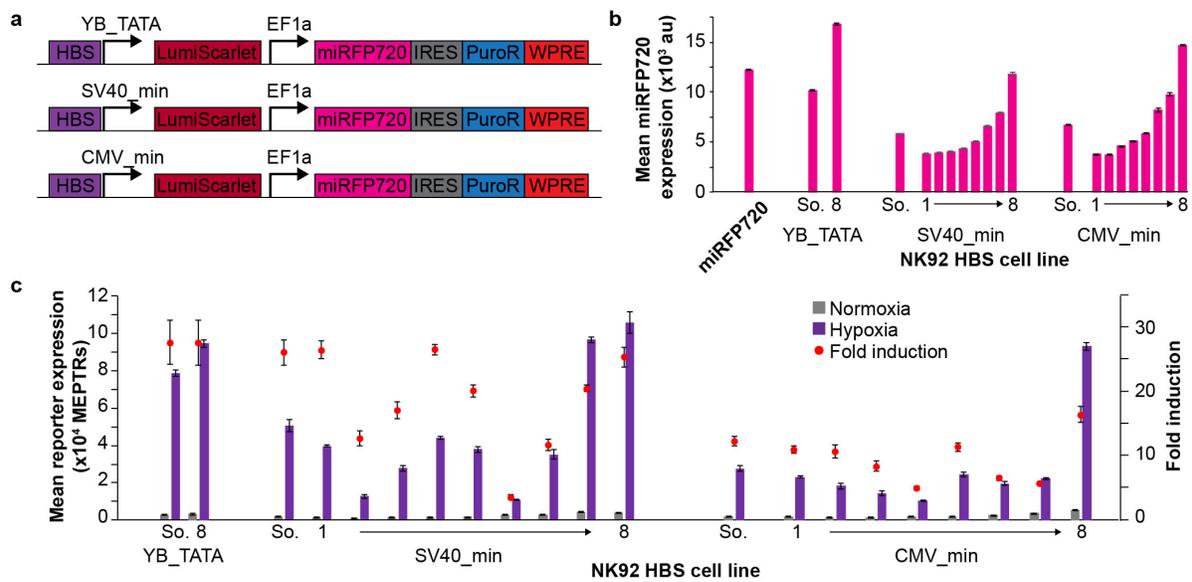


Fig 5.7 Engineering NK-92 lines with HBS with other minimal promoters. a) Lentiviral constructs for NK-92 HBS lines b) Evaluation of sorted cell lines c) Functional assay of NK-92 lines transduced with HBS constructs with varying minimal promoters after hypoxic and normoxic culture.

to biosensors. However, these lines performed no better than the HBS with the YB_TATA minimal promoter, and, surprisingly, the CMV_min-based construct actually showed the lowest ON state (**Fig. 5.7c**). Sorting on miRFP720 expression did show the expected differences in miRFP720 expression for each line (**Fig. 5.7b**); however, the trends observed for the YB_TATA minimal promoter did not hold for these two constructs (**Fig. 5.7c**). The ON and OFF states varied in no predictable fashion, but each line with the highest miRFP720 expression showed ON states higher than the parental line. Thus, while sorting based on constitutive miRFP720 expression is a useful strategy to increase the brightness of a constitutive fluorescent protein, it is not a useful strategy to increase the performance of an HBS delivered via lentivirus.

We then compared the unsorted lines and the highest-octile miRFP720 sorted line for each HBS construct. After 3-4 d of hypoxic culture, the YB_TATA based lines performed better than either of the other lines (**Fig. 5.8a**). Interestingly, in this experiment, the SV40_min HBS showed a marked improvement in performance with sorting, though in this particular repeat of the experiment, the cells for the unsorted SV40_min HBS line did not proliferate well, possibly indicating a confounding factor. Also of note was a decrease in the miRFP720 signal when the cells were cultured in hypoxia (**Fig. 5.8b**). This decrease was also seen for line expressing miRFP720 without an HBS, though the decrease in this line was less than the decreases in lines that did include an HBS. Therefore, the decrease likely comes from both the hypoxic response of the NK-92 inducing some global decrease in protein expression and promoter interference from the HBS, which is located upstream of the EF1 α promoter that drives the miRFP720 (**Fig. 5.6a**). A time course experiment revealed that the transduced NK-92 lines produced LumiScarlet with as little as 12 h of culture in hypoxia and that the reporter expression from the biosensor plateaued after 3-5 d of culture (**Fig. 5.8c**); the decreases in miRFP720 expression also occurred gradually over this time period (**Fig. 5.8d**). In summary, the YB_TATA-based HBS construct showed good performance, and no difference in the timing of the response to hypoxia were noted among HBS constructs with differing minimal promoters.

Samples from the experiment shown in **Fig. 5.8a** were also run on the IVIS imager to determine the utility of this instrument for detecting the various fluorescent and luminescent proteins these cells have. Cell number was not normalized per well; rather, equivolume amounts of each line were added to the wells

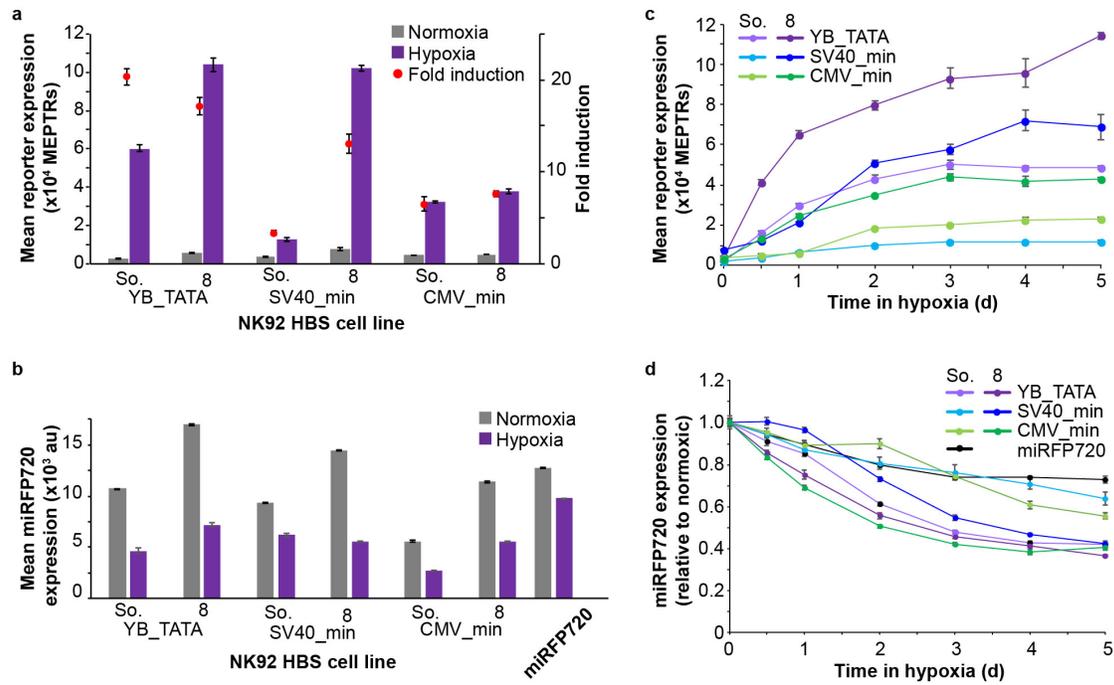


Fig 5.8 Comparison of performance of NK-92 HBS lines with various minimal promoters by flow cytometry. a) Functional assay of cell lines after 4 days of hypoxic and normoxic culture b) miRFP720 levels from the experiment in a c) Time course of hypoxia-induced LumiScarlet reporter expression in stable NK-92 cell lines d) Time course of hypoxia-induced decrease in miRFP720 expression

of a 96-well plate after 3-4 days of hypoxic or normoxic culture. The LumiScarlet was essentially not detectable via fluorescence, likely due to interference from the culture media the cells were suspended in. The miRFP720 was detectable, though some signal was evident from non-miRFP720 transduced lines, likely from dead or dying cells, as these populations of NK-92s are evidently fluorescent in this channel on the flow cytometer (**Supplementary Fig. A5.1a**). Also evident are the clumps of cells evident in the middle of the well, decreases in miRFP720 expression with hypoxic culture, and the decreased cell numbers for the SV40_min parental line. The signal from D-luciferin applied to these cells was also acquired; generally, it was not far above the noise from other light sources and not very sensitive, though in some cases a decrease was evident with hypoxic culture (**Supplementary Fig. A5.1b**). However, the signal from the hypoxia-induced LumiScarlet was bright and detectable in 5 of the 6 lines (**Supplementary Fig. A5.1c**). Notably in these experiments, there was no evidence of cross-reactivity between the two luciferases or substrates. Excitingly, the signal from hypoxia-induced LumiScarlet was detected through multiple layers of ham, hinting at promise in the future *in vivo* studies (**Supplementary Fig. A5.2**).

5.6 Discussion

In this work, we established strategies for working with and evaluating NK cells as a therapeutic platform. Though the NK-92 cell line is typically regarded as difficult to transduce, we achieved efficiencies as high as 22%, yielding up to 1×10^6 engineered cells per transduction—numbers that readily allow for selection of transduced cells by antibiotic resistance or FACS. Our protocol does not rely on use of concentrated lentivirus and has no marked toxicities or other obvious adverse effects on cell health. The killing assay protocol developed here is notable as it avoids the risks of radiation and burdensome requirements to become a radioactivity certified laboratory. Additionally, it is high throughput: most of the pipetting can be done with a multichannel pipette and the high-throughput adapter for a flow cytometer takes the samples directly from 96-well plates, allowing for analysis of hundreds of samples per hour. Analysis is done rapidly with FlowJo and provides single cell resolution. Contrast this with microcytometry, which can also provide single cell resolution but requires software capable of discriminating the cells from each other and a much larger amount of computational time to process all of the images from large assays.

With the use of counting beads, our assay report on the absolute number of cells per well, including the number of NK-92s, providing an insight into the error associated with this independent variable.

The study of BRET reporters here is also particularly useful for our future work. Historically, we have relied on Firefly and Nano Luciferases, but neither are ideal for *in vivo* work as they are not very bright or emit light that is too blue-shifted to penetrate tissue, respectively. Enabled by the very recent commercial availability of novel substrates for recently reported luciferases, we showed that LumiScarlet is a good reporter system for the study of hypoxia *in vivo*—the fluorophore oxidizes rapidly and is bright, cells show little autofluorescence in this channel allowing excellent discrimination by flow cytometry, the bright light allows for detection of as few as twenty cells on the IVIS Spectrum, and the red-shifted light emitted by the system penetrates through multiple layers of biologic tissue and can enable *in vivo* evaluation. As LumiScarlet replaces NanoLuc, we hope that AkaLuc will replace Firefly Luciferase. Future work will need to be done to develop a protocol to enable sequential readout of these *in vivo*. Key items to study include the administration route of the substrate, the distribution and washout time of the substrate, and the *in vivo* kinetics of the reaction.

Though several methods for conferring exogenous IL-2 independence in the NK-92 line have been reported, this is the first study to compare them. We found that the strategies other than the membrane-tethering of IL-2 resulted in secretion of these cytokines, which could potentially cause toxicities if used as a therapeutic. Further, we discovered that the membrane-tethered IL-2 construct can be used as a selectable marker, by culturing the cells in the absence of IL-2. This allows us to avoid using antibiotic resistance markers or fluorescent proteins to select for transduced cells, leaving these proteins available for selecting cells transduced with other viral constructs in the future. We employed this principle to generate

The most exciting finding from this work, however, is that NK-92 cells can be engineered to detect hypoxia and, in response, produce a gene of interest. Though in this work, we only evaluated this gene being a reporter gene, this will ultimately be a therapeutic cytokine, as we are currently evaluating. There was no guarantee that these biosensors would work in this cell line. In fact, some of the trends are surprising. First, the YB_TATA minimal promoter generally confers the lowest background of all of the minimal promoter options, but this low background comes with the trade-off that it will also have the lowest

inducible signal. The background from all of the minimal promoters was comparable, but the YB_TATA led to higher hypoxia-induced gene expression than did the CMV_min and original SV40_min promoters, which have been shown to produce high levels of gene expression than this promoter in other cell types (**Chapter 4**). The second surprising finding is evident in the results from the studies in which we sorted these lines based on a constitutively expressed fluorescent protein that was delivered on the same lentiviral vector as the hypoxia biosensor. For the YB_TATA based HBS line, sorting led to proportionately higher background and hypoxia induced signaling for each increasing octile of constitutively expressed fluorescent protein. However, for the other two lines, this trend did not hold, and these states fluctuated as the octile of the constitutively expressed fluorescent protein increased. Despite this, for all investigated lines, we determined that sorting for the top octile of fluorescent cells leads to an increased hypoxia induced signal and a more homogenous population. This strategy will be useful in preparing lines, so that all cells show a similar behavior and are brighter, making their detection *in vivo* more likely.

5.7 Acknowledgements

This work was supported in part by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award number F30CA203325, the National Institute of General Medical Sciences (NIGMS) of the NIH under award number T32GM008152 to Hossein Ardehali, the Northwestern University Flow Cytometry Core Facility supported by Cancer Center Support Grant (NCI 5P30CA060553), a grant from AbbVie, and the NUSeq Core of the Northwestern Center for Genetic Medicine. Imaging work was performed at the Northwestern University Center for Advanced Molecular Imaging generously supported by NCI CCSG P30 CA060553 awarded to the Robert H Lurie Comprehensive Cancer Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Chapter 6. Conclusions and future directions

As an MD/PhD student, my clinical and research interests lie in the development of new therapeutic strategies for diseases for which there is no treatment or for which the treatment is highly toxic. This latter point becomes particularly salient when combined with my interests in pediatrics: though many children with cancer are ultimately cured, the therapeutics do so at the cost of significant, long-term damage to the developing body and development of life-altering and life-threatening diseases during young adulthood. Targeted, cell-based therapies hold great promise for decreasing these serious side effects. Though several have been approved for clinical use, their reach is limited by many factors, including the availability of fundamental technologies that enable these therapies to target a broad range of diseases. My thesis addresses this problem and is best divided into two overarching efforts: developing fundamental technologies for cell-based biosensors and therapeutics (**Chapters 2-5**) and translating these cell-based devices for clinical applications (**Chapter 6**). Though time and the COVID-19 pandemic constrained me from fully translating every fundamental technology I developed, many ongoing efforts in the lab aim to harness, refine, and translate these technologies to develop novel therapeutics.

6.1 Project impact and future directions

COMET: a toolkit for controlling mammalian gene expression

In this study, we designed and characterized a toolkit for controlling gene expression in mammalian cells¹. COMET comprises activating and inhibitory transcription factors that act on a large library of engineered promoter, including those capable of integrating signals with Boolean logic. Our deep understanding of each design choice, its impacts on gene expression, and how these choices influence each other led to the development of a mathematical model, which we can use to describe these designs and guide future designs.

When we developed COMET, we planned to first employ it to transduce outputs from synthetic biosensors. Indeed, our models showed that many custom signal transduction pathways substantially enhanced the performance of such sensors and potentially the cell therapies that they compose. Joe Muldoon has led this project; together, he and I sketched out initial designs and experiments for testing these circuits and co-mentored IBIS rotation students (Jon Boucher and Tae-Eun Kim) through pilot

projects. Joe has since successfully employed these COMET-based circuits to implement various forms of signal transduction from biosensors, integrated the signals from multiple sensors into a single, logical output, and demonstrated predictive genetic circuit design in mammalian cells, a milestone achievement²⁶². Excitingly, the integration of COMET with existing MESA and other biosensors was seamless and generally resulted in very high functional performance.

Remote control of cellular transcription, by small-molecule chemical ligands, is a feature of COMET initially prototyped by a rapamycin inducible transcription factor¹. However, the availability of multiple chemically inducible transcription factors would greatly enhance the sophistication of genetic circuits that can be constructed with COMET and expand the applications of this technology. The first parts of this work are briefly mentioned in **Chapter 2**, in which we developed five new chemically inducible COMET transcription factors. These will ultimately be published along with a mathematical model, developed by Kate Dray and describing how these regulate transcription in response to the ligand, and a demonstration of their utility, in which we implement precise spatiotemporal control of gene expression with chemical ligands.

COMET has greatly enhanced our ability to engineer mammalian cells for many applications, including therapeutic ones. It has enabled many projects within our lab, has enhanced some of our biosensing technology, and has been applied to translational applications. We anticipate that it will be as useful to many other labs conducting mammalian synthetic and fundamental biology research. We are depositing approximately 250 plasmids containing COMET parts on Addgene, along with a toolkit of plasmids that enables easy adoption of this technology in other laboratories. Already, several labs have reached out to obtain these parts, and we expect that the full kit of components will be broadly distributed and widely adopted.

More broadly, we expect that components of the COMET toolkit will be useful in fundamental biology research and therapeutic technologies. tTA and Gal4 are broadly used but could easily be replaced with COMET TFs for many applications. The benefits to doing so include that COMET allows multiple synthetic pathways within any given cell, COMET parts are at baseline more potent than these workhorse TFs, and they are additionally readily tunable to customize programs for many applications. While many

COMET parts already exist, each with known parameters for how they modulate transcription, if an application called for a COMET part with a specific set of parameters and this part didn't already exist, it is likely that an existing part could be engineered to confer these parameters upon it. We have demonstrated that COMET integrates signals from multiplexed biosensors much better than tTA and Gal4 do, and we expect this finding to be translatable to the processing of signals from other sensors and native cellular pathways. When combining multiple inputs to increase the specificity of a cell-based biosensor for a pathologic environment, COMET will have enormous utility and could be easily plugged into many of the sensor systems listed in **Chapter 1**. Therefore, COMET will find widespread use across many laboratories and cell-based therapies.

Elucidation and refinement of synthetic receptor mechanisms

In this study, we investigated previously understudied components of MESA, a synthetic receptor system, conducting hypothesis driven research to elucidate which of these components were targets for enhancing receptor performance¹⁹⁹. Though tuning MESA protease kinetics and rendering the system less susceptible to intracellular proteolytic degradation did not markedly improve function, our investigation of the TMD was more impactful. Using different TMDs led to improved receptor performance and we probed into the mechanisms underlying these benefits. With this knowledge, we were able to build several new MESA systems against new ligands, with substantially less effort than had been required to build new MESA in the past (months versus years).

We also specifically identified an obstacle to building MESA against large ligands—large extracellular binding domains place the chains too far apart for them to signal. The solutions to this are readily envisioned (smaller binding domains) and implementing this strategy is currently an active area of research being carried out by Will Corcoran in our lab in collaboration with Gabe Rocklin (NU). We hope that these new binding domains will realize the full potential of MESA's modularity. Another interesting insight from this study is that while the lab's first-generation MESA required precise expression regimes for ligand-inducible signaling to occur, the data generated in this study suggest that the choices of TMD and intracellular signaling mechanism confer robustness to expression level. We are therefore investigating how

this property enables use of the more translationally relevant genomic expression of these biosensors.

Many other projects involving the translation of MESA towards clinical contexts are ongoing, led by younger graduate students. For instance, MESA-based cell therapies that sense and control inflammation in arthritis (Hailey Edelstein) and that recognize and treat various tumors (Will Corcoran, Simrita Deol) are actively being developed. Though I am not directly involved in these projects, I provide advice and mentorship on them, and my work, as well as the work of many others, with MESA over the past several years has enabled these translational advances.

More broadly, we envision MESA being used by other labs, as well, to detect antigens, both for cell-based therapies and as a tool in fundamental biology to probe changes in expression level of extracellular proteins over time. The plasmids from our first MESA study are highly requested from Addgene; submission of the plasmids from this study is ongoing and we expect that these will be in demand as well. As the only synthetic, orthogonal receptor system for studying soluble, extracellular ligands, MESA will play a role in the sensing of the various features of the TME described in **Chapter 1**, as well as markers of other pathologies we target cell-based therapies against.

Developing robust cell-based biosensors for hypoxia

In this work, we studied the properties of an HBS in the genome of HEK293FT landing pad cell. We synthesized findings from several prior works to enhance the performance of the biosensor through promoter engineering. We also discovered that, at least in HEK293FT cells, HIF1 α is a limiting reagent for the biosensor. However, when we implemented a positive feedback motif with HIFs in the B16F10 cells, this motif did not lead to improved performance. This finding emphasizes a theme from **Chapter 1**—the reliance of sensors on endogenous components can create issues with robust performance, especially when porting technologies between different cell or application contexts, as discussed further below. To address this, we are currently employing COMET to induce amplification and/or positive feedback on the signal from the HBS. This demonstration of the COMET technology, while intrinsically useful to our goals, will likely elucidate several strategies for processing the signal from synthetic biosensors and guidelines for doing so that will be useful to others in the field. Further, we have developed oxygen sensitive COMET TFs

to mitigate concerns about amplification of the background signal. This new set of oxygen sensitive TFs may have additional benefits, including design of hypoxia biosensors that do not rely on HIFs for sensing, as these are dysregulated in many cancer types.

Evidence from the studies presented in this thesis suggests that genetic circuits built with COMET could effectively modulate the response from the HBS to enhance performance. In order to study this, we have constructed approximately 40 such circuits and are integrating them into the B16F10-LP line. This project will be carried on by Kate Chambers and Jon Boucher after my return to medical school. In parallel, Katie Dreyer is building a computational model in order to understand how the HBS activates transcription, based on a model of the cell's response to hypoxia. Data from the experiments in this thesis, from our experiments with COMET, and from the testing the proposed HBS-modulating COMET circuits will be used to train this model and integrate it with the COMET model. From this, we will be able to describe the behavior of the HBS genetic circuits with COMET components, perform a sensitivity analysis to glean which parameters are most critical for biosensor performance, design a second round of genetic circuits, and predict how these circuits will modulate the signal from the HBS. The second round of circuits will be designed with the goal of conferring robustness to the choice of cell type, and we will test these predictions by evaluating biosensors predicted to be robust and to be not robust across a range of cancer cell types. Ultimately, these efforts will produce high-performing hypoxia biosensors that can be used across a range of cell types, both in a therapeutic context and to conduct fundamental research. We intend to demonstrate the application for fundamental research in cancer biology by investigating how hypoxia develops over time in a B16F10 model of melanoma *in vivo*, using one of the computationally designed robust circuits. The reporter proteins in this study were designed to be analyzed by flow cytometry and the IVIS *in vivo* imager, allowing for ready porting of this work to an *in vivo* system after *in vitro* analyses.

As exemplified with MINK in **Chapter 6**, we believe that the enhancements we have made to the HBS, and can make by processing the signal with COMET, will improve the specificity and safety profiles of cell therapies. Hypoxia is an important pathological marker not only in cancer, but also in stroke, myocardial infarction, and diabetic ulcers, and it is possible that cells could use this biosensor to sense hypoxic conditions and deliver therapeutics to provide vascular support, stimulate cardiomyocyte regrowth,

and improve blood flow to the wounds.

During these studies, we also learned much about working with cells with landing pads, including improved methodologies for recovering the rare clonal integration events that occur with some cell lines. We also validated several techniques for working with cells in hypoxia. For instance, we are now able to make educated choices about fluorophores, so that these reporters have high signal to noise ratios and oxidize rapidly and nearly-completely. We generated a B16F10-LP line that we will first use to study hypoxia in murine tumors, but could also be used to prototype, study, and optimize other biosensors. One advantage to prototyping biosensors in a cancer line is that it decouples the performance of the biosensor from the trafficking of the therapeutic cells. Though this may speed up the development of biosensing technologies, it is also important to consider that sensors may perform differently in different cell type contexts and this methodology does not study the specificity of the marker the biosensor recognizes for the TME.

Tumor microenvironment induced natural killer (MINK) cells for cancer immunotherapy

In this work we reduced to practice the fundamental concept underlying the MINK technology—NK cells can be engineered to recognize hallmarks of the TME and respond by activating targeted expression of a gene. We demonstrated that the HBS enables the NK-92 cells to report on hypoxia and that, surprisingly, some of the design rules do not translate from our studies in **Chapter 4** and from other previously published works. We anticipate that MINK will find a role in the treatment of a broad range of cancers, as the underlying technology relies on sensing hallmarks of cancer that are intrinsic to many tumors. In this initial demonstration, MINK sensed hypoxia through an HBS, but also could be VEGF as sensed by a MESA, or any of the other cues elaborated upon in **Chapter 1** with other sensors mentioned therein, in this work, or that haven't yet been developed. While MINK may very well prove to be a useful standalone therapeutic, with the cytokines it secretes stimulating MINK cells and those from the host's immune system, the cytokines or other produced molecules could be chosen as to support and synergize with other therapeutics, including CAR T-Cells, CAR macrophages, and immune checkpoint blockades.

A major part of this work was establishing the methods needed to engineer these cells and conduct *in vivo* evaluations, which are mostly new to our lab. Having validated a BRET system and an orthogonal,

far-red luciferase system will allow us to image MINK and other cell therapies that are currently in development in our lab *in vivo* with high sensitivity. In fact, the new standards I developed while conducting the studies in **Chapters 4-5** allow multimodal (microscopy, histology, flow cytometry, plate reader, and IVIS *in vitro/in vivo* imaging) analysis and will be useful for many lab members going forward. We look forward to publishing and sharing the protocols and reagents for these methods with the broader community.

Future work on this project includes studying the distribution of the hypoxia-responsive NK-92s *in vivo* and examination of where they are reporting on the hypoxia, using the BRET reporter. This will be carried out by Simrita Deol, in collaboration with members of the Bin Zhang lab (NU). The YB_TATA-based HBS line, sorted for the brightest miFRP720 cells, is the most appropriate line for this study, as it has a high fold induction and very bright reporter signal. However, this line expresses a Firefly Luciferase, and *in vivo* cell detection would be better if it were additionally transduced with the AkaLuc gene. Whether or not AkaLuc will be necessary to detect the cells *in vivo* remains to be seen, as the engineered NK-92s also express an miRFP720 gene that may allow for *in vivo* tracking. In the future, in order to better time resolve the sensing of hypoxia by MINK cells, we could use a LumiScarlet variant with a shortened half-life; alternatively, one could engineer the LumiScarlet to be unstable in the presence of oxygen, similar to the COMET TFs in **Chapter 4**. Additionally, it will be important to perform a thorough investigation of the effects of hypoxia induced cytokines produced by MINK *in vitro*, prior to a pre-clinical study, to ensure that they are secreted and bioactive, allowing them to act in *trans* on other cell types, particularly those of the hosts immune system. After this work has been completed, preclinical studies will examine the *in vivo* effects of these therapeutic cytokines as produced by MINK on tumor growth and immune composition. It may also be interesting to see whether MINK can be engineered to be synergistic with existing immunotherapies, such as CAR T cells or checkpoint inhibitors.

The MINK concept holds promise for the treatment of solid tumors with immunotherapy. Though the first prototype senses hypoxia, future MINK may be able to sense other markers of the TME, as outline in **Chapter 1**. COMET could be used to integrate the signal from multiple biosensors to increase the specificity of MINK for solid tumors. Further if on-target, off-tissue side effects are seen with MINK, one could use MESA to program MINK to recognize this organ and respond by shutting down the effector

functions, perhaps mediated through an inhibitory COMET TF. Remote control of the MINK therapeutic could also be possible with the various CID systems, leading to enhance safety through several mechanisms. These ligands could be used to enable MINK to function, perhaps by inducing expression of a biosensor or a requisite component of the signaling pathway, or to block the function of MINK, perhaps by inducing the expression of an inhibitory COMET TF or a protein that promotes apoptosis.

With MINK as an established technology, the door to many future investigations of non-T cell immunotherapies opens further. For instances, the same technologies used by MINK to sense the TME may be used in other cell types to confer specificity for the TME. This is possible because MINK do not rely on an effector function specific to NK cells—the secretion of cytokines could be done by many cell types, and then this choice can be guided by information about the cellular composition of tumors in various types of cancers rather than restricted by the need for the engineered cell to be capable of cytotoxic effector functions. As an example, macrophages could be engineered to sense hypoxia and respond by producing therapeutic cytokines. This may have benefits in tumors that show heavy recruitment of macrophages and monocytes. In conclusion, MINK is an example of the power of synthetic biology for cell-based therapies—by decreasing our reliance on endogenous proteins and functions, we increase the freedom in design choices and effector functions, so that these can be more readily selected for a broad range of diseases.

6.2 Recent progress and challenges in mammalian synthetic biology

Achievements and obstacles to mammalian cell engineering

We have achieved important milestones related to mammalian cell engineering, developing a toolkit for composing genetic programs (**Chapter 2**)¹, predictive design of genetic programs in mammalian cells²⁶², developing sensors against extracellular antigens (**Chapter 3**)¹⁹⁹ and environmental signals (**Chapter 4**), and applying these to engineer a novel NK cell-based therapy (**Chapter 6**), the field. These technologies join the ranks of other technologies for engineering mammalian cells, notably CARs and SynNotch, which have enabled many cell-based therapies currently in clinical and pre-clinical trials.

However, engineering mammalian cells is a time-consuming process. Outside of the fact that mammalian cells have doubling times that are substantially slower than bacteria and yeast, making this

type of work intrinsically slow, two major issues that slow progress currently stand out. The context in which any given technology is tested is crucial to its function – context here encompasses both cell type and DNA delivery method – and it is not always clear which options for these two choices are the best to study a new technology in.

Lack of robustness to cell type context

Cell type choices always present trade-offs. We often initially use chassis cell lines, such as HEK293FT, CHO, or HeLa, rather than more application-specific lines, as they grow rapidly and have tried and true methods for culture, DNA delivery, and analysis. However, principles elucidated and systems optimized in this cell type context may not map to other cell type contexts. One strategy to avoid this is to develop systems that are intrinsically robust to the things that vary between cell types, such as minimizing reliance on endogenously expressed components, such as certain receptors or transcription factors. Additionally, one could study this robustness computationally, by building thorough models that include parameters that would vary between cells, such as cell growth rate or concentrations of endogenous factors upon which the sensor relies, and examining system performance over the physiologic range of these parameters. One would then only move forward with designs that are robust to variations in these parameters, and principles could be gleaned from these efforts that would apply to the future design of robust systems. Other solutions include developing better methods for a broader range of cell lines, which is actively being explored by many groups and research supply companies.

Moving towards translationally relevant strategies, another choice is whether to use cells that are immortalized or primary. Though immortalized cells are cheaper, easier to work with, and may speed up the initial development phase of the research, it is again unclear how work done in these cells, that may or may not be physiologically representative of their primary counterparts, would map to primary cells, if those were ultimately to be used in a therapeutic. The solutions here are numerous: develop strategies for using immortalized cells in patients, develop strategies for faster, more high-throughput use of primary cells in laboratory research, develop immortalized cell lines that more closely mimic their primary cell counterparts, develop methods for translating work done in immortalized cells to the primary cell context, or develop

systems that are robust to the variations between these contexts.

Lack of robustness to DNA delivery context

The second important context is DNA delivery method. Frequently, transfection-based methodologies are employed for speed, practicality, and reliability. However, this methodology is often transient, not providing a therapeutically relevant strategy for most applications, the number of gene copies and thus the expression of each component is much higher than would be in a more relevant context, and there is no guarantee that the findings of studies of systems expressed from transiently transfected plasmids will map on to the behavior of these systems when they are expressed from a genomic locus. More therapeutically relevant alternatives to transfection include lentiviral integration, adenovirus or AAV transduction, HDR integration after Cas9 dsDNA breaks, or transposases. However, generating stable cell lines is a time-consuming process, and the heterogeneity arising from the random integration sites can be problematic. Solutions to this include developing systems that are robust to delivery methodology using the computational methods described above, elucidating rulesets by which behavior of transiently transfected components can be mapped into the genomic context, or developing methodologies for the more rapid generation of stable cell lines. Some of these efforts are well underway in our lab. For instance, elucidating the rules by which the quantitative framework in which COMET induces gene expression, as discovered by transient transfection, map to a framework in the genome is an active area of investigation.

Silencing of gene expression

Further complicating cell engineering is the fact that genomically integrated parts are prone to epigenetic silencing. This phenomena and strategies for mitigating it are active areas of research, as this is a pressing issue affecting not just system development but these systems when they are deployed as therapeutics. Solutions may include the use of parts (such as chromatin insulators) that mitigate the spread of these epigenetic marks, the use of chromatin-altering transcription factors to force the DNA to remain open, and others targeted at the root causes as indicated by the ongoing studies. While landing pads are a step in the right direction for this field and address many of the issues, they still are time consuming to

generate, require immortalized cell lines, and are not intrinsically immune to silencing. One finding from this work is that sorting for the top 10% of constitutive protein expressing cells, when the constructs are integrated into an LP, confers resistance to genetic silencing. As this phenomenon is heritable and stable, it is possible that there is some sort of mark made on the integration locus, which is the same locus for all cells in a population, at the time of integration that destines some cells for silencing and others for long-term stable expression. Cells with the hypothesized mark can therefore be isolated with FACS and cultured to promote long-term, stable expression and activity of integrated biosensors.

6.3 Next generation therapeutics for diseases beyond cancer

While my thesis work focuses on cancer, cell-based therapies are an exciting new class of drugs with applications that span many diseases, as the paradigm I lay out in the **Chapter 1** (sense, process, effect). Various pathologic states are marked by characteristic patterns of abnormal inflammation (i.e., sarcoidosis, atherosclerosis, psoriasis), abnormal local accumulation of metabolites (i.e., the crystalline arthropathies) or proteins (i.e., Alzheimers, Lewy Body dementia, and amyloidosis) or cells (i.e., Langerhans cell histiocytosis and cardiac fibrosis), abnormal levels of hormones or metabolites systemically (i.e., diseases of the hypothalamic pituitary axis, thyroid and parathyroid glands, and endocrine diseases of other organs), or other abnormal states that represent a departure from the tightly controlled homeostasis of the human body. Custom effector programs can be designed that counteract the abnormal inflammation, metabolize small molecule depositions, endocytose and degrade excess protein, kill infiltrative, pathologic cells, and supplement or counteract the dysregulation of the endocrine system. The technologies developed in **Chapters 2-4** will, in combination with other cell engineering and biosensing technologies, serve as the building blocks for many of these novel therapeutics, as exemplified by MINK in **Chapter 5**.

References

1. Donahue PS, Draut JW, Muldoon JJ, Edelstein HI, Bagheri N, Leonard JN. The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat Commun* **11**, 779 (2020).
2. Schwarz KA, Daringer NM, Dolberg TB, Leonard JN. Rewiring human cellular input-output using modular extracellular sensors. *Nat Chem Biol* **13**, 202-209 (2017).
3. Dolberg TB, *et al.* Computation-guided optimization of split protein systems. *bioRxiv*, (2019).
4. Kapust RB, *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993-1000 (2001).
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30 (2020).
6. Connell PP, Hellman S. Advances in radiotherapy and implications for the next century: a historical perspective. *Cancer Res* **69**, 383-392 (2009).
7. DeVita VT, Jr., Chu E. A history of cancer chemotherapy. *Cancer Res* **68**, 8643-8653 (2008).
8. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
9. Krzyszczyk P, *et al.* The growing role of precision and personalized medicine for cancer treatment. *Technology (Singap World Sci)* **6**, 79-100 (2018).
10. Zhang X, Yang H, Zhang R. Challenges and future of precision medicine strategies for breast cancer based on a database on drug reactions. *Biosci Rep* **39**, (2019).
11. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
12. McCarthy EF. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. *Iowa Orthop J* **26**, 154-158 (2006).
13. Dobosz P, Dzieciatkowski T. The Intriguing History of Cancer Immunotherapy. *Front Immunol* **10**, 2965 (2019).
14. Morales A, Eiding D, Bruce AW. Intracavitary Bacillus Calmette-Guerin in the treatment of superficial bladder tumors. *J Urol* **116**, 180-183 (1976).
15. Redelman-Sidi G, Glickman MS, Bochner BH. The mechanism of action of BCG therapy for bladder cancer--a current perspective. *Nat Rev Urol* **11**, 153-162 (2014).
16. Berraondo P, *et al.* Cytokines in clinical cancer immunotherapy. *Br J Cancer* **120**, 6-15 (2019).
17. Klapper JA, *et al.* High-dose interleukin-2 for the treatment of metastatic renal cell carcinoma : a retrospective analysis of response and survival in patients treated in the surgery branch at the National Cancer Institute between 1986 and 2006. *Cancer* **113**, 293-301 (2008).
18. Daud AI, *et al.* Phase I trial of interleukin-12 plasmid electroporation in patients with metastatic melanoma. *J Clin Oncol* **26**, 5896-5903 (2008).
19. Mortier E, *et al.* Soluble interleukin-15 receptor alpha (IL-15R alpha)-sushi as a selective and potent agonist of IL-15 action through IL-15R beta/gamma. Hyperagonist IL-15 x IL-15R alpha fusion proteins. *J Biol Chem* **281**, 1612-1619 (2006).

20. Wrangle JM, *et al.* ALT-803, an IL-15 superagonist, in combination with nivolumab in patients with metastatic non-small cell lung cancer: a non-randomised, open-label, phase 1b trial. *Lancet Oncol* **19**, 694-704 (2018).
21. Timmerman JM, *et al.* A phase I dose-finding trial of recombinant interleukin-21 and rituximab in relapsed and refractory low grade B-cell lymphoproliferative disorders. *Clin Cancer Res* **18**, 5752-5760 (2012).
22. Andtbacka RH, *et al.* Talimogene Laherparepvec Improves Durable Response Rate in Patients With Advanced Melanoma. *J Clin Oncol* **33**, 2780-2788 (2015).
23. Sharpe AH. Introduction to checkpoint inhibitors and cancer immunotherapy. *Immunol Rev* **276**, 5-8 (2017).
24. Small EJ, *et al.* Placebo-controlled phase III trial of immunologic therapy with sipuleucel-T (APC8015) in patients with metastatic, asymptomatic hormone refractory prostate cancer. *J Clin Oncol* **24**, 3089-3094 (2006).
25. Cruz CR, Bollard CM. Adoptive Immunotherapy For Leukemia With Ex vivo Expanded T Cells. *Curr Drug Targets* **18**, 271-280 (2017).
26. Franks SE, Wolfson B, Hodge JW. Natural Born Killers: NK Cells in Cancer Therapy. *Cancers (Basel)* **12**, (2020).
27. Boyiadzis MM, *et al.* Chimeric antigen receptor (CAR) T therapies for the treatment of hematologic malignancies: clinical perspective and significance. *J Immunother Cancer* **6**, 137 (2018).
28. Villadolid J, Amin A. Immune checkpoint inhibitors in clinical practice: update on management of immune-related toxicities. *Transl Lung Cancer Res* **4**, 560-575 (2015).
29. Hodi FS, *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* **363**, 711-723 (2010).
30. Weber JS, *et al.* Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* **16**, 375-384 (2015).
31. Kaufman HL, *et al.* Avelumab in patients with chemotherapy-refractory metastatic Merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. *Lancet Oncol* **17**, 1374-1385 (2016).
32. Robert C, *et al.* Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial. *Lancet* **384**, 1109-1117 (2014).
33. Migden MR, *et al.* PD-1 Blockade with Cemiplimab in Advanced Cutaneous Squamous-Cell Carcinoma. *N Engl J Med* **379**, 341-351 (2018).
34. Massard C, *et al.* Safety and Efficacy of Durvalumab (MEDI4736), an Anti-Programmed Cell Death Ligand-1 Immune Checkpoint Inhibitor, in Patients With Advanced Urothelial Bladder Cancer. *J Clin Oncol* **34**, 3119-3125 (2016).
35. Necchi A, *et al.* Atezolizumab in platinum-treated locally advanced or metastatic urothelial

- carcinoma: post-progression outcomes from the phase II IMvigor210 study. *Ann Oncol* **28**, 3044-3050 (2017).
36. Vaddepally RK, Kharel P, Pandey R, Garje R, Chandra AB. Review of Indications of FDA-Approved Immune Checkpoint Inhibitors per NCCN Guidelines with the Level of Evidence. *Cancers (Basel)* **12**, (2020).
 37. Larkin J, *et al.* Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* **373**, 23-34 (2015).
 38. Lim WA, June CH. The Principles of Engineering Immune Cells to Treat Cancer. *Cell* **168**, 724-740 (2017).
 39. Kalos M, June CH. Adoptive T cell transfer for cancer immunotherapy in the era of synthetic biology. *Immunity* **39**, 49-60 (2013).
 40. Rosenberg SA, *et al.* Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N Engl J Med* **319**, 1676-1680 (1988).
 41. Park TS, Rosenberg SA, Morgan RA. Treating cancer with genetically engineered T cells. *Trends Biotechnol* **29**, 550-557 (2011).
 42. Coccoris M, Straetemans T, Govers C, Lamers C, Sleijfer S, Debets R. T cell receptor (TCR) gene therapy to treat melanoma: lessons from clinical and preclinical studies. *Expert Opin Biol Ther* **10**, 547-562 (2010).
 43. Morgan RA, *et al.* High efficiency TCR gene transfer into primary human lymphocytes affords avid recognition of melanoma tumor antigen glycoprotein 100 and does not alter the recognition of autologous melanoma antigens. *J Immunol* **171**, 3287-3295 (2003).
 44. Zhang J, Wang L. The Emerging World of TCR-T Cell Trials Against Cancer: A Systematic Review. *Technol Cancer Res Treat* **18**, 1533033819831068 (2019).
 45. Maus MV, Grupp SA, Porter DL, June CH. Antibody-modified T cells: CARs take the front seat for hematologic malignancies. *Blood* **123**, 2625-2635 (2014).
 46. Johnson LA, June CH. Driving gene-engineered T cell immunotherapy of cancer. *Cell Res* **27**, 38-58 (2017).
 47. Weinkove R, George P, Dasyam N, McLellan AD. Selecting costimulatory domains for chimeric antigen receptors: functional and clinical considerations. *Clin Transl Immunology* **8**, e1049 (2019).
 48. Kochenderfer JN, *et al.* Eradication of B-lineage cells and regression of lymphoma in a patient treated with autologous T cells genetically engineered to recognize CD19. *Blood* **116**, 4099-4102 (2010).
 49. Cerrano M, *et al.* The Advent of CAR T-Cell Therapy for Lymphoproliferative Neoplasms: Integrating Research Into Clinical Practice. *Front Immunol* **11**, 888 (2020).
 50. Locke FL, *et al.* Long-term safety and activity of axicabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1): a single-arm, multicentre, phase 1-2 trial. *Lancet Oncol* **20**, 31-42 (2019).
 51. Schuster SJ, *et al.* Chimeric Antigen Receptor T Cells in Refractory B-Cell Lymphomas. *N Engl J Med* **377**, 2545-2554 (2017).

52. Schuster SJ, *et al.* Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *N Engl J Med* **380**, 45-56 (2019).
53. Maude SL, *et al.* Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N Engl J Med* **378**, 439-448 (2018).
54. Maude SL, *et al.* Chimeric antigen receptor T cells for sustained remissions in leukemia. *N Engl J Med* **371**, 1507-1517 (2014).
55. Abramson JS, *et al.* Lisocabtagene maraleucel for patients with relapsed or refractory large B-cell lymphomas (TRANSCEND NHL 001): a multicentre seamless design study. *Lancet* **396**, 839-852 (2020).
56. Schultz L, Mackall C. Driving CAR T cell translation forward. *Sci Transl Med* **11**, (2019).
57. Rafiq S, Hackett CS, Brentjens RJ. Engineering strategies to overcome the current roadblocks in CAR T cell therapy. *Nat Rev Clin Oncol* **17**, 147-167 (2020).
58. Lyman GH, Nguyen A, Snyder S, Gitlin M, Chung KC. Economic Evaluation of Chimeric Antigen Receptor T-Cell Therapy by Site of Care Among Patients With Relapsed or Refractory Large B-Cell Lymphoma. *JAMA Netw Open* **3**, e202072 (2020).
59. Townsend MH, Bennion K, Robison RA, O'Neill KL. Paving the way towards universal treatment with allogenic T cells. *Immunol Res* **68**, 63-70 (2020).
60. Ajina A, Maher J. Strategies to Address Chimeric Antigen Receptor Tonic Signaling. *Mol Cancer Ther* **17**, 1795-1815 (2018).
61. Eyquem J, *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113-117 (2017).
62. Wagner J, Wickman E, DeRenzo C, Gottschalk S. CAR T Cell Therapy for Solid Tumors: Bright Future or Dark Reality? *Mol Ther*, (2020).
63. Chmielewski M, Hombach AA, Abken H. Of CARs and TRUCKs: chimeric antigen receptor (CAR) T cells engineered with an inducible cytokine to modulate the tumor stroma. *Immunol Rev* **257**, 83-90 (2014).
64. Rafiq S, *et al.* Targeted delivery of a PD-1-blocking scFv by CAR-T cells enhances anti-tumor efficacy in vivo. *Nat Biotechnol* **36**, 847-856 (2018).
65. Grosser R, Cherkassky L, Chintala N, Adusumilli PS. Combination Immunotherapy with CAR T Cells and Checkpoint Blockade for the Treatment of Solid Tumors. *Cancer Cell* **36**, 471-482 (2019).
66. Morgan MA, Buning H, Sauer M, Schambach A. Use of Cell and Genome Modification Technologies to Generate Improved "Off-the-Shelf" CAR T and CAR NK Cells. *Front Immunol* **11**, 1965 (2020).
67. Abbott RC, Cross RS, Jenkins MR. Finding the Keys to the CAR: Identifying Novel Target Antigens for T Cell Redirection Immunotherapies. *Int J Mol Sci* **21**, (2020).
68. Choi BD, *et al.* CAR-T cells secreting BiTEs circumvent antigen escape without detectable toxicity. *Nat Biotechnol* **37**, 1049-1058 (2019).

69. Walsh Z, Ross S, Fry TJ. Multi-Specific CAR Targeting to Prevent Antigen Escape. *Curr Hematol Malig Rep* **14**, 451-459 (2019).
70. Park AK, *et al.* Effective combination immunotherapy using oncolytic viruses to deliver CAR targets to solid tumors. *Sci Transl Med* **12**, (2020).
71. Hicklin DJ, Ellis LM. Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. *J Clin Oncol* **23**, 1011-1027 (2005).
72. Streit M, Detmar M. Angiogenesis, lymphangiogenesis, and melanoma metastasis. *Oncogene* **22**, 3172-3179 (2003).
73. Salven P, Heikkila P, Joensuu H. Enhanced expression of vascular endothelial growth factor in metastatic melanoma. *Br J Cancer* **76**, 930-934 (1997).
74. Simonetti O, *et al.* Immunohistochemical expression of vascular endothelial growth factor, matrix metalloproteinase 2, and matrix metalloproteinase 9 in cutaneous melanocytic lesions. *Cancer* **95**, 1963-1970 (2002).
75. Skobe M, *et al.* Concurrent induction of lymphangiogenesis, angiogenesis, and macrophage recruitment by vascular endothelial growth factor-C in melanoma. *Am J Pathol* **159**, 893-903 (2001).
76. Kut C, Mac Gabhann F, Popel AS. Where is VEGF in the body? A meta-analysis of VEGF distribution in cancer. *Br J Cancer* **97**, 978-985 (2007).
77. Witsch E, Sela M, Yarden Y. Roles for growth factors in cancer progression. *Physiology (Bethesda)* **25**, 85-101 (2010).
78. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* **27**, 5904-5912 (2008).
79. Hockel M, Vaupel P. Tumor hypoxia: definitions and current clinical, biologic, and molecular aspects. *J Natl Cancer Inst* **93**, 266-276 (2001).
80. Vaupel P, Kelleher DK, Hockel M. Oxygen status of malignant tumors: pathogenesis of hypoxia and significance for tumor therapy. *Semin Oncol* **28**, 29-35 (2001).
81. Vaupel P, Okunieff P, Neuringer LJ. Blood flow, tissue oxygenation, pH distribution, and energy metabolism of murine mammary adenocarcinomas during growth. *Adv Exp Med Biol* **248**, 835-845 (1989).
82. Siemann DW. The unique characteristics of tumor vasculature and preclinical evidence for its selective disruption by Tumor-Vascular Disrupting Agents. *Cancer Treat Rev* **37**, 63-74 (2011).
83. Michiels C. Physiological and pathological responses to hypoxia. *Am J Pathol* **164**, 1875-1882 (2004).
84. Lee JY, *et al.* A novel chimeric promoter that is highly responsive to hypoxia and metals. *Gene Ther* **13**, 857-868 (2006).
85. Iglesias P, Costoya JA. A novel BRET-based genetically encoded biosensor for functional imaging of hypoxia. *Biosens Bioelectron* **24**, 3126-3130 (2009).

86. Iglesias P, Fraga M, Costoya JA. Defining hypoxic microenvironments by non-invasive functional optical imaging. *Eur J Cancer* **49**, 264-271 (2013).
87. Ede C, Chen X, Lin MY, Chen YY. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synth Biol* **5**, 395-404 (2016).
88. Gurusamy D, Clever D, Eil R, Restifo NP. Novel "Elements" of Immune Suppression within the Tumor Microenvironment. *Cancer Immunol Res* **5**, 426-433 (2017).
89. Eil R, *et al.* Ionic immune suppression within the tumour microenvironment limits T cell effector function. *Nature* **537**, 539-543 (2016).
90. Pavlova NN, Thompson CB. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab* **23**, 27-47 (2016).
91. Helmlinger G, Yuan F, Dellian M, Jain RK. Interstitial pH and pO₂ gradients in solid tumors in vivo: high-resolution measurements reveal a lack of correlation. *Nat Med* **3**, 177-182 (1997).
92. Sun C, *et al.* Spatially resolved metabolomics to discover tumor-associated metabolic alterations. *Proc Natl Acad Sci U S A* **116**, 52-57 (2019).
93. Espinoza JA, *et al.* Cytokine profiling of tumor interstitial fluid of the breast and its relationship with lymphocyte infiltration and clinicopathological characteristics. *Oncoimmunology* **5**, e1248015 (2016).
94. Kawaguchi K, *et al.* Alteration of specific cytokine expression patterns in patients with breast cancer. *Sci Rep* **9**, 2924 (2019).
95. Lippitz BE, Harris RA. Cytokine patterns in cancer patients: A review of the correlation between interleukin 6 and prognosis. *Oncoimmunology* **5**, e1093722 (2016).
96. Kim YW, Kim SK, Kim CS, Kim IY, Cho MY, Kim NK. Association of serum and intratumoral cytokine profiles with tumor stage and neutrophil lymphocyte ratio in colorectal cancer. *Anticancer Res* **34**, 3481-3487 (2014).
97. Lu C, *et al.* Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: challenges and opportunities. *Mol Cancer* **18**, 130 (2019).
98. Sabi R, Tuller T. Modelling and measuring intracellular competition for finite resources during gene expression. *J R Soc Interface* **16**, 20180887 (2019).
99. Frei T, *et al.* Characterization and mitigation of gene expression burden in mammalian cells. *Nat Commun* **11**, 4641 (2020).
100. Qian Y, Huang HH, Jiménez JI, Del Vecchio D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synth Biol* **6**, 1263-1272 (2017).
101. Darlington APS, Kim J, Jiménez JI, Bates DG. Design of a translation resource allocation controller to manage cellular resource limitations. *IFAC-PapersOnLine* **50**, 12653 - 12660 (2017).
102. Carlson ED, *et al.* Engineered ribosomes with tethered subunits for expanding biological function. *Nat Commun* **10**, 3920 (2019).

103. Darlington APS, Kim J, Jiménez JI, Bates DG. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat Commun* **9**, 695 (2018).
104. Jones RD, *et al.* An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *bioRxiv*, 867028 (2020).
105. Muldoon JJ, Donahue PS, Dolberg TB, Leonard JN. Building with intent: technologies and principles for engineering mammalian cell-based therapies to sense and respond. *Curr Opin Biomed Eng* **4**, 127-133 (2017).
106. Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. *Nature* **537**, 320-327 (2016).
107. Baker D. What has de novo protein design taught us about protein folding and biophysics? *Protein Sci* **28**, 678-683 (2019).
108. Espona-Noguera A, *et al.* Review of Advanced Hydrogel-Based Cell Encapsulation Systems for Insulin Delivery in Type 1 Diabetes Mellitus. *Pharmaceutics* **11**, (2019).
109. Agustin-Pavon C, Mielcarek M, Garriga-Canut M, Isalan M. Deimmunization for gene therapy: host matching of synthetic zinc finger constructs enables long-term mutant Huntingtin repression in mice. *Mol Neurodegener* **11**, 64 (2016).
110. Morsut L, *et al.* Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **164**, 780-791 (2016).
111. Roybal KT, *et al.* Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors. *Cell* **167**, 419-432 e416 (2016).
112. Scheller L, Strittmatter T, Fuchs D, Bojar D, Fussenegger M. Generalized extracellular molecule sensor platform for programming cellular behavior. *Nat Chem Biol* **14**, 723-729 (2018).
113. Doucette C, Vedvik K, Koepnick E, Bergsma A, Thomson B, Turek-Etienne TC. Kappa opioid receptor screen with the Tango beta-arrestin recruitment technology and characterization of hits with second-messenger assays. *J Biomol Screen* **14**, 381-394 (2009).
114. Kipniss NH, *et al.* Engineering cell sensing and responses using a GPCR-coupled CRISPR-Cas system. *Nat Commun* **8**, 2212 (2017).
115. Daringer NM, Dudek RM, Schwarz KA, Leonard JN. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synth Biol* **3**, 892-902 (2014).
116. Hou AJ, Chang ZL, Lorenzini MH, Zah E, Chen YY. TGF-beta-responsive CAR-T cells promote anti-tumor immune function. *Bioeng Transl Med* **3**, 75-86 (2018).
117. Hartfield RM, Schwarz KA, Muldoon JJ, Bagheri N, Leonard JN. Multiplexing Engineered Receptors for Multiparametric Evaluation of Environmental Ligands. *ACS Synth Biol* **6**, 2042-2055 (2017).
118. Majerle A, Gaber R, Bencina M, Jerala R. Function-based mutation-resistant synthetic signaling device activated by HIV-1 proteolysis. *ACS Synth Biol* **4**, 667-672 (2015).
119. Ho P, Ede C, Chen YY. Modularly Constructed Synthetic Granzyme B Molecule Enables Interrogation of Intracellular Proteases for Targeted Cytotoxicity. *ACS Synth Biol* **6**, 1484-1495

- (2017).
120. Siciliano V, *et al.* Engineering modular intracellular protein sensor-actuator devices. *Nat Commun* **9**, 1881 (2018).
 121. Xu C, *et al.* Computational design of transmembrane pores. *Nature* **585**, 129-134 (2020).
 122. Mohammadi P, Beerenwinkel N, Benenson Y. Automated Design of Synthetic Cell Classifier Circuits Using a Two-Step Optimization Strategy. *Cell Syst* **4**, 207-218 e214 (2017).
 123. 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).
 124. Smole A, Lainscek D, Bezeljak U, Horvat S, Jerala R. A Synthetic Mammalian Therapeutic Gene Circuit for Sensing and Suppressing Inflammation. *Mol Ther* **25**, 102-119 (2017).
 125. Gossen M, Freundlieb S, Bender G, Muller G, Hillen W, Bujard H. Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769 (1995).
 126. Gossen M, Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547-5551 (1992).
 127. Kakidani H, Ptashne M. GAL4 activates gene expression in mammalian cells. *Cell* **52**, 161-167 (1988).
 128. Stanton BC, *et al.* Systematic transfer of prokaryotic sensors and circuits to mammalian cells. *ACS Synth Biol* **3**, 880-891 (2014).
 129. Karzenowski D, Potter DW, Padidam M. Inducible control of transgene expression with ecdysone receptor: gene switches with high sensitivity, robust expression, and reduced size. *Biotechniques* **39**, 191-192, 194, 196 passim (2005).
 130. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
 131. Li Y, *et al.* Modular construction of mammalian gene circuits using TALE transcriptional repressors. *Nat Chem Biol* **11**, 207-213 (2015).
 132. Boch J, *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512 (2009).
 133. Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A* **100**, 12271-12276 (2003).
 134. Lohmueller JJ, Armel TZ, Silver PA. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res* **40**, 5180-5187 (2012).
 135. Gilbert LA, *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451 (2013).
 136. Weinberg BH, *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotechnol* **35**, 453-462 (2017).

137. Xie M, *et al.* beta-cell-mimetic designer cells provide closed-loop glycemic control. *Science* **354**, 1296-1301 (2016).
138. Baeumler TA, Ahmed AA, Fulga TA. Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors. *Cell Rep* **20**, 2639-2653 (2017).
139. Krawczyk K, Scheller L, Kim H, Fussenegger M. Rewiring of endogenous signaling pathways to genomic targets for therapeutic cell reprogramming. *Nat Commun* **11**, 608 (2020).
140. Nielsen AA, *et al.* Genetic circuit design automation. *Science* **352**, aac7341 (2016).
141. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707-723 (2017).
142. Bonaventura P, *et al.* Cold Tumors: A Therapeutic Challenge for Immunotherapy. *Front Immunol* **10**, 168 (2019).
143. Whiteside TL. Immune responses to malignancies. *J Allergy Clin Immunol* **125**, S272-283 (2010).
144. Disis ML. Immune regulation of cancer. *J Clin Oncol* **28**, 4531-4538 (2010).
145. Cohen J. IL-12 deaths: explanation and a puzzle. *Science* **270**, 908 (1995).
146. Car BD, Eng VM, Lipman JM, Anderson TD. The toxicology of interleukin-12: a review. *Toxicol Pathol* **27**, 58-63 (1999).
147. Srivastava N, McDermott D. Update on benefit of immunotherapy and targeted therapy in melanoma: the changing landscape. *Cancer Manag Res* **6**, 279-289 (2014).
148. Tugues S, *et al.* New insights into IL-12-mediated tumor suppression. *Cell Death Differ* **22**, 237-246 (2015).
149. Leonard JP, *et al.* Effects of single-dose interleukin-12 exposure on interleukin-12-associated toxicity and interferon-gamma production. *Blood* **90**, 2541-2548 (1997).
150. Liao W, Lin JX, Leonard WJ. IL-2 family cytokines: new insights into the complex roles of IL-2 as a broad regulator of T helper cell differentiation. *Curr Opin Immunol* **23**, 598-604 (2011).
151. Lieschke GJ, Rao PK, Gately MK, Mulligan RC. Bioactive murine and human interleukin-12 fusion proteins which retain antitumor activity in vivo. *Nat Biotechnol* **15**, 35-40 (1997).
152. Bessard A, Sole V, Bouchaud G, Quemener A, Jacques Y. High antitumor activity of RLI, an interleukin-15 (IL-15)-IL-15 receptor alpha fusion protein, in metastatic melanoma and colorectal cancer. **8**, 2736-2745 (2009).
153. Silva DA, *et al.* De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186-191 (2019).
154. Zhou T, *et al.* IL-18BP is a secreted immune checkpoint and barrier to IL-18 immunotherapy. *Nature* **583**, 609-614 (2020).
155. Serakinci N, Cagsin H. Programming hMSCs into Potential Genetic Therapy in Cancer. *Crit Rev Eukaryot Gene Expr* **29**, 343-350 (2019).

156. Dirkx AE, Oude Egbrink MG, Wagstaff J, Griffioen AW. Monocyte/macrophage infiltration in tumors: modulators of angiogenesis. *J Leukoc Biol* **80**, 1183-1196 (2006).
157. Fang F, Xiao W, Tian Z. NK cell-based immunotherapy for cancer. *Semin Immunol* **31**, 37-54 (2017).
158. Daher M, Rezvani K. Next generation natural killer cells for cancer immunotherapy: the promise of genetic engineering. *Curr Opin Immunol* **51**, 146-153 (2018).
159. Geller MA, Miller JS. Use of allogeneic NK cells for cancer immunotherapy. *Immunotherapy* **3**, 1445-1459 (2011).
160. Gregoire C, *et al.* The trafficking of natural killer cells. *Immunol Rev* **220**, 169-182 (2007).
161. Sta Maria NS, Barnes SR, Jacobs RE. In vivo monitoring of natural killer cell trafficking during tumor immunotherapy. *Magn Reson Insights* **7**, 15-21 (2014).
162. Cheng M, Chen Y, Xiao W, Sun R, Tian Z. NK cell-based immunotherapy for malignant diseases. *Cell Mol Immunol* **10**, 230-252 (2013).
163. Ames E, Murphy WJ. Advantages and clinical applications of natural killer cells in cancer immunotherapy. *Cancer Immunol Immunother* **63**, 21-28 (2014).
164. Miller JS, *et al.* Successful adoptive transfer and in vivo expansion of human haploidentical NK cells in patients with cancer. *Blood* **105**, 3051-3057 (2005).
165. Klingemann H, Boissel L, Toneguzzo F. Natural Killer Cells for Immunotherapy - Advantages of the NK-92 Cell Line over Blood NK Cells. *Front Immunol* **7**, 91 (2016).
166. Bashor CJ, Horwitz AA, Peisajovich SG, Lim WA. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* **39**, 515-537 (2010).
167. Lim WA. Designing customized cell signalling circuits. *Nature reviews Molecular cell biology* **11**, 393-403 (2010).
168. Nielsen AA, Segall-Shapiro TH, Voigt CA. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr Opin Chem Biol* **17**, 878-892 (2013).
169. Khalil AS, *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012).
170. Maeder ML, *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* **31**, 294-301 (2008).
171. Chavez A, *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326-328 (2015).
172. Ai HW, Shaner NC, Cheng Z, Tsien RY, Campbell RE. Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **46**, 5904-5910 (2007).
173. Duportet X, *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res* **42**, 13440-13451 (2014).

174. Bintu L, *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720-724 (2016).
175. Chu VT, *et al.* Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes. *BMC Biotechnol* **16**, 4 (2016).
176. Xu Q, Schlabach MR, Hannon GJ, Elledge SJ. Design of 240,000 orthogonal 25mer DNA barcode probes. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2289-2294 (2009).
177. Konermann S, *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).
178. Goodwin EC, Rottman FM. The 3'-flanking sequence of the bovine growth hormone gene contains novel elements required for efficient and accurate polyadenylation. *J Biol Chem* **267**, 16330-16334 (1992).
179. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS one* **3**, e3647 (2008).
180. Xu D, Farmer A, Collett G, Grishin NV, Chook YM. Sequence and structural analyses of nuclear export signals in the NESdb database. *Mol Biol Cell* **23**, 3677-3693 (2012).
181. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).
182. Gibson DG, Smith HO, Hutchison CA, 3rd, Venter JC, Merryman C. Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* **7**, 901-903 (2010).
183. Stubbs L, Sun Y, Caetano-Anolles D. Function and Evolution of C2H2 Zinc Finger Arrays. *Subcell Biochem* **52**, 75-94 (2011).
184. Jones KA, Kadonaga JT, Luciw PA, Tjian R. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232**, 755-759 (1986).
185. Ptashne M. How eukaryotic transcriptional activators work. *Nature* **335**, 683-689 (1988).
186. Hansen J, Mailand E, Swaminathan KK, Schreiber J, Angelici B, Benenson Y. Transplantation of prokaryotic two-component signaling pathways into mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15705-15710 (2014).
187. Beal J, *et al.* Model-driven engineering of gene expression from RNA replicons. *ACS synthetic biology* **4**, 48-56 (2015).
188. Angelici B, Mailand E, Haefliger B, Benenson Y. Synthetic Biology Platform for Sensing and Integrating Endogenous Transcriptional Inputs in Mammalian Cells. *Cell Rep* **16**, 2525-2537 (2016).
189. Gill G, Ptashne M. Negative effect of the transcriptional activator GAL4. *Nature* **334**, 721-724 (1988).
190. Natesan S, Rivera VM, Molinari E, Gilman M. Transcriptional squelching re-examined. *Nature* **390**, 349-350 (1997).

191. Alon U. *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC (2006).
192. Swirnoff AH, Milbrandt J. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol Cell Biol* **15**, 2275-2287 (1995).
193. Elrod-Erickson M, Benson TE, Pabo CO. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-464 (1998).
194. Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809-817 (1991).
195. Beerli RR, Segal DJ, Dreier B, Barbas CF, 3rd. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14628-14633 (1998).
196. Rivera VM, *et al.* A humanized system for pharmacologic control of gene expression. *Nature medicine* **2**, 1028-1032 (1996).
197. Bashor CJ, *et al.* Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* **364**, 593-597 (2019).
198. Muldoon J. COMET v1.0.1. *Zenodo*, <http://dx.doi.org/10.5281/zenodo.3533966> (2019).
199. Edelstein HI, *et al.* Elucidation And Refinement Of Synthetic Receptor Mechanisms. *Synthetic Biology*, (2020).
200. Muldoon JJ, Donahue PS, Dolberg TB, Leonard JN. Building with intent: technologies and principles for engineering mammalian cell-based therapies to sense and respond. *Curr Opin Biomed Eng* **4**, 127–133 (2017).
201. Guedan S, Ruella M, June CH. Emerging Cellular Therapies for Cancer. *Annu Rev Immunol* **37**, 145-171 (2019).
202. Roybal KT, Lim WA. Synthetic Immunology: Hacking Immune Cells to Expand Their Therapeutic Capabilities. *Annu Rev Immunol* **35**, 229-253 (2017).
203. Kojima R, Aubel D, Fussenegger M. Building sophisticated sensors of extracellular cues that enable mammalian cells to work as "doctors" in the body. *Cell Mol Life Sci*, (2020).
204. Schwarz KA, Leonard JN. Engineering cell-based therapies to interface robustly with host physiology. *Adv Drug Deliv Rev* **105**, 55-65 (2016).
205. Chen LC, Chen YY. Outsmarting and outmuscling cancer cells with synthetic and systems immunology. *Curr Opin Biotechnol* **60**, 111-118 (2019).
206. Ishizuka S, Lai CY, Otsu M, Nakauchi H, Nagamune T, Kawahara M. Designing Motif-Engineered Receptors To Elucidate Signaling Molecules Important for Proliferation of Hematopoietic Stem Cells. *ACS Synth Biol* **7**, 1709-1714 (2018).
207. Engelowski E, *et al.* Synthetic cytokine receptors transmit biological signals using artificial ligands. *Nat Commun* **9**, 2034 (2018).

208. Qudrat A, Truong K. Engineering Synthetic Proteins to Generate Ca(2+) Signals in Mammalian Cells. *ACS Synth Biol* **6**, 582-590 (2017).
209. Conklin BR, *et al.* Engineering GPCR signaling pathways with RASSLs. *Nat Methods* **5**, 673-678 (2008).
210. Kojima R, Scheller L, Fussenegger M. Nonimmune cells equipped with T-cell-receptor-like signaling for cancer cell ablation. *Nat Chem Biol* **14**, 42-49 (2018).
211. Qudrat A, Mosabbir AA, Truong K. Engineered Proteins Program Mammalian Cells to Target Inflammatory Disease Sites. *Cell Chem Biol* **24**, 703-711 e702 (2017).
212. Qudrat A, Truong K. Autonomous Cell Migration to CSF1 Sources via a Synthetic Protein-Based System. *ACS Synth Biol* **6**, 1563-1571 (2017).
213. Chung HK, *et al.* A compact synthetic pathway rewires cancer signaling to therapeutic effector release. *Science* **364**, (2019).
214. Barnea G, *et al.* The genetic design of signaling cascades to record receptor activation. *Proc Natl Acad Sci U S A* **105**, 64-69 (2008).
215. Maze A, Benenson Y. Artificial signaling in mammalian cells enabled by prokaryotic two-component system. *Nat Chem Biol* **16**, 179-187 (2020).
216. Krawczyk K, Scheller L, Kim H, Fussenegger M. Rewiring of endogenous signaling pathways to genomic targets for therapeutic cell reprogramming. *Nature Communications* **11**, (2020).
217. Roybal KT, *et al.* Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits. *Cell* **164**, 770-779 (2016).
218. Yang Z, Yu Z, Cai Y, Du R, Cai L. Engineering of an enhanced synthetic Notch receptor by reducing ligand-independent activation. *Communications Biology* **3**, 116 (2020).
219. Brenner M, Cho JH, Wong WW. Synthetic biology: Sensing with modular receptors. *Nat Chem Biol* **13**, 131-132 (2017).
220. Tang JC, *et al.* A nanobody-based system using fluorescent proteins as scaffolds for cell-specific gene manipulation. *Cell* **154**, 928-939 (2013).
221. Edelstein HI, *et al.* Source Data for "Elucidation and refinement of synthetic receptor mechanisms". *Zenodo*, (2020).
222. Xia Z, Liu Y. Reliable and global measurement of fluorescence resonance energy transfer using fluorescence microscopes. *Biophys J* **81**, 2395-2402 (2001).
223. Edelstein HI, Kang AK. MESA FRET v1.1. *Zenodo*, (2020).
224. Zal T, Gascoigne NR. Photobleaching-corrected FRET efficiency imaging of live cells. *Biophys J* **86**, 3923-3939 (2004).
225. Liang J, Choi J, Clardy J. Refined structure of the FKBP12-rapamycin-FRB ternary complex at 2.2 Å resolution. *Acta Crystallogr D Biol Crystallogr* **55**, 736-744 (1999).

226. Murase K, Hirano Y, Sun TP, Hakoshima T. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **456**, 459-463 (2008).
227. Yin P, *et al.* Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat Struct Mol Biol* **16**, 1230-1236 (2009).
228. Fuh G, *et al.* Structure-function studies of two synthetic anti-vascular endothelial growth factor Fabs and comparison with the Avastin Fab. *J Biol Chem* **281**, 6625-6631 (2006).
229. Dougherty W, Cary SM, Parks TD. Molecular genetic analysis of a plant virus polyprotein cleavage sequence: a model. *Virology* **171**, 356-364 (1989).
230. Kapust RB, Tozser J, Copeland TD, Waugh DS. The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* **294**, 949-955 (2002).
231. Nunn CM, *et al.* Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. *J Mol Biol* **350**, 145-155 (2005).
232. Alabanza L, *et al.* Function of Novel Anti-CD19 Chimeric Antigen Receptors with Human Variable Regions Is Affected by Hinge and Transmembrane Domains. *Mol Ther* **25**, 2452-2465 (2017).
233. Bridgeman JS, Hawkins RE, Bagley S, Blaylock M, Holland M, Gilham DE. The optimal antigen response of chimeric antigen receptors harboring the CD3zeta transmembrane domain is dependent upon incorporation of the receptor into the endogenous TCR/CD3 complex. *J Immunol* **184**, 6938-6949 (2010).
234. Biggs MJ, Milone MC, Santos LC, Gondarenko A, Wind SJ. High-resolution imaging of the immunological synapse and T-cell receptor microclustering through microfabricated substrates. *J R Soc Interface* **8**, 1462-1471 (2011).
235. Yokosuka T, Saito T. Dynamic regulation of T-cell costimulation through TCR-CD28 microclusters. *Immunol Rev* **229**, 27-40 (2009).
236. Westerfield JM, Barrera FN. Membrane receptor activation mechanisms and transmembrane peptide tools to elucidate them. *J Biol Chem* **295**, 1792-1814 (2020).
237. Chen LI, Webster MK, Meyer AN, Donoghue DJ. Transmembrane domain sequence requirements for activation of the p185c-neu receptor tyrosine kinase. *J Cell Biol* **137**, 619-631 (1997).
238. Lomize AL, Pogozheva ID. TMDOCK: An Energy-Based Method for Modeling alpha-Helical Dimers in Membranes. *J Mol Biol* **429**, 390-398 (2017).
239. Doura AK, Fleming KG. Complex interactions at the helix-helix interface stabilize the glycoporphin A transmembrane dimer. *J Mol Biol* **343**, 1487-1497 (2004).
240. Finger C, Escher C, Schneider D. The single transmembrane domains of human receptor tyrosine kinases encode self-interactions. *Sci Signal* **2**, ra56 (2009).
241. Banning C, *et al.* A flow cytometry-based FRET assay to identify and analyse protein-protein interactions in living cells. *PLoS One* **5**, e9344 (2010).
242. Goedhart J, van Weeren L, Hink MA, Vischer NO, Jalink K, Gadella TW, Jr. Bright cyan fluorescent protein variants identified by fluorescence lifetime screening. *Nat Methods* **7**, 137-139 (2010).

243. Day RN, Davidson MW. The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* **38**, 2887-2921 (2009).
244. Koushik SV, Chen H, Thaler C, Puhl HL, 3rd, Vogel SS. Cerulean, Venus, and VenusY67C FRET reference standards. *Biophys J* **91**, L99-L101 (2006).
245. Berney C, Danuser G. FRET or no FRET: a quantitative comparison. *Biophys J* **84**, 3992-4010 (2003).
246. Bell CA, Tynan JA, Hart KC, Meyer AN, Robertson SC, Donoghue DJ. Rotational coupling of the transmembrane and kinase domains of the Neu receptor tyrosine kinase. *Mol Biol Cell* **11**, 3589-3599 (2000).
247. Kirchhofer A, *et al.* Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* **17**, 133-138 (2010).
248. Gao Y, Xiong X, Wong S, Charles EJ, Lim WA, Qi LS. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* **13**, 1043-1049 (2016).
249. Algar WR, Hildebrandt N, Vogel SS, Medintz IL. FRET as a biomolecular research tool - understanding its potential while avoiding pitfalls. *Nat Methods* **16**, 815-829 (2019).
250. Chan FK, Chun HJ, Zheng L, Siegel RM, Bui KL, Lenardo MJ. A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science* **288**, 2351-2354 (2000).
251. Hochreiter B, Garcia AP, Schmid JA. Fluorescent proteins as genetically encoded FRET biosensors in life sciences. *Sensors (Basel)* **15**, 26281-26314 (2015).
252. Koh MY, Powis G. Passing the baton: the HIF switch. *Trends Biochem Sci* **37**, 364-372 (2012).
253. Uchida T, *et al.* Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. *J Biol Chem* **279**, 14871-14878 (2004).
254. Gordan JD, Simon MC. Hypoxia-inducible factors: central regulators of the tumor phenotype. *Curr Opin Genet Dev* **17**, 71-77 (2007).
255. Holmquist-Mengelbier L, *et al.* Recruitment of HIF-1alpha and HIF-2alpha to common target genes is differentially regulated in neuroblastoma: HIF-2alpha promotes an aggressive phenotype. *Cancer Cell* **10**, 413-423 (2006).
256. Jochmanová I, Zelinka T, Widimský J, Jr., Pacak K. HIF signaling pathway in pheochromocytoma and other neuroendocrine tumors. *Physiol Res* **63**, S251-262 (2014).
257. Löfstedt T, *et al.* Hypoxia inducible factor-2alpha in cancer. *Cell Cycle* **6**, 919-926 (2007).
258. Semenza GL. HIF-1 mediates the Warburg effect in clear cell renal carcinoma. *J Bioenerg Biomembr* **39**, 231-234 (2007).
259. Tiburcio PD, Choi H, Huang LE. Complex role of HIF in cancer: the known, the unknown, and the unexpected. *Hypoxia (Auckl)* **2**, 59-70 (2014).
260. Jun JC, Rathore A, Younas H, Gilkes D, Polotsky VY. Hypoxia-Inducible Factors and Cancer. *Curr Sleep Med Rep* **3**, 1-10 (2017).

261. Iommarini L, Porcelli AM, Gasparre G, Kurelac I. Non-Canonical Mechanisms Regulating Hypoxia-Inducible Factor 1 Alpha in Cancer. *Front Oncol* **7**, 286 (2017).
262. Muldoon JJ, *et al.* Model-guided design of mammalian genetic programs. *bioRxiv*, 2020.2009.2030.320853 (2020).
263. Gaidukov L, *et al.* A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res* **46**, 4072-4086 (2018).
264. Fitzgerald M, Livingston M, Gibbs C, Deans TL. Rosa26 docking sites for investigating genetic circuit silencing in stem cells. *Synthetic Biology*, (2020).
265. Hu CJ, Sataur A, Wang L, Chen H, Simon MC. The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1alpha and HIF-2alpha. *Mol Biol Cell* **18**, 4528-4542 (2007).
266. Subach OM, Cranfill PJ, Davidson MW, Verkhusha VV. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS One* **6**, e28674 (2011).
267. Shaner NC, *et al.* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat Methods* **10**, 407-409 (2013).
268. Bajar BT, *et al.* Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci Rep* **6**, 20889 (2016).
269. Shemetov AA, Oliinyk OS, Verkhusha VV. How to Increase Brightness of Near-Infrared Fluorescent Proteins in Mammalian Cells. *Cell Chem Biol* **24**, 758-766.e753 (2017).
270. Oliinyk OS, Shemetov AA, Pletnev S, Shcherbakova DM, Verkhusha VV. Smallest near-infrared fluorescent protein evolved from cyanobacteriochrome as versatile tag for spectral multiplexing. *Nat Commun* **10**, 279 (2019).
271. Shcherbakova DM, Cox Cammer N, Huisman TM, Verkhusha VV, Hodgson L. Direct multiplex imaging and optogenetics of Rho GTPases enabled by near-infrared FRET. *Nat Chem Biol* **14**, 591-600 (2018).
272. Chu VT, *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-548 (2015).
273. Place TL, Domann FE, Case AJ. Limitations of oxygen delivery to cells in culture: An underappreciated problem in basic and translational research. *Free Radic Biol Med* **113**, 311-322 (2017).
274. Wenger RH, Kurtcuoglu V, Scholz CC, Marti HH, Hoogewijs D. Frequently asked questions in hypoxia research. *Hypoxia (Auckl)* **3**, 35-43 (2015).
275. Erapanedi R, Belousov VV, Schäfers M, Kiefer F. A novel family of fluorescent hypoxia sensors reveal strong heterogeneity in tumor hypoxia at the cellular level. *Embo j* **35**, 102-113 (2016).
276. Misra T, *et al.* A genetically encoded biosensor for visualising hypoxia responses in vivo. *Biol Open* **6**, 296-304 (2017).
277. Yuan Y, Hilliard G, Ferguson T, Millhorn DE. Cobalt inhibits the interaction between hypoxia-

- inducible factor- α and von Hippel-Lindau protein by direct binding to hypoxia-inducible factor- α . *J Biol Chem* **278**, 15911-15916 (2003).
278. Iyer NV, Leung SW, Semenza GL. The human hypoxia-inducible factor 1 α gene: HIF1A structure and evolutionary conservation. *Genomics* **52**, 159-165 (1998).
279. Hon WC, *et al.* Structural basis for the recognition of hydroxyproline in HIF-1 α by pVHL. *Nature* **417**, 975-978 (2002).
280. Iglesias P, Penas C, Barral-Cagiao L, Pazos E, Costoya JA. A Bio-inspired Hypoxia Sensor using HIF1 α -Oxygen-Dependent Degradation Domain. *Sci Rep* **9**, 7117 (2019).
281. Grégoire C, *et al.* The trafficking of natural killer cells. *Immunological Reviews* **220**, 169-182 (2007).
282. Gong JH, Maki G, Klingemann HG. Characterization of a human cell line (NK-92) with phenotypical and functional characteristics of activated natural killer cells. *Leukemia* **8**, 652-658 (1994).
283. Zhang C, *et al.* Chimeric Antigen Receptor-Engineered NK-92 Cells: An Off-the-Shelf Cellular Therapeutic for Targeted Elimination of Cancer Cells and Induction of Protective Antitumor Immunity. *Front Immunol* **8**, 533 (2017).
284. Jochems C, *et al.* An NK cell line (haNK) expressing high levels of granzyme and engineered to express the high affinity CD16 allele. *Oncotarget* **7**, 86359-86373 (2016).
285. Yeh HW, Karmach O, Ji A, Carter D, Martins-Green MM, Ai HW. Red-shifted luciferase-luciferin pairs for enhanced bioluminescence imaging. *Nat Methods* **14**, 971-974 (2017).
286. Yeh HW, *et al.* ATP-Independent Bioluminescent Reporter Variants To Improve in Vivo Imaging. *ACS Chem Biol* **14**, 959-965 (2019).
287. Arakaki LS, Ciesielski WA, Thackray BD, Feigl EO, Schenkman KA. Simultaneous optical spectroscopic measurement of hemoglobin and myoglobin saturations and cytochrome aa3 oxidation in vivo. *Appl Spectrosc* **64**, 973-979 (2010).
288. Sandell JL, Zhu TC. A review of in-vivo optical properties of human tissues and its impact on PDT. *J Biophotonics* **4**, 773-787 (2011).
289. Kuchimaru T, *et al.* A luciferin analogue generating near-infrared bioluminescence achieves highly sensitive deep-tissue imaging. *Nat Commun* **7**, 11856 (2016).
290. Tam YK, Maki G, Miyagawa B, Hennemann B, Tonn T, Klingemann HG. Characterization of genetically altered, interleukin 2-independent natural killer cell lines suitable for adoptive cellular immunotherapy. *Hum Gene Ther* **10**, 1359-1373 (1999).
291. Nagashima S, *et al.* Stable transduction of the interleukin-2 gene into human natural killer cell lines and their phenotypic and functional characterization in vitro and in vivo. *Blood* **91**, 3850-3861 (1998).
292. Konstantinidis KV, Alici E, Aints A, Christensson B, Ljunggren HG, Dilber MS. Targeting IL-2 to the endoplasmic reticulum confines autocrine growth stimulation to NK-92 cells. *Exp Hematol* **33**, 159-164 (2005).
293. Jounaidi Y, Cotten JF, Miller KW, Forman SA. Tethering IL2 to Its Receptor IL2R β Enhances Antitumor Activity and Expansion of Natural Killer NK92 Cells. *Cancer Res* **77**, 5938-5951 (2017).

294. Iwano S, *et al.* Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* **359**, 935-939 (2018).
295. Clackson T, *et al.* Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc Natl Acad Sci U S A* **95**, 10437-10442 (1998).
296. Kapust RB, *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993–1000 (2001).

Appendix 1. Disposition of reagents

A list of plasmids generated during this thesis work (approximately 1,400) will be deposited in our Leonard Lab Inventory electronic archive. Published plasmids have descriptions and maps published with the respective work and will be made available on Addgene. All published and otherwise useful plasmids will also be banked in DNA and bacterial stocks in the Leonard Lab Inventory.

A list of mammalian cell lines generated during this thesis work (approximately 300) will be deposited in our Leonard Lab Inventory electronic archive. All published and otherwise useful mammalian cell lines will also be banked in duplicate in the Leonard Lab Inventory.

Appendix 2. Supplementary information for Chapter 2

A version of this appendix was previously published as supplemental information for:

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

It has been reproduced here for completeness.

The work in Supplementary Figures A2.3 and A2.8 was done by Joseph Muldoon.

The work in Supplementary Figure A2.5 was done by Joseph Draut.

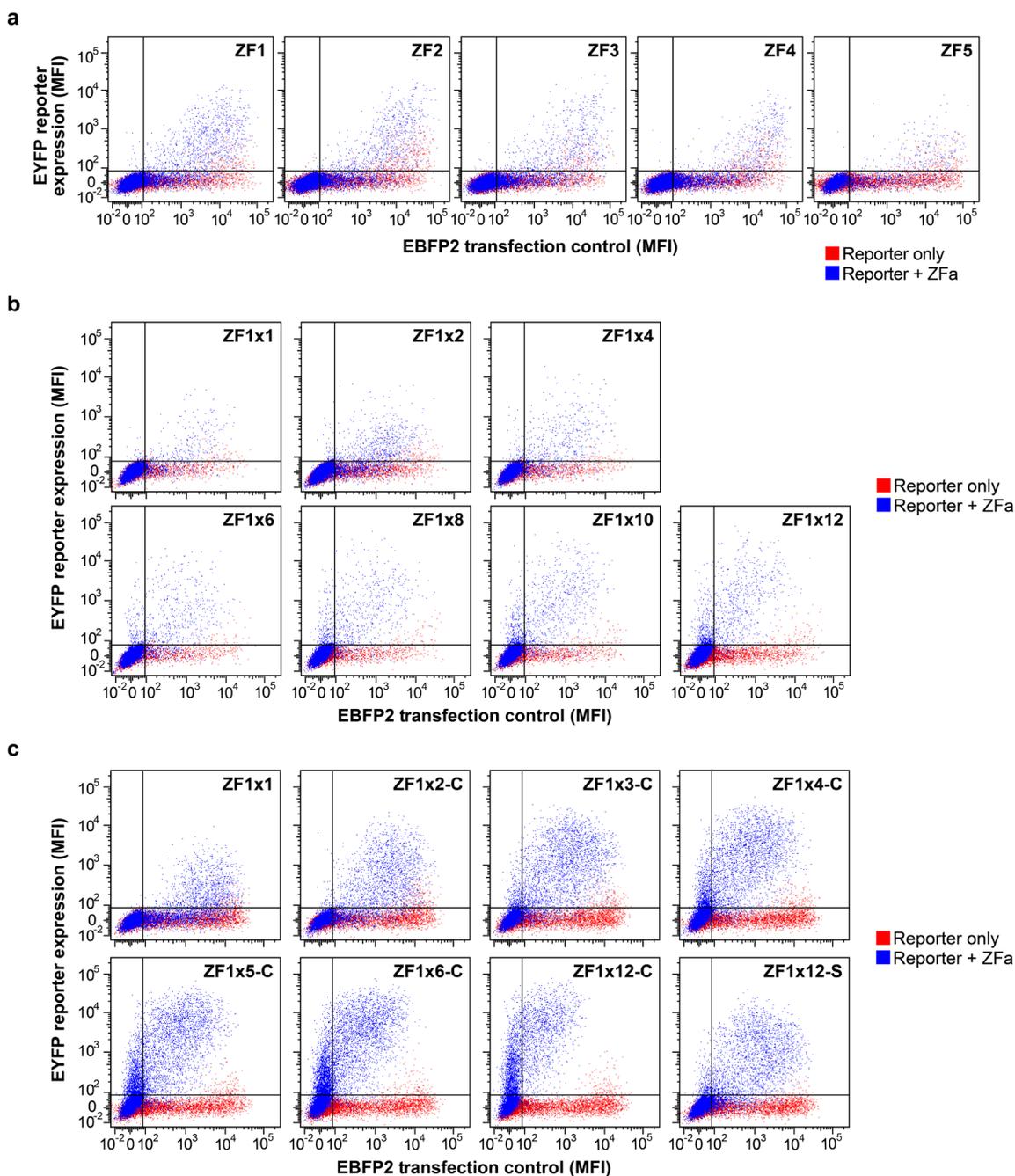
The work in Supplementary Figures A2.5, A2.9, and A2.10 was done by Hailey Edelstein.

Contents:

Supplementary Figures A2.1– A2.16

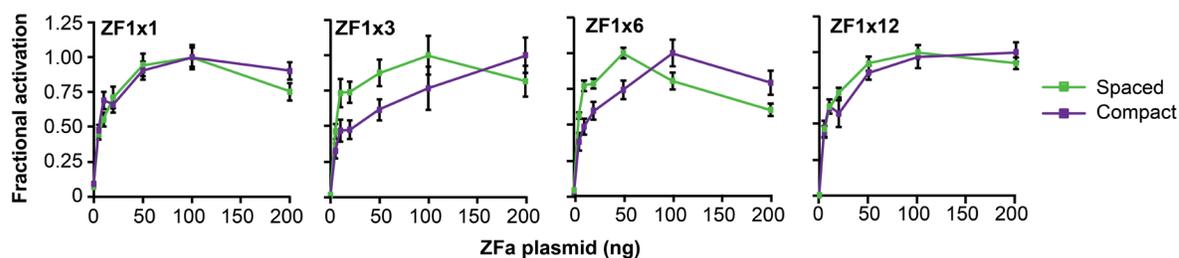
Supplementary Tables A2.1– A2.8

Supplementary Note A2.1

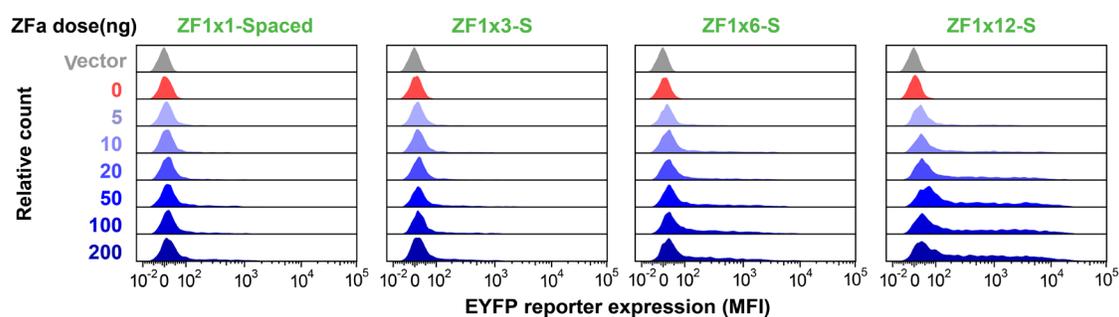


Supplementary Fig. A2.1 Effects of the choice of ZF and the number and spacing of ZF binding sites. (a) Representative flow cytometry plots from Fig. 2.1b. The joint distribution shows that inducible EYFP expression is higher in more highly transfected cells (i.e., those with more EBFP2 expression). Constitutively expressed EBFP2 is the transfection control. (b) Representative flow cytometry plots from Fig. 2.1c. (c) Representative flow cytometry plots from Fig. 2.1e. Squelching is suggested in conditions with strong COMET promoters (i.e., higher EYFP expression), which exhibit lower EBFP2 fluorescence than do conditions with weaker COMET promoters.

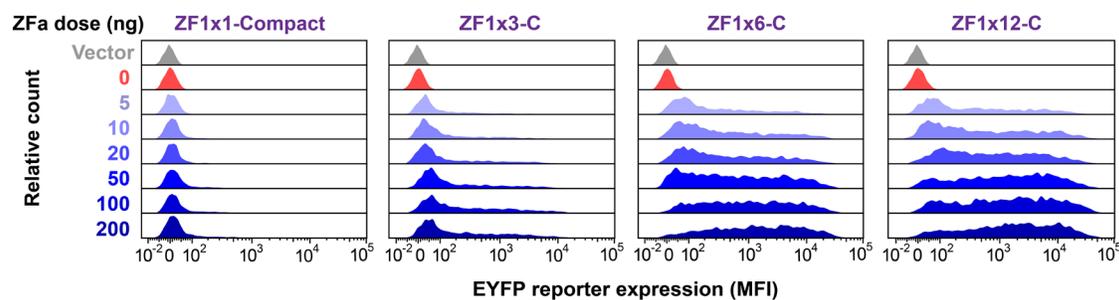
a



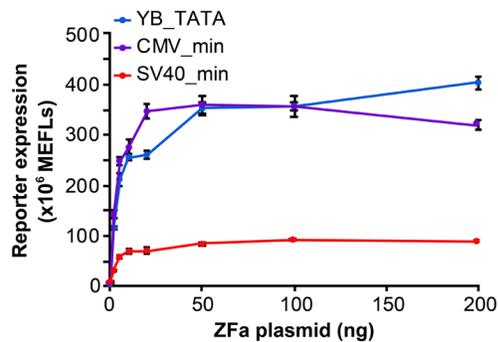
b



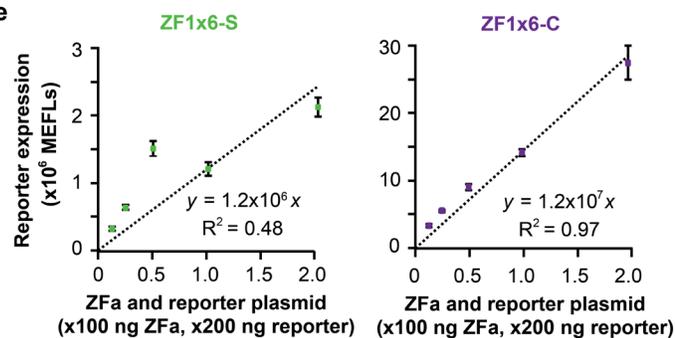
c



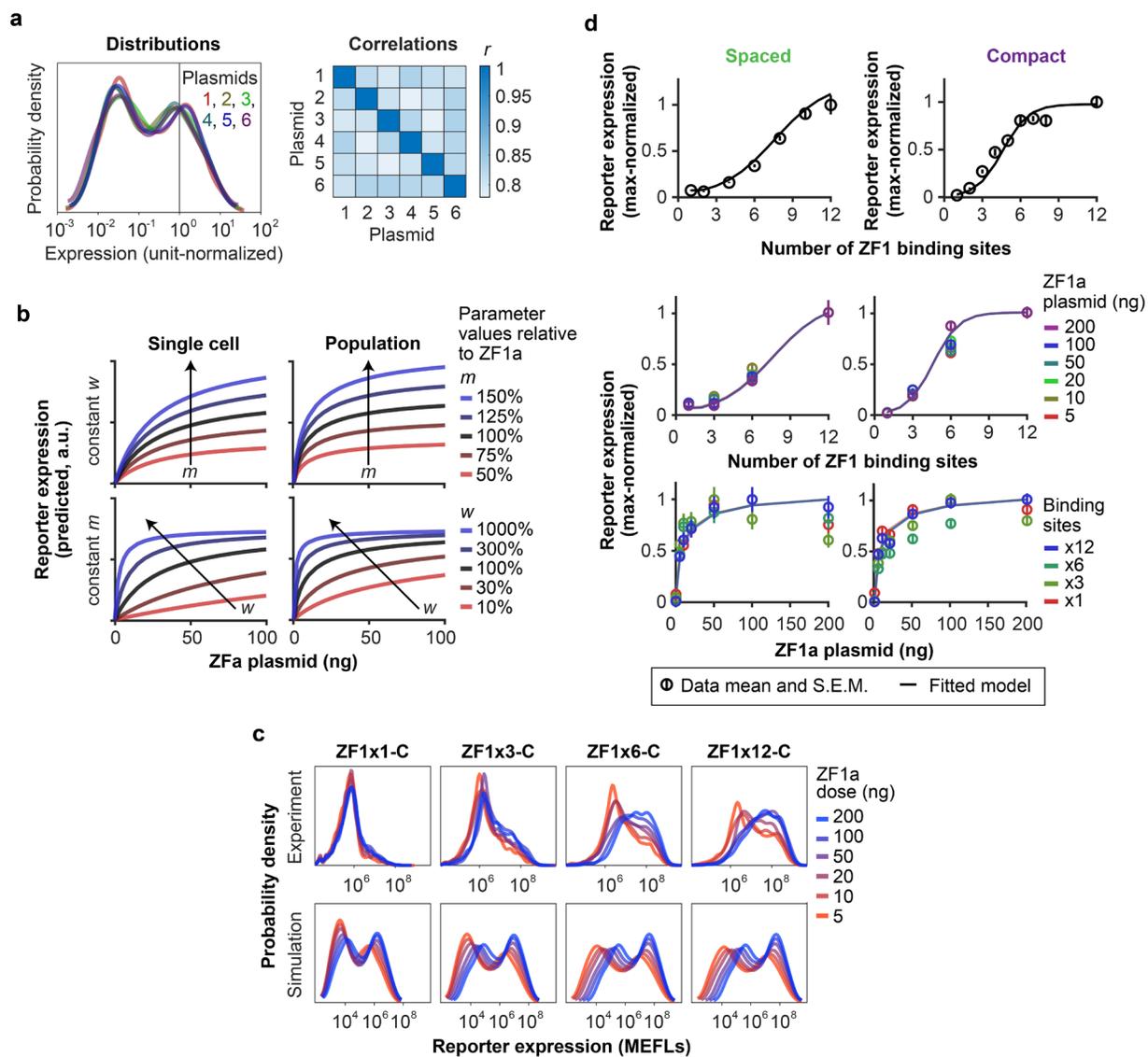
d



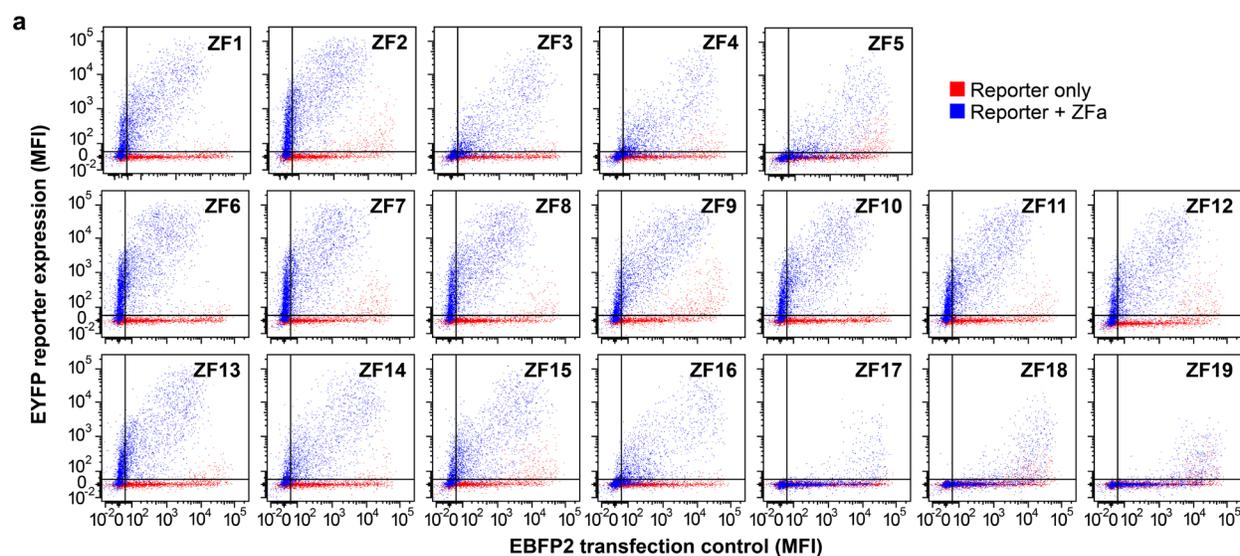
e



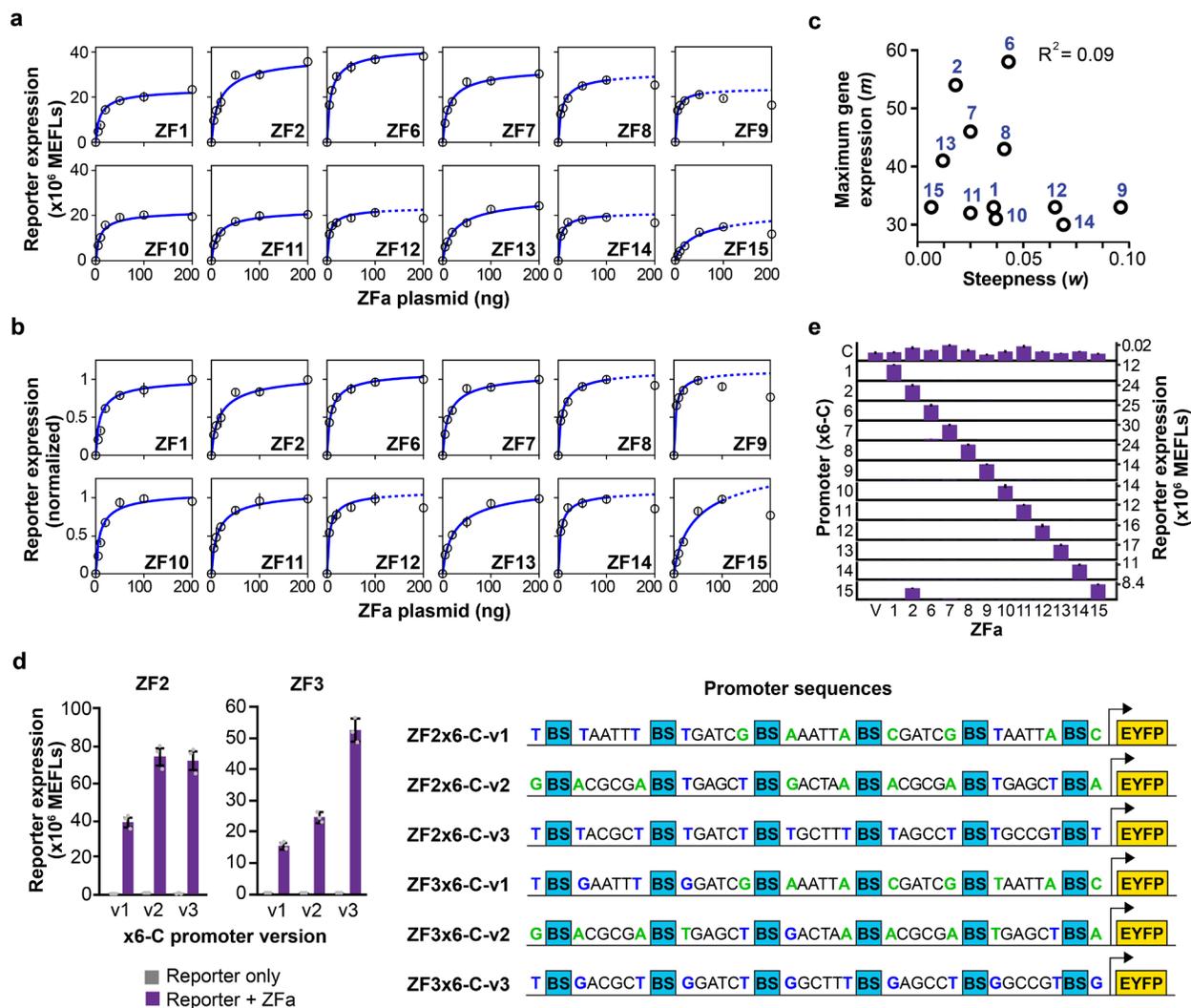
Supplementary Fig. A2.2 Differences between spaced and compact promoters. (a) Fractional activation for dose response profiles in **Fig. 2.2a**. Fractional activation was determined by dividing each data point by the maximum reporter expression induced by the ZFa on a given reporter. In several conditions, notably ZF1x6-S, excess ZFa (above 50 ng plasmid) resulted in a decrease in reporter expression. In this case, we hypothesize that unbound ZFa competes with bound ZFa for endogenous cofactors required for transcription. (b) Representative flow cytometry histograms for experiments with spaced reporters in **Fig. 2.2a**. Data were gated on single, transfected cells (**Supplementary Fig. A2.15**). Reporter expression increases with ZFa dose and number of binding sites. (c) Representative flow cytometry histograms from experiments with compact reporters in **Fig. 2.2a**. Data were gated on single, transfected cells. Reporter expression increases with ZFa dose and number of binding sites. Compared to the case of spaced promoters (panel b), cases with compact promoters exhibit a greater fraction of cells that are distinguishably ON, i.e., expressing more EYFP than cells without ZFa. (d) Investigating different minimal promoters with COMET. ZF1a dose responses were conducted using ZF1x6-C promoters with either the YB_TATA, CMV, or SV40 minimal promoters. The SV40 minimal promoter produced low levels of gene expression, while the YB_TATA minimal promoter conferred a maximal gene expression level similar to that of the CMV minimal promoter. Although the CMV minimal promoter was more responsive at lower levels of ZFa expression, this promoter also had higher leaky gene expression (in the absence of ZFa) than did the YB_TATA minimal promoter (quantified in **Supplementary Table A2.10** with fitted parameters). Thus, the YB_TATA minimal promoter resulted in higher fold inductions than did the CMV minimal promoter (approximately 220-fold for YB_TATA, as compared to 60-fold for CMV_min) without sacrificing maximal gene expression. (e) Investigating maximal inducible EYFP expression. Cells were transfected with ZF1a plasmid and with reporter plasmid containing a ZF1x6-S (*left*) or ZF1x6-C (*right*) promoter. The ZFa plasmid and reporter plasmid were maintained at a ratio of 1:2 (ZFa:reporter) as the doses were scaled. On the x-axis, a value of 1 denotes a condition with 100 ng of ZF1a plasmid and 200 ng of reporter plasmid. In previous experiments, reporter expression typically plateaued at the level indicated by plasmid doses corresponding to 1 on the x-axis and could not be increased by the addition of more ZFa plasmid (**Supplementary Fig. A2.2a**). However, doubling the amount of *both* ZFa plasmid and reporter plasmid led to twice the reporter expression, which indicates the amount of plasmid was the limiting factor in gene expression as opposed to a downstream step such as translation. For the compact promoter, reporter expression scaled linearly with dose of these components (at a fixed ratio) (one-tailed permutation test $p = 0.001$), but no strong linear correlation was observed for the spaced promoter ($p = 0.10$). Error bars depict S.E.M. Source data are provided as a **Source Data** file.



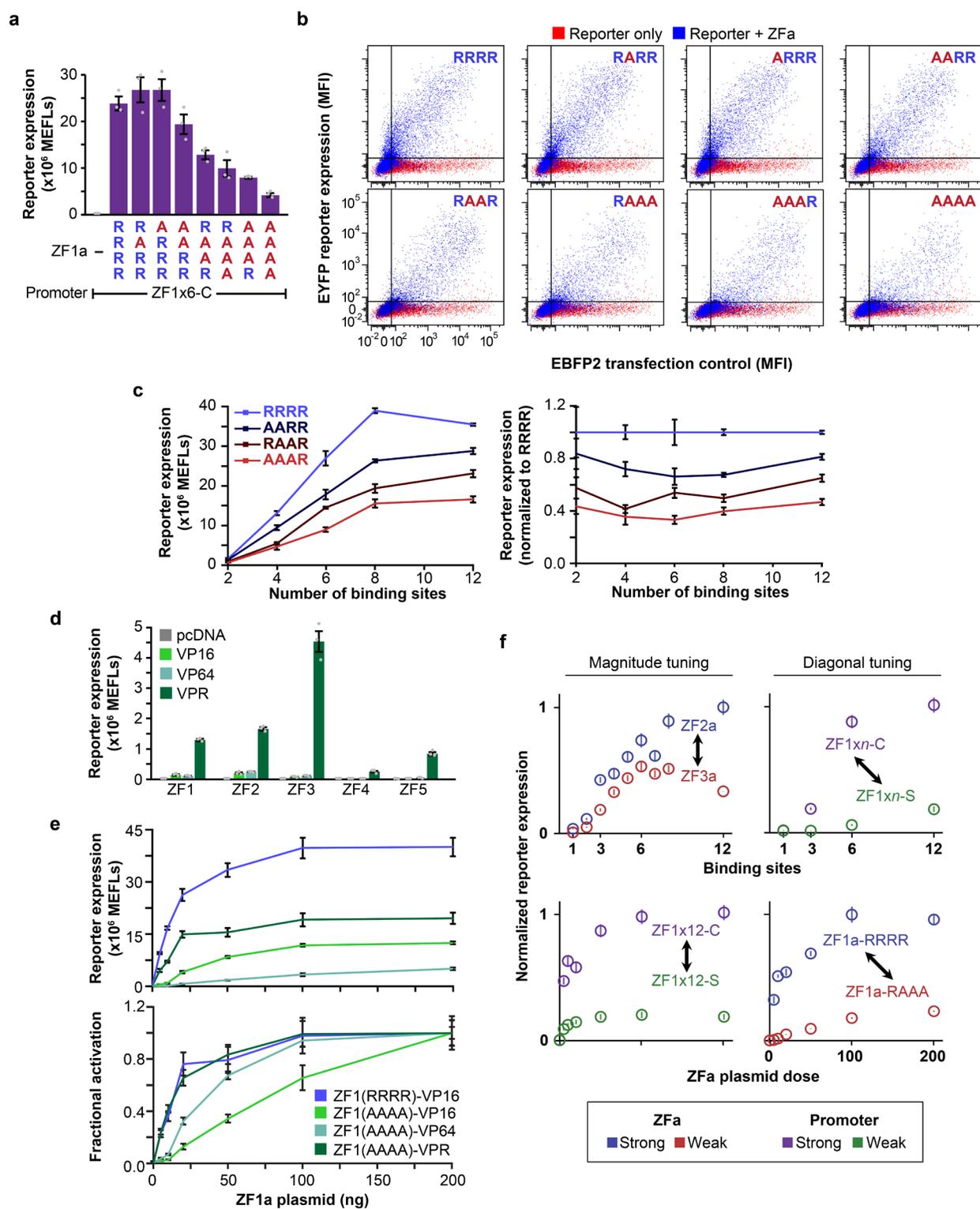
Supplementary Fig. A2.3 A model for ZFa-regulated gene expression. (a) A statistical model for cell heterogeneity is presented. Accounting for variation in gene expression enables more accurate explanations and predictions for genetic programs. An *in silico* population was generated for 200 cells and multiple plasmids using a previously described method¹¹⁷. Properties of simulated cell populations are summarized here. *Left*: marginal distributions from the model show intercellular variation in the relative expression of a constitutive gene matching expected distributions for single, transfected cells. Each distribution is normalized to a mean value of 1 arbitrary unit (a.u.) at the vertical line. *Right*: pairwise correlations capture intracellular variation, shown for the relative expression of six constitutively expressed genes on separate plasmids. (b) Advantages of using a population model. Simulated dose responses are shown for hypothetical ZFa with varying m or w parameters (rows). Outcomes are shown for a single cell and for the mean of a heterogeneous population (columns), where a single cell refers to a cell expressing the mean amount of each component. The comparison shows how, for the same model parameter values, the presence of population-level heterogeneity can lead to greater observed reporter expression. As a result, fitting models to experimental data (which include population-level heterogeneity) using a standard homogeneous approach (which corresponds to the single cell case) can lead to poor estimates of model parameters. Approaches that account for cell heterogeneity mitigate this problem. Y-axes are linearly scaled. (c) Experimental (flow cytometry) and simulated distributions of reporter expression for different ZF1a doses and numbers of binding sites. The model captures the observed bimodal log-Gaussian distributions, and the fact that at increasing ZFa doses, the probability density shifts from the lower mode to the upper mode. Simulated reporter expression is presented in internally consistent model-a.u. which are linearly scaled to align with experimental MEFLs. While MEFLs are absolute units, the magnitude of reporter induction varies between experiments, and this is why simulated distributions are scaled. (d) Effects of the number and spacing of binding sites and the ZFa dose. Cross-sections from experimental data (circles) and simulated landscapes in **Fig. 2.2a** were normalized to the maximum value within each cross-section. We observed that these normalized curves followed characteristic profiles. Lines show endpoint (long-time) mean reporter expression from dynamical population simulations. *Upper*: reporter expression varies sigmoidally with the number of binding sites (described in **Methods**). Spaced and compact promoters have distinct profiles. *Middle*: the scaled sigmoidal shape of the response to the number of binding sites holds across ZFa plasmid doses. *Lower*: the scaled concave shape of the response to ZFa dose holds across the number and spacing of binding sites. Error bars depict S.E.M. Source data are provided as a **Source Data** file.



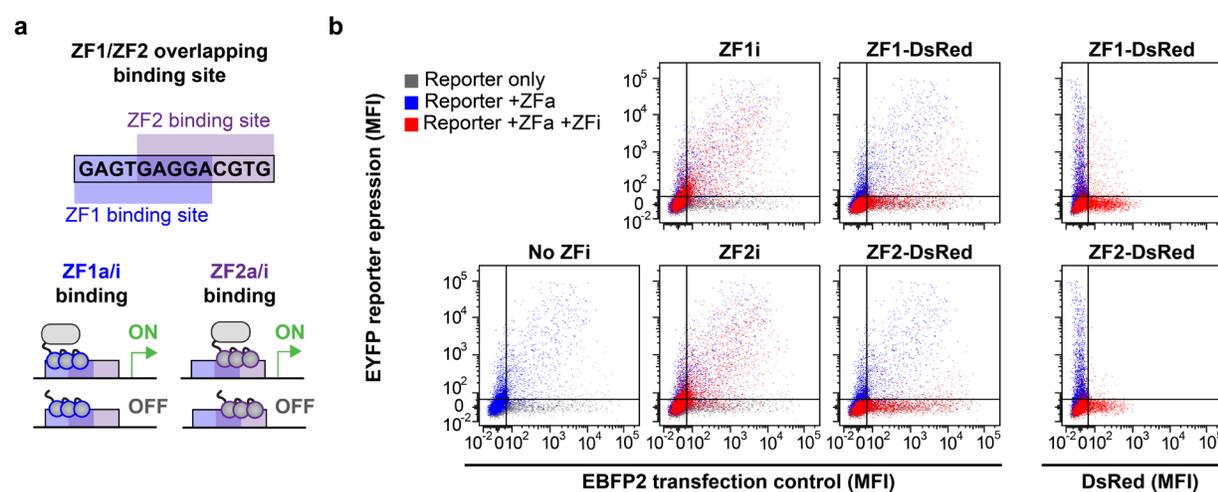
Supplementary Fig. A2.4 Characterization of the panel of ZFa. (a) Representative flow cytometry plots for Fig. 2.3a. We note several observations. (1) ZFa-induced EFYP increases with the transfection control protein (EBFP2). (2) Reporter-ZFa pairs that induce similar gene expression in the presence of a ZFa can vary in ZFa-independent gene expression (e.g., ZF6a and ZF7a). Thus, a different b (background) term in the COMET model is used for each promoter. (3) For many conditions, some cells that would be considered “not transfected” based on EBFP2 expression do express EYFP (e.g., the ZF2a case), indicating that ZFa are potent even when cells receive low amounts of plasmid. (4) Squelching is suggested by the observed decrease in EBFP2 expression between the “reporter only” and “reporter + ZFa” conditions for ZFa that induce high reporter expression.



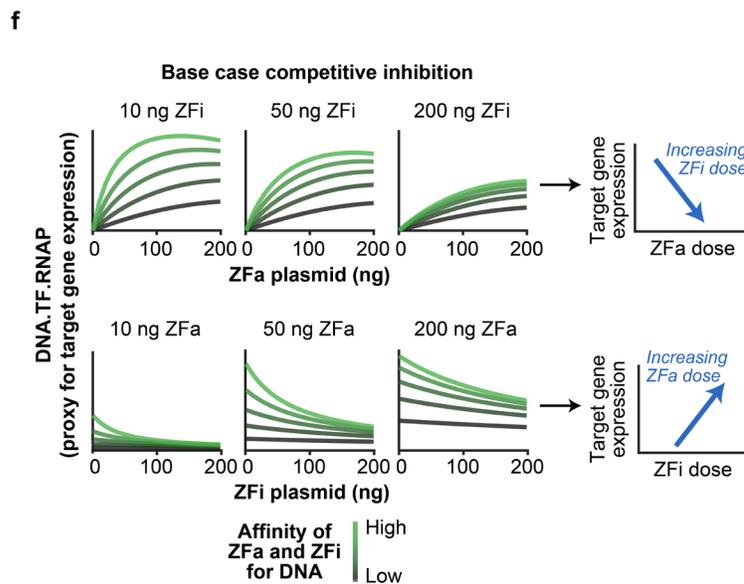
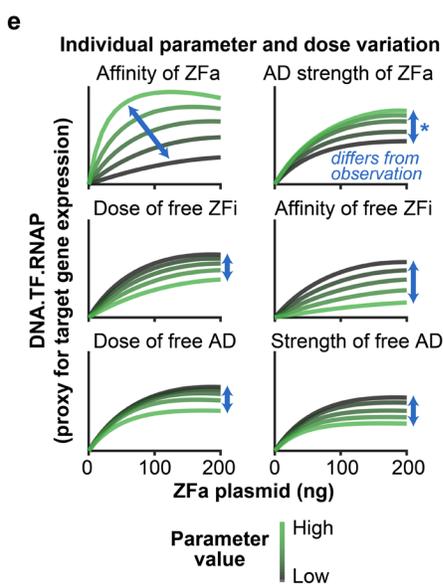
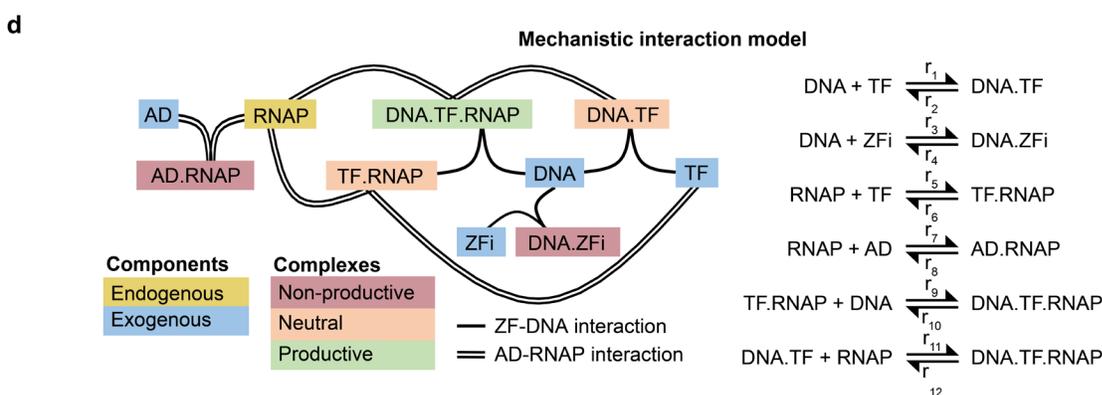
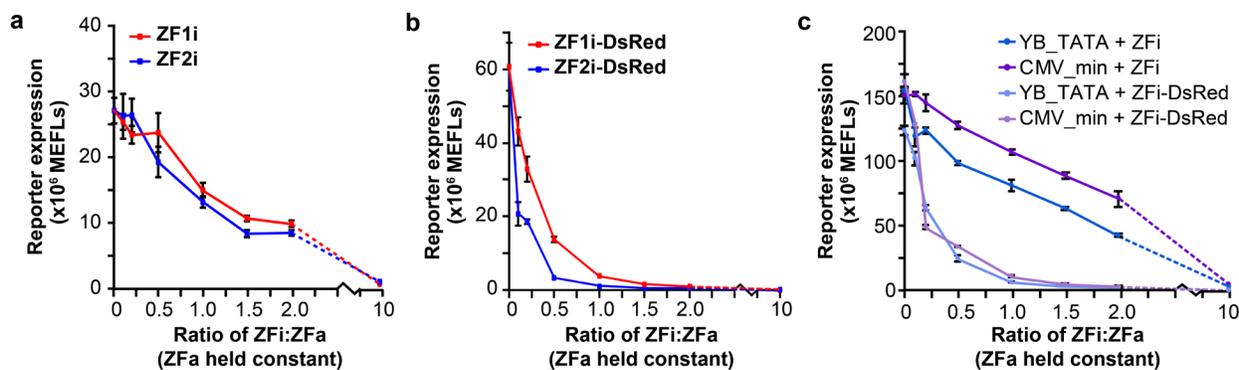
Supplementary Fig. A2.5 Characterization of the panel of ZFa. (a–c) Dose responses and model fits for the twelve strongest ZFa on x6-C promoters. Data are shown in MEFLs in (a) and normalized to the maximum of each reporter in (b). (c) Correlations in fitted parameters m and w for the twelve strongest ZFa on x6-C promoters. There is no significant correlation between m and w across the panel of ZFa (two-tailed permutation test $p = 0.34$). (d) ZFa induce transcription to varying extents depending on the flanking nucleotides. (Right) Comparison of compact promoters containing six binding sites with different nucleotides in the intervening base pairs between binding sites, including different flanking nucleotides. Blue denotes flanking nucleotides that were previously reported to confer strong ZF binding¹⁶⁹. Green denotes flanking nucleotides with previously unreported effects. “BS” indicates the ZF binding site. (Left) For ZF2a and ZF3a, reporter expression varied with changes to spacer sequences. Reporter v1 constructs were not used elsewhere in this work, reporter v2 constructs were used in **Fig. 2.3a** and **Fig. 2.3c**, and reporter v3 constructs were used in **Fig. 2.3b**. Gene expression was significantly affected by the version of the ZF2 reporter used (ANOVA $p < 0.001$) and the version of the ZF3 reporter used (ANOVA $p < 0.001$). (E) Bar graph representation of the data in **Fig. 2.3c**; each series is plotted on the range from null reporter expression (0 MEFLs, horizontal line) to the value indicated. Error bars depict S.E.M. Source data are provided as a **Source Data** file.



Supplementary Fig. A2.6 Properties of ZFa with ZF mutants and AD variants. (a) Effect of DNA affinity mutations on ZFa-induced reporter expression, using 100 ng of ZFa plasmid. ZF mutations affected reporter expression (ANOVA $p < 0.001$). (b) Representative flow cytometry plots from (a); the “reporter only” sample is the same across panels. (c) Effects of ZF mutations on reporter expression using promoters with varying numbers of binding sites. The WT ZF1a and three mutant ZF1a variants were each transfected along with compact promoters containing different numbers of binding sites. (*Left*) Reporter expression is presented in absolute units. Each ZFa reaches essentially maximum reporter expression with promoters containing 8–12 binding sites. Gene expression was affected by both the number of ZF binding sites (two-factor ANOVA $p < 0.001$) and the ZF mutations ($p < 0.001$). (*Right*) Reporter expression was normalized to the expression at each dose from the WT case. (d) VP16, VP64, and VPR activation domains (ADs) were fused to the panel of five ZF domains characterized in **Fig. 2.1b** and transfected into cells with a reporter plasmid containing a cognate x1 promoter driving an EYFP reporter. For each ZFa, VPR led to more reporter expression than VP16, and VP64 was similar to VP16. The ordering of reporter expression with VP16-ZF did not directly correspond to the ordering of reporter expression with VPR-ZF. Reporter expression was affected by both the choice of ZF (two-factor ANOVA $p < 0.001$) and the AD ($p < 0.001$), with an interaction between these two variables ($p < 0.001$). (e) Reporter expression data from **Fig. 2.4f** are shown with a linear x-axis, in absolute units (*left*) or normalized units (*right*). For each ZFa series, data were normalized to the maximum expression observed for the ZFa. The TFs vary in dose response profiles: WT ZF-VP16 and mutant ZF-VPR both reached maximal reporter expression at relatively low doses, mutant ZF-VP64 had a rare sigmoidal response, and mutant ZF-VP16 had a large linear range. The results show how the AD variant and ZF mutant can be chosen to customize response profiles. (f) Representative experimental datasets, corresponding to the summary schematic in **Fig. 2.4g**, were obtained from **Figs. 2.2a, 2.3b, and 2.4b**. Error bars depict S.E.M. Source data are provided as a **Source Data** file.

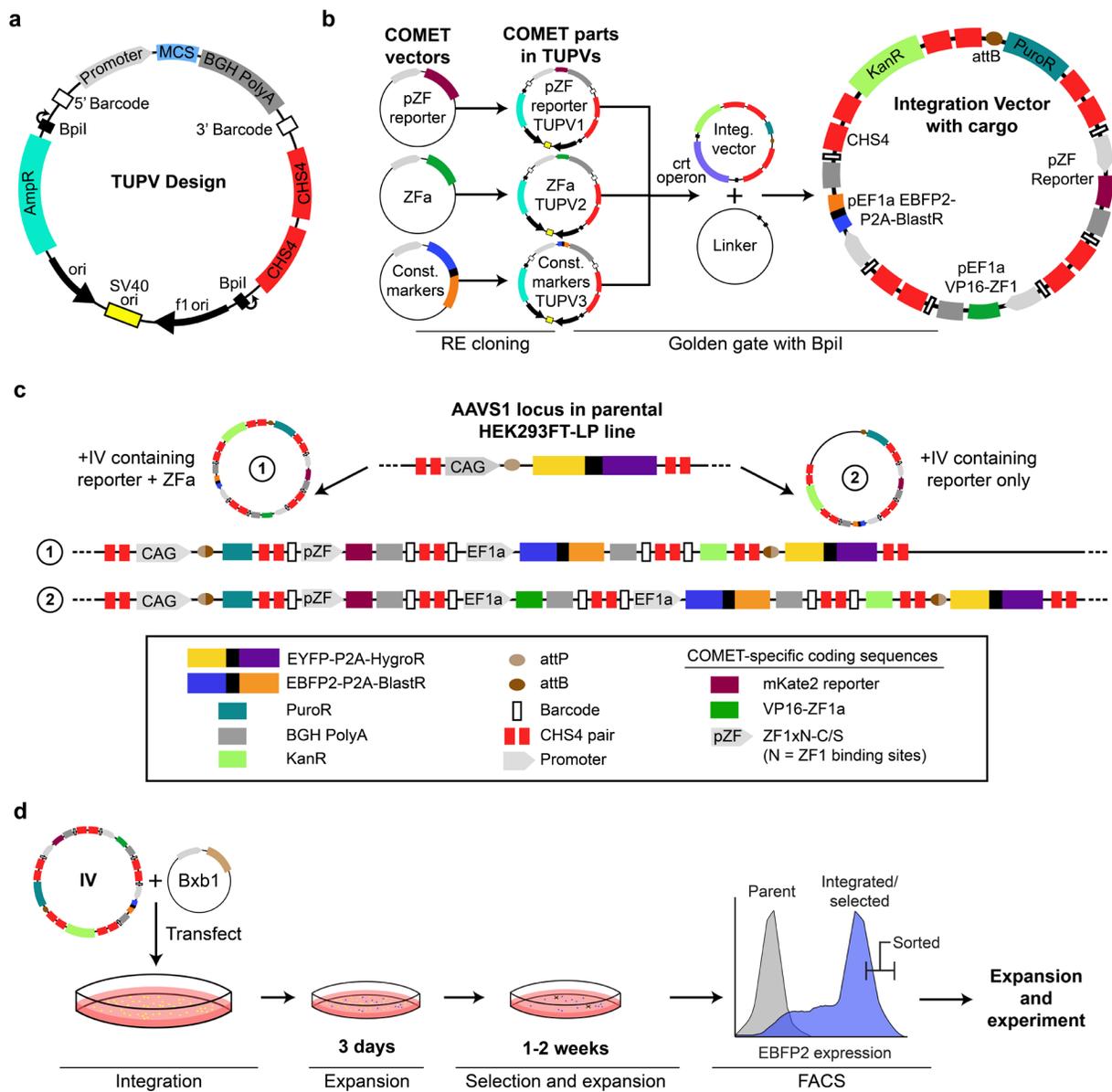


Supplementary Fig. A2.7 Investigation of ZFi-mediated and ZFi-DsRed-mediated inhibition. (a) The cartoon depicts a hybrid promoter regulated by a ZFa and a ZFi. Shown are potential states that any one overlapping binding site could take. The promoters that were evaluated experimentally contain multiple hybrid binding sites. The last five nucleotides of the ZF1 binding site are the same as the first five of the ZF2 binding site, and this property was used to construct a hybrid reporter by arranging six sites for ZF1 each separated by six base pairs, where the first four base pairs of this linker are the last four base pairs of the ZF2 site. (b) Representative flow cytometry plots from **Fig. 2.5b**.



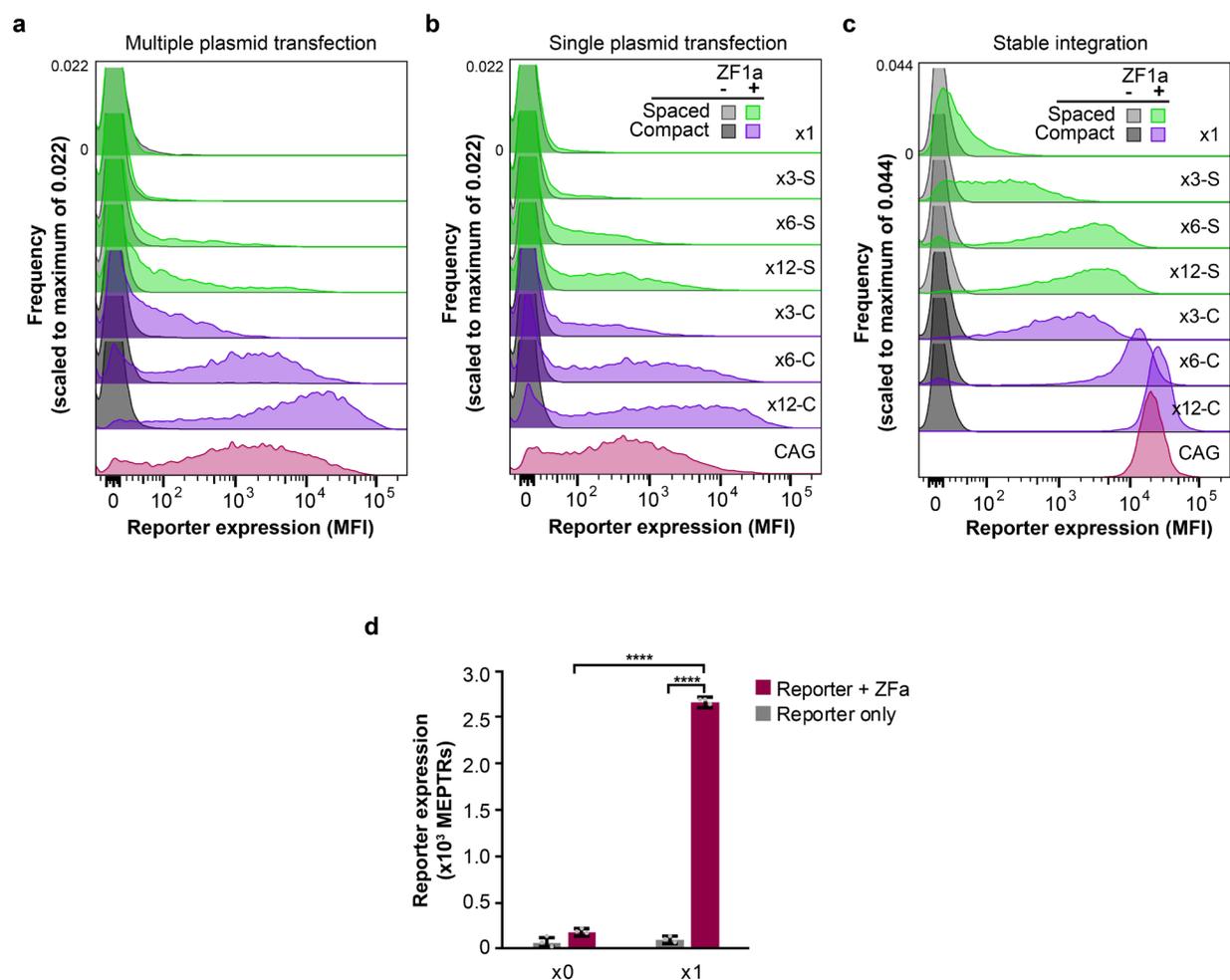
Supplementary Fig. A2.8 Further investigation of ZFi-mediated and ZFi-DsRed-mediated inhibition and the COMET mechanism.

(a, b) ZFi (a) and ZFi-DsRed (b) dose responses. In these experiments, the doses of ZFa plasmid and reporter plasmid were held constant while the dose of inhibitor plasmid was increased. Increasing the dose of ZFi led to decreased reporter expression (two-factor ANOVA $p < 0.001$), while the choice of ZF domain did not ($p = 0.35$). Increasing the dose of ZFi-DsRed also led to decreased reporter expression (two-factor ANOVA $p < 0.001$) and using the ZF2 domain resulted in more potent inhibition than using ZF3 ($p < 0.001$). (c) Dose responses of ZFi and ZFi-DsRed on the ZF1x6-C promoter with either the YB_TATA or CMV minimal promoters, as in (a) and (b). Both ZFi and ZFi-DsRed inhibited the YB_TATA and stronger CMV minimal promoters (two-factor ANOVAs $p < 0.001$). In both cases, the CMV minimal promoter was less inhibited than the YB_TATA promoter for a given dose of ZFi or ZFi-DsRed ($p < 0.001$), possibly due to the higher w value conferred on the ZFa response by the CMV minimal promoter. (d) A model for gene expression that uses more mechanistic detail than does the concise COMET model. The schematic on the left depicts reversible associations and dissociations (lines) between components (color-coded) as a network, and the corresponding biochemical reactions are listed on the right. The DNA variable represents a reporter gene with a promoter containing one binding site (x1). ZFi and free AD species can form non-productive complexes with DNA and RNAP, respectively. Here, the RNAP variable represents the ensemble of endogenous factors involved in transcription, and not specifically RNAPII (which does not directly physically interact with the AD). (e, f) Steady-state simulated values of DNA.TF.RNAP are a proxy for reporter expression here. Simulations were run for a single cell, i.e., homogeneously rather than for a heterogeneous population. (e) The abundance and properties of ZFa, ZFi, and free AD (not implemented experimentally) are expected to have the effects shown on reporter expression for a x1 promoter. The arrow in each panel indicates whether tuning affects the dose response trend vertically or diagonally. Although the simulated diagonal tuning of ZFa affinity agrees with experiments in **Fig. 2.4**, the effects of AD strength do not match those observed (denoted by an asterisk)—in that the x6-compact (*multi-site*) promoter experiments and associated transfer function model fits show diagonal tuning while the *single-site* promoter simulations show vertical tuning. This difference is consistent with the phenomenon of cooperativity through TF-mediated RNAPII recruitment at multi-site promoters (**Fig. 2.2, Methods**). That is, for a cooperative (multi-site) promoter, when a weak activator is present at a sufficiently high dose, it can disrupt the dual mechanism by which the inhibitor operates, thereby greatly increasing transcription. Conversely, when a weak inhibitor is present at such a promoter at a sufficiently high dose, the inhibitor can disrupt activator binding via the dual mechanism, thereby greatly decreasing transcription. In both cases, the TF plasmid doses over which these transitions occur differ based upon TF strength, and therefore TF plasmid dose responses at cooperative promoters are described by diagonal tuning rather than vertical tuning. We note that additionally, at a very high dose or affinity of ZFa, the mechanistic interaction model used here shows a non-monotonic dose response due to the formation of TF.RNAP and DNA.TF at the expense of DNA.TF.RNAP. Non-monotonicity is considered a non-ideal behavior (observed in a subset of ZFa plasmid dose response experiments in **Supplementary Fig. A2.5a,b**); this is not represented in the concise COMET model but can be captured in the mechanistic interaction model used here. The lower two rows show the simulated effects as DNA and RNAP are sequestered into non-productive complexes. (f) Simulated effects of competition between an activator and inhibitor. With a x1 promoter, the inhibitor mechanism is competitive without loss of cooperativity. Rows show simulated dose responses with respect to one TF (ZFa or ZFi) while the other TF is held constant. The arrow in each row indicates the diagonal tuning that results from altering both the ZFa and ZFi affinity for DNA. Error bars depict S.E.M. Source data are provided as a **Source Data** file.

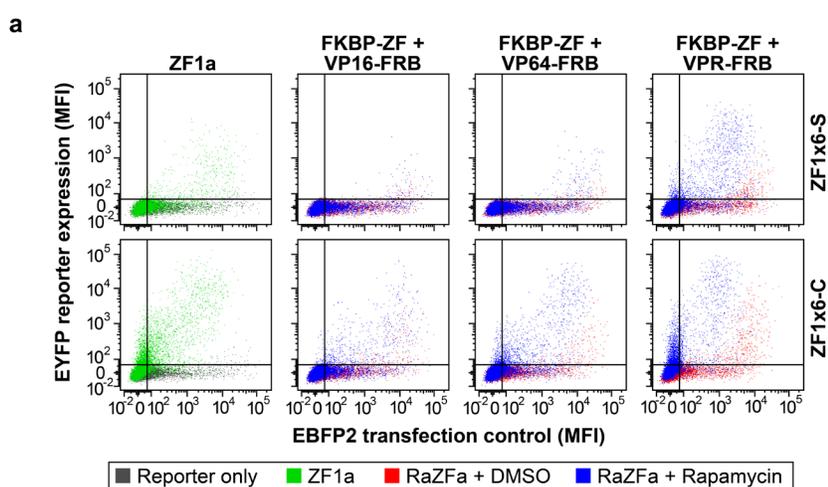


Supplementary Fig. A2.9 Workflow for landing pad (LP) integration and stable cell line generation.

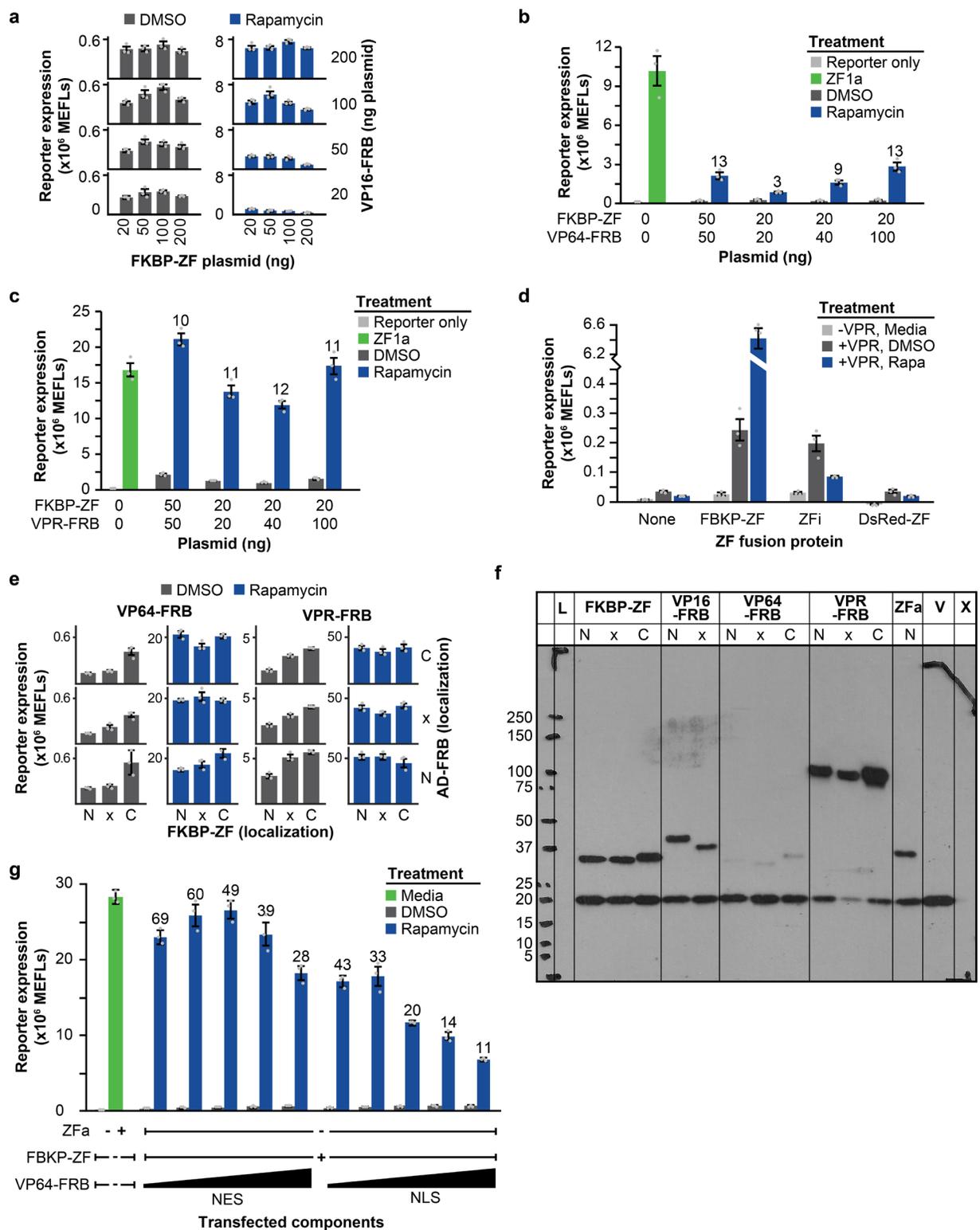
(a) The pcDNA-based transcription unit positioning vector (TUPV) permits transfer of COMET components into the mMoClo system¹⁷³ through restriction enzyme (RE) cloning. Notable features include the pcDNA multiple cloning site (MCS) for coding sequences and convenient RE sites placed at the 5' side of the promoter. Barcodes unique to each TUPV, found 5' of the promoter and 3' of the terminator, enable sequencing of the TUPV contents after TUPVs are combined into an integration vector¹⁷⁶. Two repeats of the CHS4 insulator were included to mitigate transcriptional readthrough¹⁷⁴. Details of these TUPVs and their features are in **Methods** and **Supplementary Tables A2.5, A2.6**. (b) After COMET and other components are transferred into TUPVs by RE cloning, they can be combined into a single plasmid with one to nine transcription units through a Bpil-mediated golden gate reaction. The integration vector serves as the backbone for the reaction and contains an attB site (for downstream integration into the LP) 5' of a promoter-less puromycin resistance gene, dual CHS4 insulators, a *crtRed* operon for Red-White colony screening that is replaced by the inserts from the TUPVs and Linker Vector, and dual CHS4 insulators to protect the 3' end of the LP locus after integration. The end result is the integration vector with cargo comprising multiple transcription units. (c) Integration of the LP cargo into the LP by BxB1-mediated recombination of the attB site on the IV and the genomic attP site integrates cargo into the AAVS1 safe harbor locus in the HEK293FT-LP cell line. Cells with successful integrations are no longer yellow (EYFP+) or hygromycin resistant; instead, they are blue (EBFP+) and resistant to puromycin and blasticidin. (d) After IV delivery by transient transfection and integration via BxB1, cells are expanded, selected with puromycin and blasticidin, and then sorted via FACS based upon EBFP2 expression (a typical gate used for this step is illustrated on the histogram). Additional details of this methodology and rationale are in **Methods**.



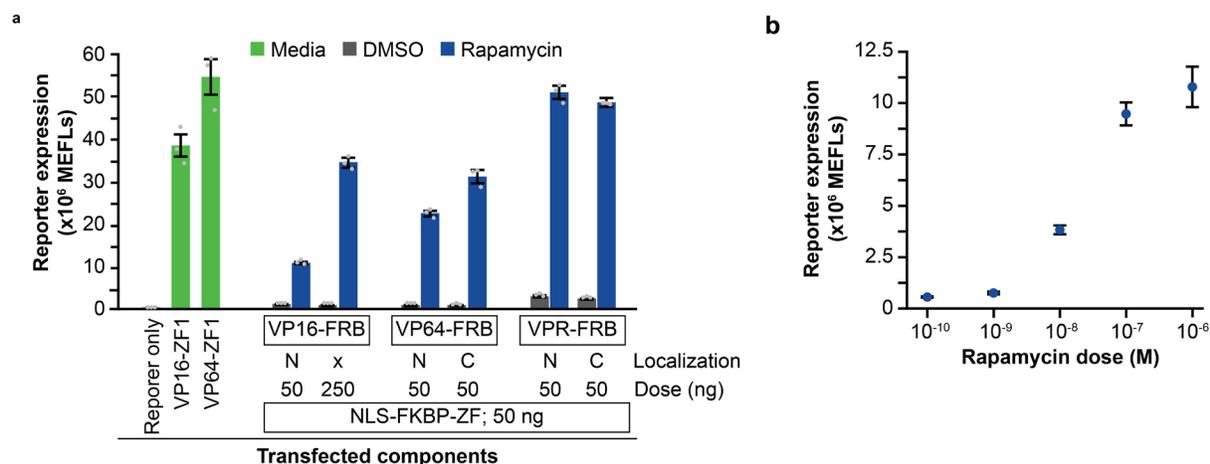
Supplementary Fig. A2.10 Characterization of genomically integrated COMET TFs. Representative histograms of reporter expression shown in **Fig. 2.6** are presented here with truncation of the y-axis (frequency) to better visualize differences between populations in each expression context: (a) multiple plasmid transfection, (b) single plasmid transfection, and (c) stable integration. Most trends observed in the stable genomic integration context appear to follow the trends observed in the context of transient transfection; these trends include an increase in reporter expression with addition of binding sites, and an increase in reporter expression for compact versus spaced promoter architectures. (d) A genomically-integrated promoter containing a single integrated ZF1 binding site induces low but significant reporter expression compared to background (one-tailed Welch's *t*-test **** $p < 0.0001$). All cell lines except for the ZF1x0 +ZFa line were sorted by FACS (outlined in **Methods**) before expansion and assaying. Data in all panels reflect reporter expression for cells that express EBFP2—a constitutive fluorescent marker driven from a gene that is genomically integrated with the cargo. Error bars depict S.E.M. Source data are provided as a **Source Data** file.



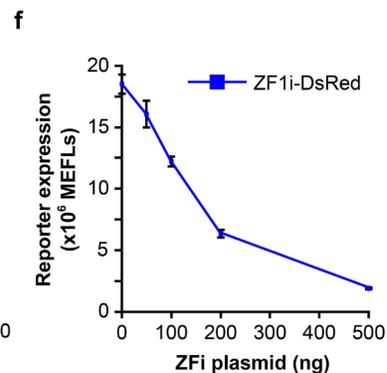
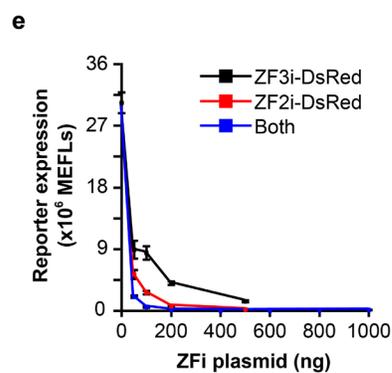
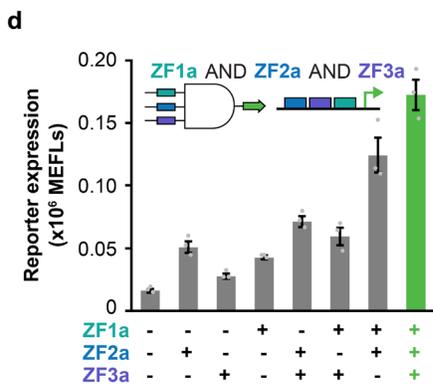
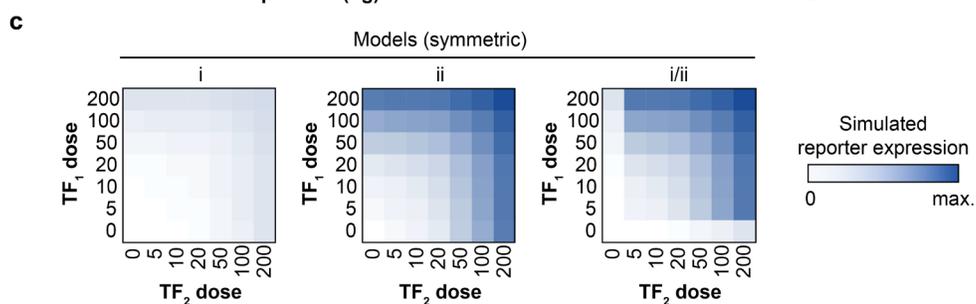
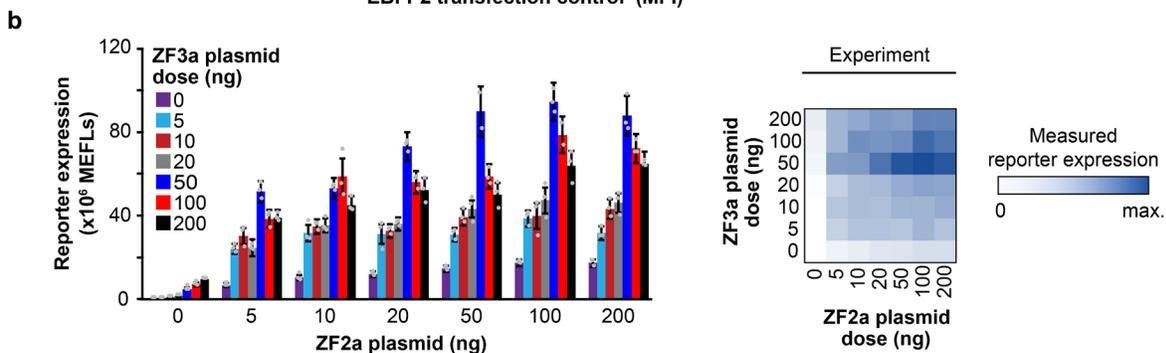
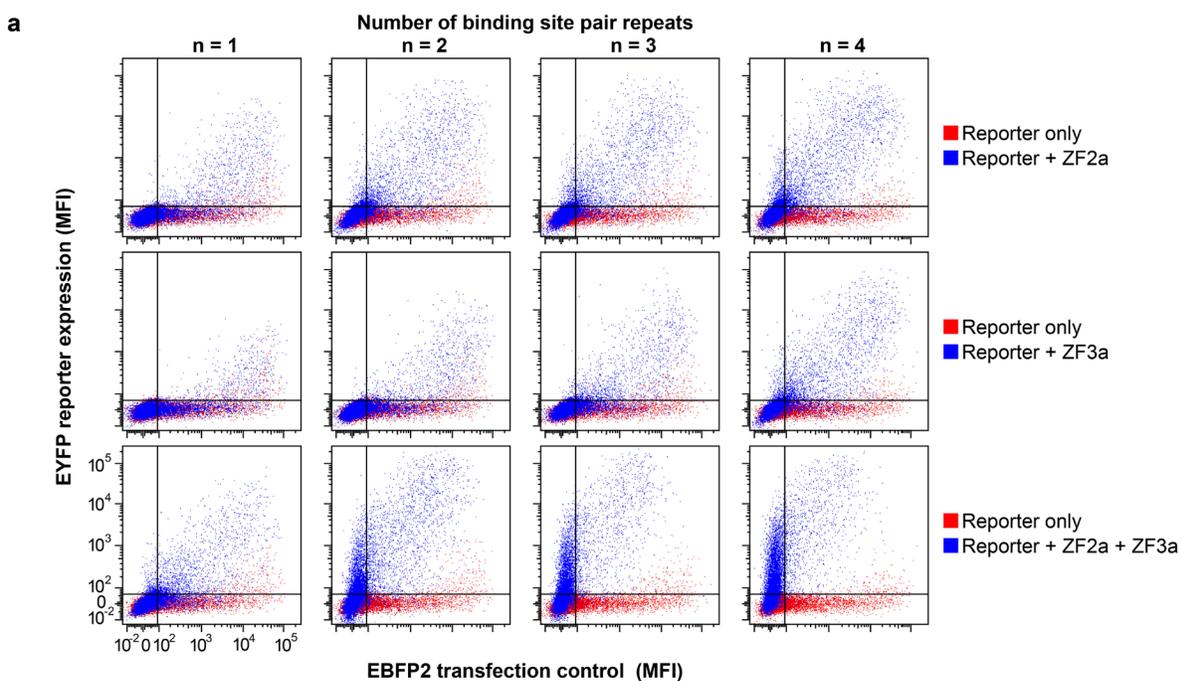
Supplementary Fig. A2.11 Characterization and tuning of RaZFa activity. (a) Representative flow cytometry plots from **Fig. 2.7b**. Rapamycin induces gene expression via RaZFa from COMET promoters similar to how a ZFa induces gene expression from these promoters. The effects of AD strength, as first seen in **Fig. 2.4**, apply similarly to RaZFa.



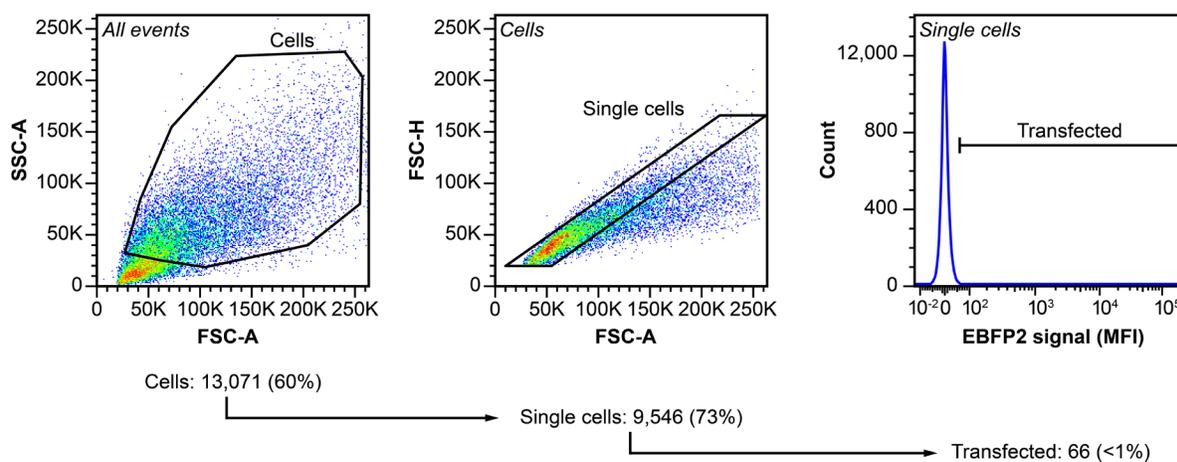
Supplementary Fig. A2.12 Characterization and tuning of RaZFa activity. (a) Data from **Fig. 2.7c** are shown as a bar graph. (b–c) Effect of component doses on the performance of (b) VP64-based RaZFa and (c) VPR-based RaZFa. Unlike the case of VP16-based RaZFa (**Fig. 7c**), decreasing the FKBP-ZF dose and increasing the AD-FRB dose did not lead to an increase in fold induction for VPR-based RaZFa. For all RaZFa, reporter expression was significantly higher with rapamycin than with the DMSO vehicle (one-tailed Welch's *t*-test, all $p < 0.01$). (d) Investigation of rapamycin-independent RaZFa activity. With a ZF1x6-C promoter, VPR-FRB was transfected alone or with FKBP-ZF, ZFi, or ZFi-DsRed, and cells were treated with rapamycin or DMSO vehicle. (e) Data from **Fig. 7d** are shown as a bar graph. (f) Western blot of RaZFa components with tags for subcellular localization (N: nuclear, NLS; C: cytoplasmic, NES). All components (upper bands) contain an N-terminal 3x-FLAG tag. Cells were co-transfected with a 3x-FLAG-tagged NanoLuciferase (lower band) as a loading and transfection control. VP64-FRB is expressed less than the other AD-FRB fusions, and VPR-FRB is expressed more. The NES tag increases protein expression level. (g) Increasing the dose of VP64-FRB above the doses tested in (b). To investigate whether the low expression of VP64-FRB limited the rapamycin-induced activation of reporter expression, VP64-FRB with a NES or NLS was transfected at increasing doses with FKBP-ZF (20 ng of FKBP-ZF; 100, 200, 400, 600, or 800 ng of VP64-FRB). While rapamycin-inducible reporter expression increased with VP64-FRB dose, so did rapamycin-independent reporter expression, resulting in lower fold induction at high VP64-FRB doses. At high doses, squelching (as evidenced by decreased rapamycin-inducible reporter expression) was also evident and was more pronounced for the NLS-tagged than the NES-tagged VP64-FRB. For all RaZFa, reporter expression was significantly higher in the presence of rapamycin compared to the DMSO vehicle cases (one-tailed Welch's *t*-test, all $p < 0.01$). Error bars depict S.E.M. Source data are provided as a **Source Data** file.



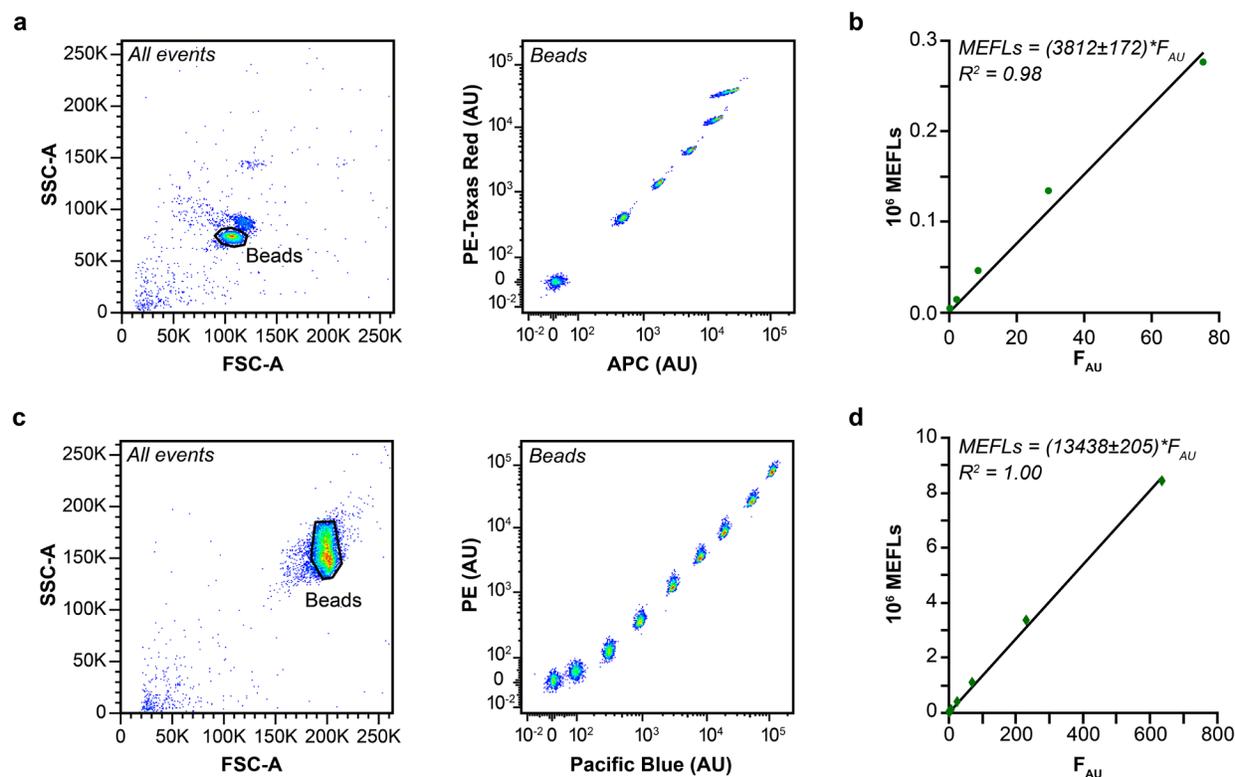
Supplementary Fig. A2.13 Results of RaZFa tuning. (a) Comparison of initial RaZFa performance with performance after optimization. Cells were transfected with various doses of RaZFa components, with various subcellular localization tags. For VP16-based RaZFa, increasing the dose of the VP16-FRB component five-fold and removing the localization tag increased reporter fold induction from 11 to 38 (~360% increase). For VP64-based RaZFa, localizing the VP64-FRB component to the cytoplasm rather than the nucleus increased fold induction from 25 to 40 (60% increase). For VPR-based RaZFa, localization of the VPR-FRB domain to the cytoplasm had a more modest effect on fold induction (17 to 20; 18% increase). Despite the substantial induced state, VPR-based RaZFa also exhibit high background. (b) Rapamycin dose response of VP64-based RaZFa. The range for inducibility spans approximately 1 nM to 1 μ M rapamycin. Error bars depict S.E.M. Source data are provided as a **Source Data** file.



Supplementary Fig. A2.14 Implementing Boolean logic with COMET. (a) Representative flow cytometry plots from **Fig. 2.8b**. The “reporter only” sample is the same across panels for a given promoter. (b) Data from **Fig. 2.8c** are presented as a bar graph and a heat map. (c) Model simulations corresponding to **Fig. 2.8d** are displayed as heatmaps. These simulations depict different types of AND gates broadly (**Methods**) and are not parameterized to data. The hybrid model (i/ii) depicting a transition from x3-S along the perimeter to x6-C in the interior of the landscape captures COMET AND gate behavior better than standard non-hybrid models. As discussed in **Supplementary Fig. A2.8**, this compact model does not encompass mechanisms that could lead to squelching, which is why experimentally observed diminishment in reporter output at high ZFa doses is not captured. (d) Three-input promoter with one repeat of the three-site motif. Cells were transfected with the reporter (containing one site each for ZF1, ZF2, and ZF3) and combinations of the ZFa plasmids. The difference between any two ZFa and all three ZFa was modest; this promoter is not ideal for implementing Boolean logic, since the *effective architecture* switches only from x2-S or x2-C to x3-C (when both ZFa inputs are present, rather than one or the other). AND gate activation is considered statistically significant if reporter expression with all three ZFa present is greater than the sum of reporter expression with the three ZFa expressed individually, and greater than the sum of reporter expression with two ZFa co-expressed and the other expressed individually. This criterion was met for all conditions tested (one-tailed Welch’s *t*-test, $p < 0.05$) except for (ZF3a)+(ZF1a+ZF2a) vs. (ZF1a+ZF2a+ZF3a) ($p = 0.17$). (e, f) Inhibiting the three-repeat AND gate using ZFi-DsRed. The promoter has alternating sites for ZF2 and ZF3 and employs the strategy shown in **Supplementary Fig. A2.7a**, in which a ZF2 site overlaps with a ZF1 site. Plasmids for ZF2a, ZF3a, the reporter, and ZFi-DsRed (dose response) were transfected into cells. Although high doses of ZF1i-DsRed were necessary for inhibition, ZF2i-DsRed and ZF3i-DsRed had more potent effects, and ZF2i-DsRed and ZF3i-DsRed in combination conferred strong inhibition even at low plasmid doses. Results from this experiment informed the dose choices for the experiment in **Fig. 2.8f**, which uses 50 ng of each ZFa plasmid and 100 ng of each ZFi-DsRed plasmid. Error bars depict S.E.M. Source data are provided as a **Source Data** file.



Supplementary Fig. A2.15 Flow cytometry gating. The plots show a sample of cells transfected with the modified pcDNA vector used in this study without any coding sequence in the multiple cloning site. These cells were *not* transfected with an EBFP2 expression plasmid. In the gating procedure, HEK293FT cells were identified based on the FSC-A vs SSC-A profile. From this population, single cells were identified based on the FSC-A vs FSC-H profile. The transfected population was defined as all single cells with a greater EBFP2 signal than the sample of single cells that was transfected with vector-only DNA (EBFP2 was used as a transfection control in this experiment and the vast majority of the experiments in this paper). This gate was drawn such that it did not encompass more than 1% of this non-fluorescent population of cells.



Supplementary Fig. A2.16 Profile of fluorescent calibration beads. Rainbow Calibration Particles (RCP; a, b) have six fluorescent bead populations, while UltraRainbow Calibration Particles (URCP; c, d) have nine fluorescent bead populations and are brighter than RCP. (a, c) Beads were identified based on the FSC-A vs. SSC-A profile. The beads are fluorescent in the majority of fluorescent channels on the flow cytometer. For each experiment, two channels were used to identify the bead populations. (b, d) The mean intensity of each population in the FITC (EYFP) channel in arbitrary units (F_{AU}) was recorded and plotted against the manufacturer-supplied number of fluorophores on the beads for each population (MEFLs). To generate the calibration curve, a linear regression was performed with the constraint that the y-intercept equals zero. In each experiment, MFI are converted to MEFLs by using the multiplier on F_{AU} obtained from the regression. The magnitude of the multiplier, along with corresponding uncertainty (\pm one standard error), is reported in parentheses on each plot (b, d).

Supplementary Table A2.1. ZFa and fitted parameters for x6-C promoters.

Zinc finger ¹	Reference number ²	Binding site ³	b^5	m^5	w^5
ZF1	43-8	a GAG TGA GGA c	0.08	33	0.036
ZF2	37-12	t GAG GAC GTG t	0.25	54	0.018
ZF3	158-2	t GTA GAT GGA g	n.d.		
ZF4	97-4	a TTA TGG GAG a			
ZF5	92-1	a GAT GTA GCC t			
ZF6	150-4	g GTG TAG GGG t	0.02	58	0.043
ZF7	172-5	a GGA GGG GCT c	0.11	46	0.025
ZF8 ⁴	173-3	a GAT GAA GCT g	0.07	43	0.041
ZF9 ⁴	42-10	a GAC GCT GCT c	0.46	33	0.096
ZF10	13-6	a GAA GAT GGT g	0.01	31	0.037
ZF11	36-4	c GAA GAC GCT g	0.08	32	0.025
ZF12 ⁴	62-1	g GCC GAA GAT a	0.15	33	0.065
ZF13	21-16	a TTA GAA GTG a	0.04	41	0.012
ZF14 ⁴	14-3	g GAC GAC GGC a	0.20	30	0.069
ZF15 ⁴	129-3	c GGG GAC GTC a	0.18	33	0.007
ZF16	54-8	a TGG GTG GCA t	n.d.		
ZF17	55-1	c TGG GGT GCC c			
ZF18	93-10	c TTT GTT GGC a			
ZF19	151-1	t GCA GGA GGT g			

¹All ZFa are WT (RRRR) with the VP16 AD. n.d. = no data. ZFa are numbered by order of appearance in this work. This is the nomenclature used in another study¹⁶⁹; we provide this information to facilitate cross-referencing.³ Each ZF binds one nucleotide triplet indicated by upper-case letters. Lower-case letters indicate nucleotides flanking the binding site; we observed that these flanking residues may influence binding affinity (**Supplementary Fig. A2.5**). The flanking nucleotides listed here are known to confer strong binding affinity¹⁶⁹, however some constructs in this study contain other nucleotides (promoter sequences are provided in **Supplementary Table A2.2**).⁴ These ZFa exhibited squelching at high ZFa plasmid doses. ⁵**Supplementary Fig. A2.5** contains the response profiles that were used to fit these parameters.

Supplementary Table A2.2. Fitted parameters for modifications to promoter architecture and ZFa domains.

Figure	ZF	AD	Promoter	<i>b</i>	<i>m</i>	<i>w</i>
2a	ZF1	VP16	ZF1x1	0.08	1.0	0.036
			ZF1x3-S		1.3	
			ZF1x6-S		3	
			ZF1x12-S		8.5	
			ZF1x3-C		7.1	
			ZF1x6-C v1		33	
			ZF1x12-C		41	
S2d	ZF1	VP16	ZF1x6-C CMV_Min	0.26	33	0.058
			ZF1x6-C SV40_Min	0.43	7.5	0.046
4b	ZF1(RARR)	VP16	ZF1x6-C v1	0.08	26	0.018
	ZF1(ARRR)				19	0.010
	ZF1(AARR)				15	0.011
	ZF1(RAAR)				13	0.0043
	ZF1(RAAA)				13	0.0023
	ZF1(AAAR)				7	0.0040
	ZF1(AAAA)				7	0.0017
4f	ZF1(AAAA)	VP64	ZF1x6-C v1	0.08	24	0.012
		VPR			78	0.020

Supplementary Table A2.3. Instrument specifications for analytical flow cytometry

Instrument	Fluorescent protein	Parameter	Excitation laser	Filter set
BD LSR II	EBFP2, BFP	DAPI	Violet 407 nm	450/50
BD LSR II	EYFP	FITC	Blue 488 nm	505LP, 530/30
BD LSR II	DsRed Express2	PE-Texas Red	Yellow Green 561 nm	600LP, 610/20
BD LSR Fortessa	EBFP2, BFP	Pacific Blue	Violet 405 nm	450/50
BD LSR Fortessa	EYFP	FITC	Blue 488 nm	505LP, 530/30
BD LSR Fortessa	DsRed Express2	PE-Texas Red	Light Green 552 nm	600LP, 610/20
BD LSR Fortessa	mKate2	PE-Texas Red	Light Green 552 nm	600LP, 610/20

Supplementary Table A2.4. Instrument specifications for FACS

Instrument	Fluorescent protein	Parameter	Excitation laser	Filter set
BD FACS Aria	EBFP2	Pacific Blue	Violet 407 nm	450/50
BD FACS Aria	EYFP	FITC	Blue 488 nm	505LP, 525/30

Supplementary Table A2.5. Parameters used for standard models of activation

Model	Landscape	b or α_0	m or α	w	n	ρ	
i	Upper	0	1	0.005	1	0	
	Lower				1	1	
ii	Upper				2	0	
	Lower				2	1	
iii	Upper		spaced		0.036	1	1
	Lower		compact			1	1
iv	Upper		spaced			2	1
	Lower		compact			2	1
ZF-TF	Upper	0.08	spaced	n/a (1)		n/a	
	Lower		compact	n/a (1)		n/a	

Supplementary Table A2.6. Parameters for the removal of cooperativity by COMET inhibitors

Inhibitor	Promoter	l	u
ZFi	x6-compact	n/a	n/a
	x6-spaced	0	1.5
ZFi-DsRed	x6-compact	0.5	2
	x6-spaced	0	1

Supplementary Table A2.7. Parameters used for standard models of AND gates

Model	a_1 and a_2	a_{12}	w_1 and w_2	ρ
i	0.2	1	0.004	0
ii	1	1		0
iii	0.2	1		5
iv	1	1		5

Supplementary Table A2.8. Promoter states for the COMET four-input gate

ZF2a	-				+				-				+			
ZF3a	-				-				+				+			
ZF2i-DsRed	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ZF3i-DsRed	-		+		-		+		-		+		-		+	
Effective # of sites	x0	x0	x0	x0	x3	x1	x3	x1	x3	x3	x1	x1	x6	x3	x3	x1
Effective spacing	n/a	n/a	n/a	n/a	S	n/a	S	n/a	S	S	n/a	n/a	C	S	S	n/a
Competitive inhibition	n/a	n/a	n/a	n/a	N	Y	N	Y	N	N	Y	Y	N	Y/N	N/Y	Y/Y

Supplementary Note A2.1. Nomenclature used in this manuscript.

The ZF sequences used here were previously described by Khalil and colleagues¹⁶⁹. We refer to ZF domains as ZF1, ZF2, etc. by order of appearance in this work for clarity. **Supplementary Table A2.1** provides a summary to facilitate cross-referencing of this nomenclature with that which was previously used by Khalil and colleagues. Addgene pages associated with our plasmids refer to ZFs by the original nomenclature.¹⁶⁹

For ZF fusion proteins (e.g., ZFi-DsRed or VP16-ZF), the order of domains in the name does not indicate whether the protein is an N-terminal or C-terminal fusion. DsRed-Express2, FKBP, and all ADs fused to a ZF domain were fused to the N-terminus of the ZF domain. For AD-FRB, the AD is fused to the N-terminus of the FRB domain. Unless explicitly stated otherwise, all ZF proteins and ZF fusion proteins begin with a 3X-FLAG tag and SV40 NLS. More detailed information on the sequence of these constructs is provided in **Supplementary Tables A2.3, A2.4, A2.8**. Throughout, we refer to DsRed-Express2 as “DsRed” for brevity; this abbreviation always refers to DsRed-Express2 and never to either the original DsRed or DsRed2.

Appendix 3. Supplementary information for Chapter 3

A version of this appendix was previously published as Supplementary Information for:

Hailey I Edelstein, Patrick S Donahue, Joseph J Muldoon, Anthony K Kang, Taylor B Dolberg, Lauren M Battaglia, Everett R Allchin, Mihe Hong, Joshua N Leonard, Elucidation And Refinement Of Synthetic Receptor Mechanisms, *Synthetic Biology*, <https://doi.org/10.1093/synbio/ysaa017>

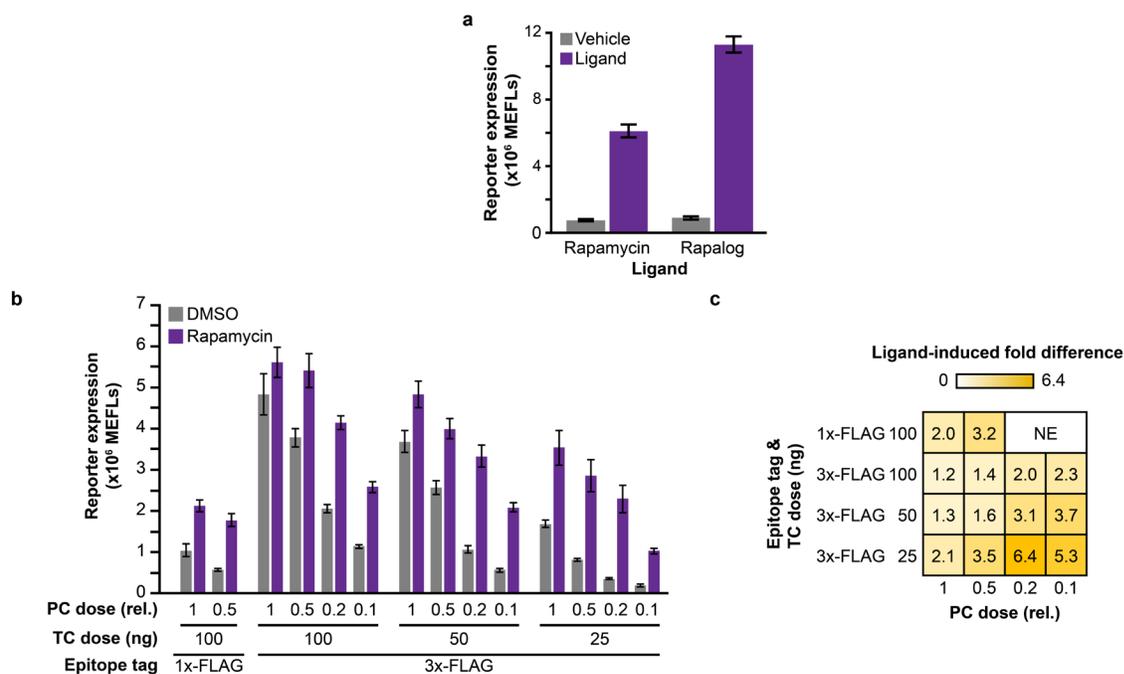
This is an abbreviated version of the SI, containing only the Supplementary Figures that I contributed to. The complete SI can be found online at the journal website above.

Contents:

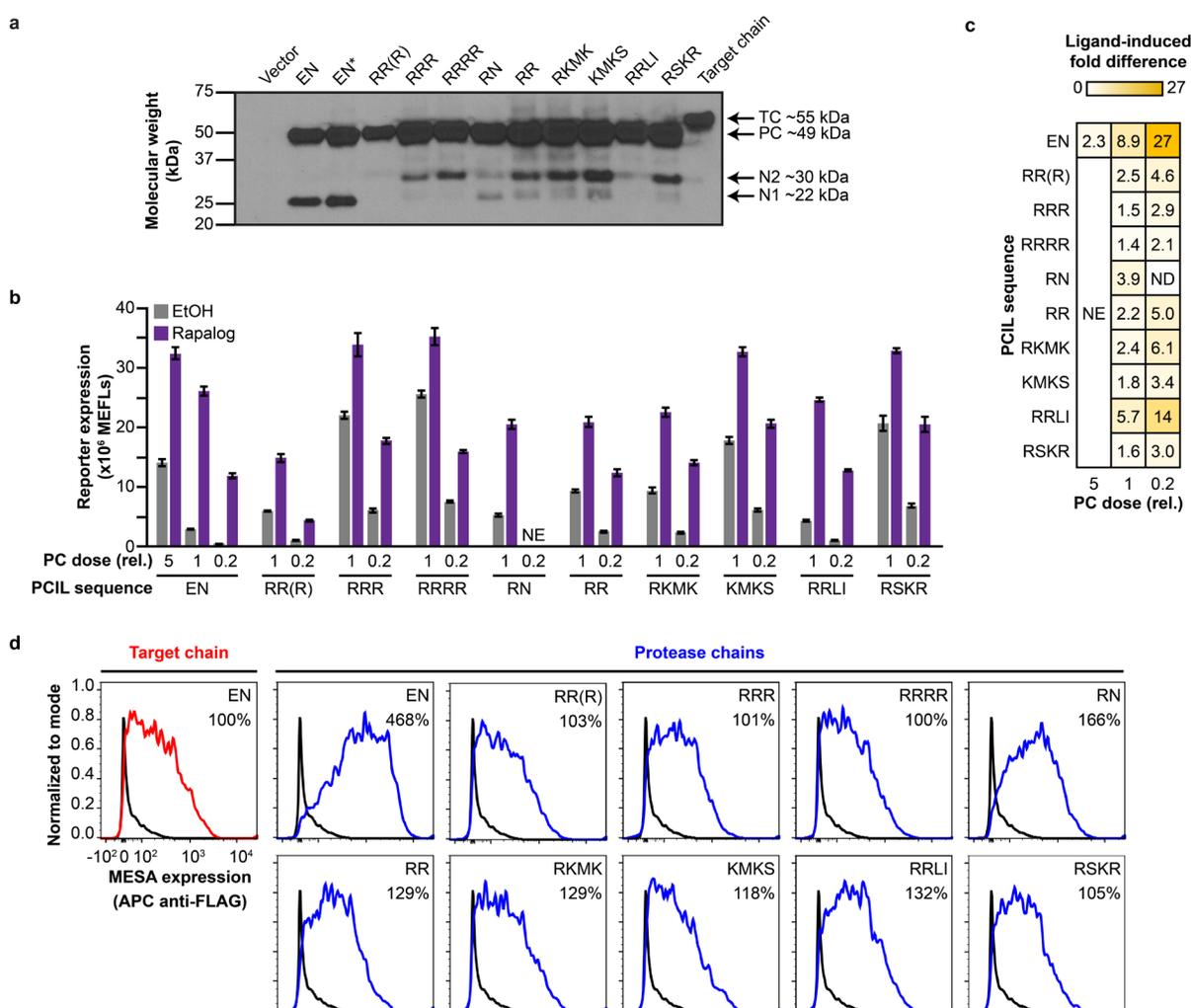
Supplementary Figures A3.1-A3.6

Supplementary Tables A3.1–A3.5

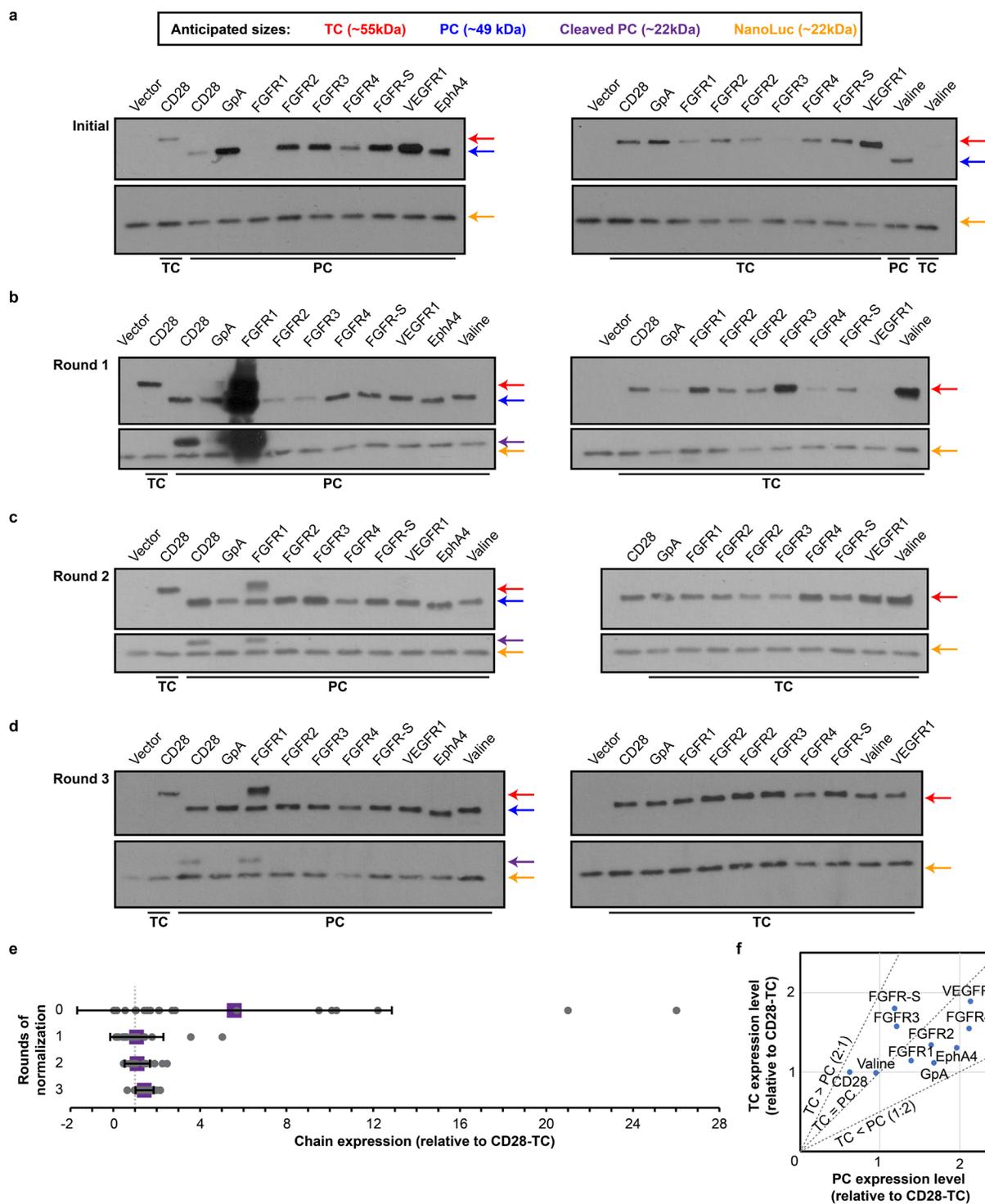
Supplementary Notes A3.1–A3.3



Supplementary Fig. A3.1. Comparisons across MESA assays. (a) Comparison between rapalog and rapamycin ligands. The rapalog AP21967 is a synthetic rapamycin analog that binds only to FRB containing the T2098L mutation (which are contained in all reported rapamycin-sensing MESA receptors). Rapalog is less immunosuppressive than rapamycin, due to diminished interference with the endogenous mTOR pathway²⁹⁵. Rapalog produces higher ligand-induced reporter expression than does rapamycin. We noted some cytotoxicity by microscopy for cells treated with rapamycin, which is consistent with this outcome. The vehicle for rapamycin is 50% DMSO/50% water. The vehicle for rapalog is 100% EtOH. Stock and final working ligand concentrations used in cell culture are listed in **Supplementary Table A3.1**. (b,c) To increase the sensitivity of detection on Western blots and via flow cytometry, we replaced the 1x-FLAG epitope tag on the N-terminus of each MESA chain with a 3x-FLAG tag. The use of 100 ng of plasmid encoding each MESA chain led to higher levels of background and ligand-induced signaling for the 3x-FLAG constructs than for the 1x-FLAG constructs. As this phenomenon is similar to what has been observed when MESA receptors are overexpressed, we performed a combinatorial dose scanning experiment, reducing the plasmid doses of TC and PC (labels show the PC dose relative to the TC dose, where a value of unity denotes equal amounts of each plasmid). The use of 25 ng of 3x-FLAG tagged TC plasmid yielded similar fold differences compared to 1x-FLAG tagged constructs (c), though both the background and ligand-induced signaling remained higher with the 3x-FLAG constructs. Abbreviation: NE, not evaluated. Bars are the means of three biological replicates, and error bars depict S.E.M.



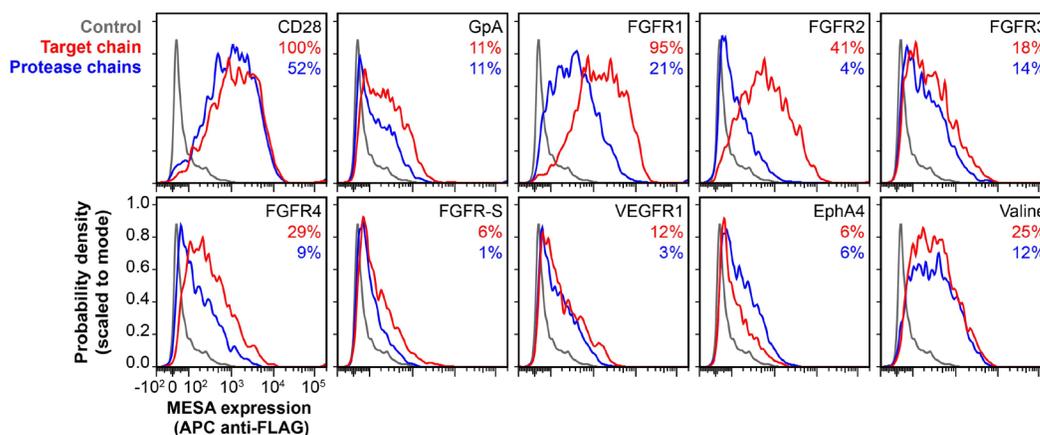
Supplementary Fig. A3.2. Investigating the role of PCIL on PC stability. (a) Based on a prior observation that many native juxtamembrane domains are positively charged, we substituted the PCIL with linkers comprising varying numbers of arginine residues or sequences from native receptors (EN is the original PCIL; EN* contains the original PCIL and the D81N mutation in the TEVp; the parenthetical R in the fourth lane indicates that this R replaces the first amino acid TEVp coding sequence). While in many cases a diminishment of the originally observed cleavage product (N1) was observed, we also observed new cleavage products (N2) with many of the constructs. (b,c) In a functional assay, with 25 ng of each 3x-FLAG tagged MESA chain plasmid, PCIL substitution led to decreased receptor performance through increased background. As this outcome was consistent with what would be expected if the number of intact PCs per cell were increased (far left condition), we decreased the PC plasmid dose by 5-fold (to 5 ng) while holding the TC plasmid dose constant (at 25 ng). Though this change did result in lower background and ligand-induced signaling, only one case (RRLL) exhibited fold differences greater than that of the original construct (EN); however, the new outcome was not ideal as it came at a cost of decreased ligand-induced signaling. The data shown here include the data reported in Fig. 3.1f in addition to samples with adjusted DNA doses. Bars are the means of three biological replicates, and error bars depict S.E.M. Abbreviation: NE, not evaluated. (d) Surface staining profiles of cells expressing 1x-FLAG tagged MESA PCs with various PCIL sequences. All PCIL sequences investigated led to decreased expression relative to the original construct (EN). Percentages indicate PC expression relative to TC expression (the left plot). A stained, vector-only (no MESA chain) sample is shown in each plot as a black-lined histogram.



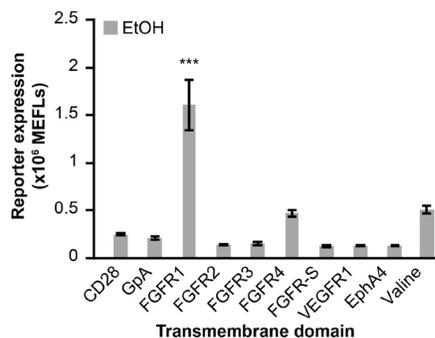
Supplementary Fig. A3.3. Normalization of protein expression for MESA containing various TMDs.

(a) Western blot of MESA containing various TMDs, using the same plasmid dose for each construct. Successive rounds of normalization were performed by quantifying the expression of each chain (normalizing first to the intensity from the co-transfected 3x-FLAG tagged NanoLuciferase in each sample to control for loading and transfection efficiency, and then normalizing to the normalized intensity from the CD28-TMD TC as an internal control) and varying the plasmid dose in each successive round. Details are provided in **Methods**. Anticipated sizes for applicable bands are indicated by the colored arrows, as described by the legend, which also applies to **b–d**. Western blots after (b) one round, (c) two rounds, and (d) three rounds of normalization. The Western blot in **d** is the same as in **Fig. 3.2b**. Plasmid doses in **d** were used in **Fig. 3.2c,d**. (e) The sequential normalization procedure resulted in more similar protein expression levels across with each round of quantification. Each chain's relative expression level is represented by a grey dot. The mean relative expression level of all chains is depicted by the purple square. Error bars represent the S.D. The vertical dotted grey line at 1 relative unit corresponds to the expression level of the CD28-TMD TC base case. (f) Ratio of TC:PC expression for each chain, using the final plasmid doses. All constructs were expressed within a 1:2 and 2:1 TC:PC ratio.

a

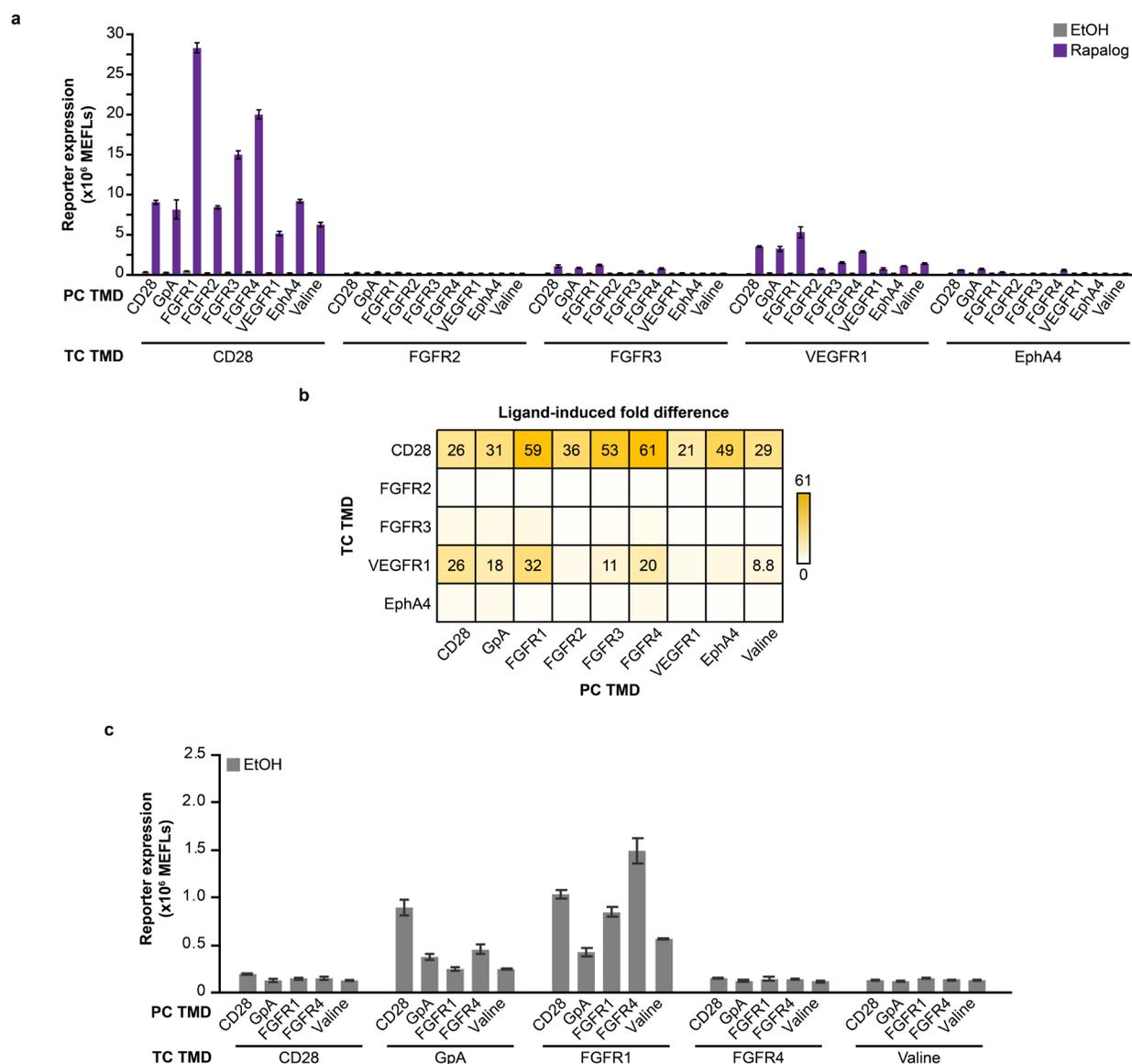


b

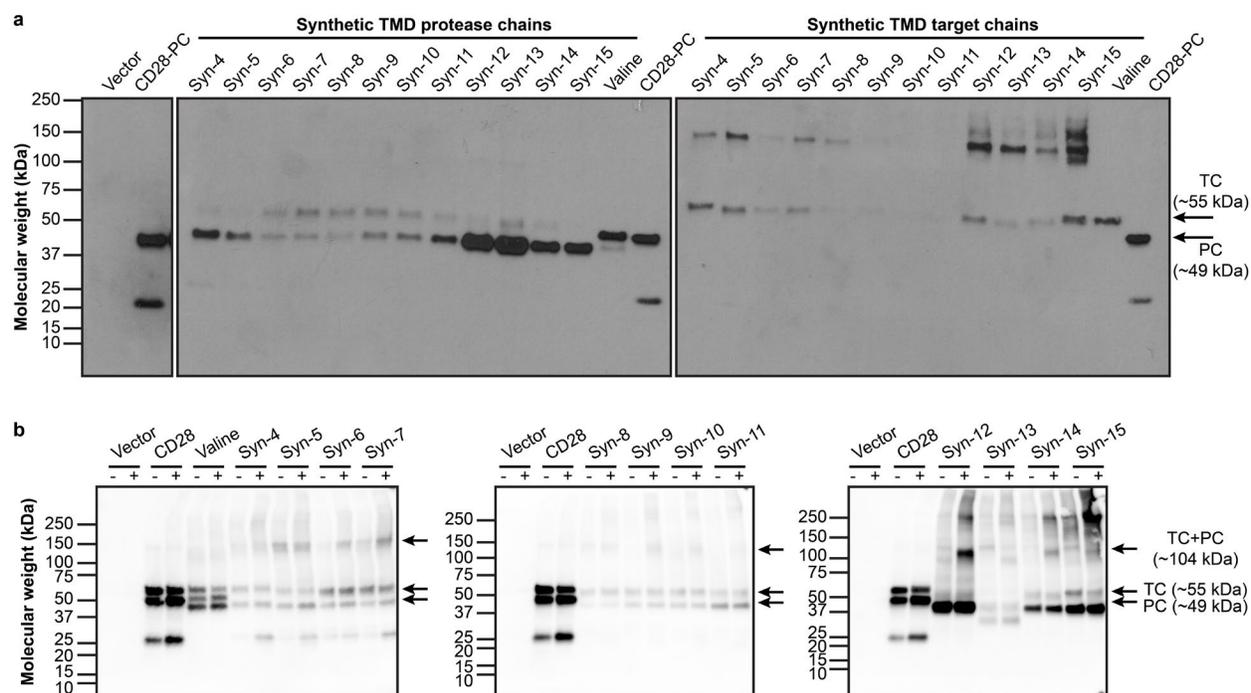


Supplementary Fig. A3.4. Surface stain and background signal of MESA containing various TMDs.

(a) Surface stain of MESA chains containing various TMDs, using the plasmid doses that yielded approximately equal whole-cell protein expression in **Supplementary Fig. A3.3d**. Percentages are the receptor surface expression relative to the CD28-TMD TC. The control sample in all plots is a stained, vector-only (no MESA) sample. (b) Reporter expression for the samples treated with vehicle (EtOH) shown in **Fig. 2c**. Data are reproduced here with a y-axis scaled to highlight differences between background reporter expression conferred by pairs of receptors with different matched TMDs. Comparing all vehicle-treated TMDs, only FGFR1 exhibits significantly greater background signal than the others (one-way ANOVA, *** $p < 0.001$). Statistical analysis is in **Supplementary Note A3.1**. Bars are the means of three biological replicates, and error bars depict S.E.M.



Supplementary Fig. A3.5. Evaluating TCs that showed low ligand-induced signaling with TMD-matched PC. (a) TCs that produced little or no ligand-induced signaling in Fig. 3.2c were tested with PCs containing matched or mixed TMDs. CD28-TMD TC was used as a positive control, as this chain cleaved well with the matched CD28-TMD PC. Most combinations produced little or no ligand-induced signaling, except for those with the VEGFR1-TMD containing TC. However, even in these cases, ligand-induced signaling remained below those seen with the CD28-TMD. (b) Corresponding ligand-induced fold difference values are shown for conditions that yielded a significant difference between the vehicle and ligand-treated conditions. Statistical analysis is in **Supplementary Note A3.3**. (c) Reporter expression for the samples treated with vehicle (EtOH) shown in Fig. 3.2d. Data are reproduced here with a y-axis scaled to highlight differences between background reporter expression conferred by pairs of receptors with different mixed TMDs. Background signal varies with choice of TC TMD and PC TMD and the interaction between these variables, but TC choice explains most of this variance. Statistical analysis (two-way ANOVA) is in **Supplementary Note A3.2**. Throughout all panels, bars are the means of three biological replicates, and error bars depict S.E.M.



Supplementary Fig. A3.6. Investigating ligand-induced accumulation of MESA receptors containing synthetic TMDs. (a) Western blots of individual MESA chains containing synthetic TMDs show a range of expression, with higher expression for cases in which dimerizing motifs are located near the outer leaflet of the membrane. (b) Some co-transfected MESA chains containing matched, synthetic TMDs and a catalytically inactive TEVp on the PC (D81N mutation²⁹⁶) show modestly higher expression with rapalog treatment (+) than with vehicle-only treatment (-). The expression level trends observed across the vehicle-only conditions are similar to the expression level trends observed across the rapalog conditions. Therefore, we conclude that chain expression alone does not explain the trends in **Fig. 3.5b**.

Supplementary Table A3.1. Ligand concentrations for small molecule-sensing MESA systems

MESA	Ligand	Vehicle (% in stock)	Stock concentration	Vehicle (% in culture)	Working concentration
Rapamycin	Rapamycin (Santa Cruz Biotechnology sc-3504)	50% DMSO, 50% H ₂ O	0.1 mM	0.05% DMSO	0.1 μM
	Rapalog (Takara AP21967)	100% EtOH	0.1 mM	0.1% EtOH	0.1 μM
Gibberellin	GA3-AM (Tocris 5407)	100% DMSO	10 mM	0.1% DMSO	10 μM
Abscisic acid	ABA (GoldBio A-050)	100% EtOH	100 mM	0.1% EtOH	100 μM

Supplementary Table A3.2. Rapamycin-sensing MESA expression normalization doses¹

Plasmid #	MESA components	Initial (S6a)	Round 1 (S6b)	Round 2 (S6c)	Round 3/Final (S6d, 2b)
pPD801	PC CD28-TMD	25	18	32.3	21.9
pPD802	PC GpA-TMD	25	1	1.9	3.7
pPD803	PC FGFR1-TMD	25	750	75.0	88.2 ²
pPD804	PC FGFR2-TMD	25	2	9.2	9.6
pPD805	PC FGFR3-TMD	25	2	14.6	8.7
pPD806	PC FGFR4-TMD	25	9	7.9	16.3
pPD807	PC FGFR-S-TMD	25	3	6.9	6.5
pPD808	PC VEGFR1-TMD	25	1	1.6	2.0
pPD809	PC EphA4-TMD	25	4	9.2	9.6
pPD810	TC CD28-TMD	25	25	14.8	28.5
pPD811	TC GpA-TMD	25	9	25.0	25.0
pPD812	TC FGFR1-TMD	25	50	19.1	13.6
pPD813	TC FGFR2-TMD	25	15	34.7	37.6
pPD814	TC FGFR3-TMD	25	24	16.2	22.0
pPD815	TC FGFR4-TMD	25	181	13.5	24.9
pPD816	TC FGFR-S-TMD	25	16	36.2	78.9
pPD817	TC VEGFR1-TMD	25	12	38.5	20.5
pPD818	TC EphA4-TMD	25	2	13.9	14.2
pPD819	PC Valine-TMD	25	12	18.5	8.2
pPD832	TC Valine-TMD	25	750	211.4	85.8

¹Plasmid doses in this table are those used in a single well of a 24-well plate. As transfection doses scale with the volume of media, for the western blots shown in these figures (**Fig 3.2b**, **Supplementary Fig. A3.3**), a mass four times the doses shown in this table were used (i.e., 100 ng of pPD810 TC CD28-TMD).

²88.2 ng of plasmid (used per well of a 24-well plate) was used to calculate the plasmid dose for **Fig. 3.2b**. This resulted in multiple full-length PC bands (49 kDa and approx. 55 kDa, the latter of which we speculate is glycosylated). The 49 kDa band had a similar intensity as the other 49 kDa PC bands from MESA with other TMDs. To keep the total amount of full-length PC in the cells constant, we normalized the sum of the 49 and 55 kDa bands to the intensities of the other 49 kDa bands. This resulted in a dose of 33.5 ng (per well of a 24-well plate) of FGFR1-TMD PC used in functional assays (**Fig. 3.2c and 3.2d**).

Supplementary Table A3.3. Instrument specifications for analytical flow cytometry to quantify reporter expression

Instrument	Fluorescent protein	Parameter/ Channel name	Excitation laser	Filter set
BD LSR Fortessa	EBFP2	Pacific Blue	Violet, 405 nm	450/50
	EYFP	FITC	Blue, 488 nm	505LP, 530/30
	DsRed2	PE-Texas Red	Light Green, 552 nm	600LP, 610/20

Supplementary Table A3.4. Instrument specifications for analytical flow cytometry to quantify surface expression

Instrument	Fluorescent protein	Parameter/ Channel name	Excitation laser	Filter set
BD LSR Fortessa	EBFP2	Pacific Blue	Violet, 405 nm	450/50
	Phycoerythrin	PE	Light Green, 552 nm	575/26
	miRFP670	APC	Red, 640 nm	670/30
	Allophycocyanin	APC	Red, 640 nm	670/30

Supplementary Table A3.5. Instrument specifications for confocal microscopy to quantify FRET

Instrument	Fluorescent protein	Excitation laser	Beam splitter	Sensor	Acquisition band-pass filter
Leica SP5 II Laser Scanning Confocal Microscope	mCerulean (D)	Argon 458 nm	458/514 double dichroic	HyD*	465–500 nm
	mVenus (A)	Argon 514 nm		HyD*	520–560 nm
	FRET	Argon 458 nm		HyD*	520–560 nm

*Abbreviation: HyD, hybrid detector

Supplementary Note A3.1

Below are the outcomes from one-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of PC C-terminal peptide on the measured reporter expression, evaluated separately for vehicle- and ligand-treated conditions.

Functional signaling assay in **Fig. 3.1c**,

Vehicle treatment

- PC C-terminal peptide $p = 3.9 \times 10^{-13}$
 - All differences were significant (all $p < 0.05$) except for:
 - Comparisons between PCs that each contain a PRS-based peptide
 - PCs that contain the AIP_Y or AIP_A peptide and all of the PRS-based peptides
 - PC-AIP and PC-AIP_K or PC-AIP_M
 - PC-AIP_A and PC-AIP_Y
 - PC-AIP_M and PC-AIP_K
 - Notably, included within the comparisons that are significantly different (all $p \ll 0.01$) are comparisons between each PC variant with a C-terminal peptide and the PC without a C-terminal peptide.

Ligand treatment

- PC C-terminal peptide $p = 7.8 \times 10^{-19}$
 - All differences were significant (all $p < 0.05$) except for:
 - Comparisons between PCs that each contain a PRS-based peptide
 - PC-AIP and PC-AIP_K

Below are the outcomes from one-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of TMD on the measured reporter expression, evaluated separately for vehicle- and ligand-treated conditions.

Functional signaling assay in **Fig. 3.2c**, **Supplementary Fig. A3.4b**

Vehicle treatment

- TMD $p = 2.0 \times 10^{-9}$
 - All differences were not significant (all $p > 0.05$) except for comparisons between each TMD and the FGFR1 TMD ($p \ll 0.001$).

Ligand treatment

- TMD $p = 2.4 \times 10^{-18}$
 - All differences were significant (all $p < 0.05$) except for:
 - CD28 and Valine
 - Comparisons between EphA4 and FGFR2, FGFR3, FGFRS, and VEGFR1
 - Comparisons between FGFRS and FGFR2, FGFR3, and VEGFR1
 - Comparisons between FGFR2, FGFR3 and VEGFR1
 - Notably, included within the comparisons that are significantly different (all $p \ll 0.001$) are comparisons between each TMD and CD28, excluding Valine.

Supplementary Note A3.2

Below are the outcomes from two-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of PC C-terminal peptide or PC inner linker (PCIL), ligand treatment, or their interaction on the measured reporter expression.

Functional signaling assay in Fig. 3.1c

- PC C-terminal peptide $p = 1.6 \times 10^{-33}$
 - All differences were significant (all $p < 0.01$) except for AIP vs AIP_K and all comparisons between PRS-containing samples (all $p > 0.9$).
- Treatment $p = 1.4 \times 10^{-37}$
 - The samples with no PC C-terminal peptide and with any AIP PC C-terminal peptide show a significant increase in reporter expression upon ligand treatment (all $p < 0.01$).
 - The samples with any PRS PC C-terminal peptide show no significant increase in reporter expression upon ligand treatment (all $p = 1$).
- Interaction between TMD and ligand treatment $p = 3.6 \times 10^{-31}$

Functional signaling assay in Fig. 3.1f

- PCIL $p = 1.3 \times 10^{-29}$
- Receptors with EN performed similarly to receptors with RKMK, RN, RR, and RRLI but significantly differently than all other receptors ($p < 0.05$). Treatment $p = 2.5 \times 10^{-34}$
 - All PCILs produced significant increases in reporter expression upon ligand treatment (all $p < 0.01$)
 - Compared to the base case EN PCIL:
 - RR(R), RN, and RRLI show no significant change to background signal ($p > 0.05$) and either no change ($p > 0.05$) or a significant decrease ($p < 0.05$) in ligand-induced signal.
 - RRR, RRRR, RKMK, and KMKS show a significant increase to background signal ($p < 0.01$) and either no change ($p > 0.05$) or a significant increase to ligand-induced signal ($p < 0.01$).
 - RR shows a significant increase to background signal ($p < 0.01$) and a significant decrease to ligand-induced signal ($p < 0.01$).
- Interaction between TMD and ligand treatment $p = 3.5 \times 10^{-11}$

Below are the outcomes from two-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of TMD, ligand treatment, or their interaction on the measured reporter expression.

Functional signaling assay in Fig. 3.2c

- TMD $p = 1.7 \times 10^{-32}$
 - The following TMDs performed significantly differently than the CD28 TMD: FGFR1, FGFR2, FGFR3, FGFR4, FGFR-S, VEGFR1 (all $p < 0.01$).
 - The GpA ($p = 0.13$) and Valine ($p = 0.93$) TMDs did not perform significantly differently than the CD28 TMD.
 - No vehicle-only conditions were different than the CD28-TMD vehicle-only condition (all $p > 0.3$).
 - Of the ligand-treated conditions, only the Valine-TMD ligand ($p = 0.6$) condition was not significantly different from the CD28-TMD ligand-treated condition ($p < 0.01$ for all conditions except GpA-TMD where $p = 0.015$).
- Treatment $p = 1.8 \times 10^{-31}$
 - The following TMDs did show significantly different reporter expression upon ligand treatment: CD28, GpA, FGFR1, FGFR4, Valine (all $p < 0.01$).

- The following TMDs did not show significantly different reporter expression upon ligand treatment: FGFR2, FGFR3, EphA4, VEGFR1 (all $p = 1$).
- Interaction between TMD and ligand treatment $p = 1.5 \times 10^{-30}$

Below are the outcomes from two-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of TMD, ligand treatment, or their interaction on the measured NFRET values.

NFRET assay in **Fig. 3.4a**

- TMD $p = 5.2 \times 10^{-10}$
 - All differences were significant (all $p \ll 0.01$).
- Treatment $p = 1.6 \times 10^{-20}$
 - Vehicle vs. ligand was significant (all $p \ll 0.01$).
- Interaction between TMD and ligand treatment $p = 8.9 \times 10^{-6}$

NFRET assay in **Fig. 3.4b**

- TMD $p = 1.6 \times 10^{-32}$
 - All differences were significant (all $p \ll 0.01$).
- Ligand treatment dose $p = 3.3 \times 10^{-66}$
 - All differences were significant except for 50 vs. 100 ($p = 0.15$), 50 vs. 1000 ($p = 0.06$), 100 vs. 500 ($p = 0.99$), 100 vs. 1000 ($p = 1$), and 500 vs. 1000 ($p = 1$).
- Interaction between TMD and ligand treatment $p = 7.0 \times 10^{-12}$

Below are the outcomes from two-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of position of the first dimerizing residue (from the inner linker) in the synthetic TMDs, ligand treatment, or their interaction on the measured reporter expression.

Functional signaling assay in **Fig. 3.5b**

- Position of first dimerizing residue $p = 1.9 \times 10^{-14}$
 - All differences were significant (all $p < 0.05$) except for:
 - Comparisons between positions 10-15
 - Positions 4 and 6
 - Positions 8 and 9
- Treatment $p = 2.3 \times 10^{-5}$
 - Only TMDs with the first dimerizing residue in positions 4-9 produced significant increases in reporter expression upon ligand treatment (all $p \ll 0.01$)
 - Moving the TMD from positions 4-5, 5-6, 6-7, 7-8, 9-10 resulted in significant differences in ligand-induced reporter expression (all $p \ll 0.01$)
- Interaction between TMD and ligand treatment $p = 1.0 \times 10^{-4}$

Below are the outcomes from two-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of TC TMD, PC TMD, or their interaction on the measured reporter expression, evaluated separately for vehicle- and ligand-treated conditions.

Functional signaling assay in **Fig. 3.2d, Supplementary Fig. 3.5c**

Vehicle treatment

- TC TMD $p = 1.6 \times 10^{-37}$
 - All differences were significant (all $p \ll 0.01$) except for comparisons between FGFR4, Valine, and CD28.
- PC TMD $p = 3.8 \times 10^{-17}$
 - All differences were significant (all $p \ll 0.01$) except for:

- Comparisons between FGFR1, Valine, and CD28
- FGFR4 vs. CD28
- Interaction between TC TMD and PC TMD $p = 3.0 \times 10^{-20}$
- Variance explained by TC TMD $\omega^2 = 0.64$
- Variance explained by PC TMD $\omega^2 = 0.10$
- Variance explained by the interaction $\omega^2 = 0.22$

Ligand treatment

- TC TMD $p = 9.8 \times 10^{-53}$
 - All differences were significant (all $p \ll 0.01$).
- PC TMD $p = 1.9 \times 10^{-44}$
 - All differences were significant (all $p \ll 0.01$).
- Interaction between TC TMD and PC TMD $p = 6.9 \times 10^{-33}$
- Variance explained by TC TMD $\omega^2 = 0.58$
- Variance explained by PC TMD $\omega^2 = 0.27$
- Variance explained by the interaction $\omega^2 = 0.14$

Functional signaling assays in **Fig. 3.6a**

Vehicle treatment

- TC TMD $p = 1.7 \times 10^{-33}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR4 ($p = 0.97$).
- PC TMD $p = 7.3 \times 10^{-29}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR1 ($p = 0.99$).
- Interaction between TC TMD and PC TMD $p = 2.1 \times 10^{-22}$
- Variance explained by TC TMD $\omega^2 = 0.45$
- Variance explained by PC TMD $\omega^2 = 0.28$
- Variance explained by the interaction $\omega^2 = 0.24$

Ligand treatment

- TC TMD $p = 8.8 \times 10^{-30}$
 - All differences were significant (all $p \ll 0.01$).
- PC TMD $p = 2.1 \times 10^{-19}$
 - All differences were significant (FGFR1 vs. CD28 $p = 0.04$, all other $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.16$).
- Interaction between TC TMD and PC TMD $p = 2.7 \times 10^{-9}$
- Variance explained by TC TMD $\omega^2 = 0.63$
- Variance explained by PC TMD $\omega^2 = 0.21$
- Variance explained by the interaction $\omega^2 = 0.10$

Functional signaling assays in **Fig. 3.6b**

Vehicle treatment

- TC TMD $p = 6.2 \times 10^{-47}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR4 ($p = 0.61$).
- PC TMD $p = 2.7 \times 10^{-44}$
 - All differences were significant (Valine vs. FGFR1 $p = 0.02$, all other $p \ll 0.01$) except for FGFR4 vs. CD28 ($p = 0.26$).
- Interaction between TC TMD and PC TMD $p = 6.2 \times 10^{-33}$
- Variance explained by TC TMD $\omega^2 = 0.45$
- Variance explained by PC TMD $\omega^2 = 0.35$
- Variance explained by the interaction $\omega^2 = 0.19$

Ligand treatment

- TC TMD $p = 2.6 \times 10^{-51}$
 - All differences were significant (Valine vs. FGFR4 $p = 0.03$, all other $p \ll 0.01$).
- PC TMD $p = 5.9 \times 10^{-39}$

- All differences were significant (all $p \ll 0.01$).
- Interaction between TC TMD and PC TMD $p = 7.0 \times 10^{-23}$
- Variance explained by TC TMD $\omega^2 = 0.70$
- Variance explained by PC TMD $\omega^2 = 0.22$
- Variance explained by the interaction $\omega^2 = 0.07$

Functional signaling assays in **Fig. 3.6c**

Vehicle treatment

- TC TMD $p = 7.9 \times 10^{-55}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR4 ($p = 1$).
- PC TMD $p = 5.6 \times 10^{-46}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR1 ($p = 0.41$).
- Interaction between TC TMD and PC TMD $p = 8.5 \times 10^{-33}$
- Variance explained by TC TMD $\omega^2 = 0.61$
- Variance explained by PC TMD $\omega^2 = 0.27$
- Variance explained by the interaction $\omega^2 = 0.12$

Ligand treatment

- TC TMD $p = 2.1 \times 10^{-50}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR4 ($p = 0.97$).
- PC TMD $p = 3.3 \times 10^{-51}$
 - All differences were significant (all $p \ll 0.01$).
- Interaction between TC TMD and PC TMD $p = 3.8 \times 10^{-26}$
- Variance explained by TC TMD $\omega^2 = 0.44$
- Variance explained by PC TMD $\omega^2 = 0.48$
- Variance explained by the interaction $\omega^2 = 0.07$

Functional signaling assays in **Fig. 3.6d**

Vehicle treatment

- TC TMD $p = 7.2 \times 10^{-36}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR1 ($p = 0.99$).
- PC TMD $p = 1.2 \times 10^{-39}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.78$).
- Interaction between TC TMD and PC TMD $p = 6.3 \times 10^{-23}$
- Variance explained by TC TMD $\omega^2 = 0.34$
- Variance explained by PC TMD $\omega^2 = 0.49$
- Variance explained by the interaction $\omega^2 = 0.15$

Ligand treatment

- TC TMD $p = 4.7 \times 10^{-39}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR4 ($p = 0.14$) and FGFR1 vs. CD28 ($p = 1$).
- PC TMD $p = 7.1 \times 10^{-41}$
 - All differences were significant (all $p < 0.05$) except for GpA vs. FGFR4 ($p = 0.19$) and FGFR1 vs. CD28 ($p = 0.71$).
- Interaction between TC TMD and PC TMD $p = 8.7 \times 10^{-21}$
- Variance explained by TC TMD $\omega^2 = 0.40$
- Variance explained by PC TMD $\omega^2 = 0.48$
- Variance explained by the interaction $\omega^2 = 0.11$

Below are the outcomes from two-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of CTEVp TMD, NTEVp TMC, or their interaction on the measured reporter expression in the presence of either ligand or vehicle.

Functional signaling assays in **Fig. 3.6e**

Vehicle treatment

- CTEVp TMD $p = 7.1 \times 10^{-9}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.08$).
- NTEVp TMD $p = 4.6 \times 10^{-11}$
 - All differences were significant except for GpA vs. FGFR1 ($p = 0.36$) and FGFR4 vs. FGFR1 ($p = 0.31$).
- Interaction between CTEVp TMD and NTEVp TMD $p = 1.8 \times 10^{-22}$
- Variance explained by CTEVp TMD $\omega^2 = 0.06$
- Variance explained by NTEVp TMD $\omega^2 = 0.09$
- Variance explained by the interaction $\omega^2 = 0.82$

Ligand treatment

- CTEVp TMD $p = 9.3 \times 10^{-8}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. CD28 ($p = 0.86$), GpA vs. FGFR4 ($p = 0.24$), and FGFR4 vs. CD28 ($p = 0.051$).
- NTEVp TMD $p = 4.1 \times 10^{-6}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.54$), GpA vs. FGFR1 ($p = 1$), and FGFR4 vs. FGFR1 ($p = 0.56$).
- Interaction between CTEVp TMD and NTEVp TMD $p = 3.5 \times 10^{-3}$
- Variance explained by CTEVp TMD $\omega^2 = 0.35$
- Variance explained by NTEVp TMD $\omega^2 = 0.23$
- Variance explained by the interaction $\omega^2 = 0.13$

Functional signaling assays in **Fig. 3.6f**

Vehicle treatment

- CTEVp TMD $p = 3.0 \times 10^{-24}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.75$) and FGFR4 vs. FGFR1 ($p = 0.06$).
- NTEVp TMD $p = 8.1 \times 10^{-22}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR4 vs. FGFR1 ($p = 0.98$).
- Interaction between CTEVp TMD and NTEVp TMD $p = 1.4 \times 10^{-29}$
- Variance explained by CTEVp TMD $\omega^2 = 0.20$
- Variance explained by NTEVp TMD $\omega^2 = 0.14$
- Variance explained by the interaction $\omega^2 = 0.65$

Ligand treatment

- CTEVp TMD $p = 9.5 \times 10^{-23}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.08$)
- NTEVp TMD $p = 7.5 \times 10^{-18}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR4 vs. FGFR1 ($p = 0.30$) and GpA vs. FGFR1 ($p = 0.36$).
- Interaction between CTEVp TMD and NTEVp TMD $p = 4.7 \times 10^{-13}$
- Variance explained by CTEVp TMD $\omega^2 = 0.54$
- Variance explained by NTEVp TMD $\omega^2 = 0.25$
- Variance explained by the interaction $\omega^2 = 0.17$

Supplementary Note A3.3

Below are the outcomes from three-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of TC TMD, PC TMD, ligand treatment, or their pairwise interactions on the measured NFRET values or reporter expression.

Functional signaling assays in **Fig. 3.2d**

- TC TMD $p = 2.5 \times 10^{-92}$
 - All differences were significant (all $p \ll 0.01$) except for Valine vs. FGFR4 ($p = 0.21$).
- PC TMD $p = 2.8 \times 10^{-75}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR1 vs. CD28 ($p = 0.06$), FGFR4 vs. FGFR1 ($p = 0.83$), and Valine vs. GpA ($p = 0.10$).
- Ligand treatment $p = 2.6 \times 10^{-124}$
 - Vehicle vs. ligand was significant for all TMD pairs (all $p \ll 0.01$).
- Interaction between TC TMD and PC TMD $p = 8.9 \times 10^{-56}$
- Interaction between TC TMD and ligand treatment $p = 8.6 \times 10^{-89}$
- Interaction between PC TMD and ligand treatment $p = 5.0 \times 10^{-73}$
- Interaction between TC TMD, PC TMD, and ligand treatment $p = 4.2 \times 10^{-54}$

Functional assays in **Supplementary Fig. A3.5**

- TC TMD $p = 2.3 \times 10^{-169}$
- PC TMD $p = 1.9 \times 10^{-93}$
- Ligand $p = 2.1 \times 10^{-137}$
 - Significant ligand-induced signaling was observed for all receptor pairs that included a CD28-TMD on the TC (all $p \ll 0.01$).
 - Significant ligand-induced signaling was not observed for any receptor pairs that included an EphA4, FGFR2, or FGFR3-TMD on the TC (all $p > 0.15$).
 - Significant ligand-induced signaling was observed for receptor pairs that included a VEGFR1-TMD on the TC and a either a CD28, GpA, FGFR1, FGFR3, FGFR4, or Valine-TMD on the PC (all $p \ll 0.02$).
 - Significant ligand-induced signaling was not observed for all receptor pairs that included a VEGFR1-TMD on the TC and either a FGFR2, EphA4, or VEGFR1-TMD on the PC (all $p > 0.5$).
- Interaction between TC TMD and PC TMD $p = 3.2 \times 10^{-115}$
- Interaction between TC TMD and ligand treatment $p = 1.7 \times 10^{-167}$
- Interaction between PC TMD and ligand treatment $p = 6.6 \times 10^{-92}$
- Interaction between TC TMD, PC TMD, and ligand treatment $p = 6.5 \times 10^{-114}$

NFRET assays in **Fig. 3.4c**

- FRB chain TMD $p = 1.0 \times 10^{-12}$
 - All differences were significant (all $p \ll 0.01$).
- FKBP chain TMD $p = 3.9 \times 10^{-15}$
 - All differences were significant (all $p \ll 0.01$).
- Ligand treatment $p = 6.0 \times 10^{-59}$
 - Vehicle vs. ligand was significant for all TMD pairs (all $p \ll 0.01$).
- Interaction between FRB chain TMD and FKBP chain TMD $p = 5.2 \times 10^{-29}$
- Interaction between FRB chain TMD and ligand treatment $p = 5.5 \times 10^{-11}$
- Interaction between FKBP chain TMD and ligand treatment $p = 9.4 \times 10^{-11}$
- Interaction between FRB chain TMD, FKBP chain TMD, and ligand treatment $p = 1.1 \times 10^{-10}$

Functional signaling assays in **Fig. 3.6a**

- TC TMD $p = 2.1 \times 10^{-50}$
 - All differences were significant (all $p \ll 0.01$).

- PC TMD $p = 3.4 \times 10^{-31}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR1 vs. CD28 ($p = 0.14$). and GpA vs. FGFR4 ($p = 0.83$).
- Ligand treatment $p = 2.2 \times 10^{-74}$
 - Vehicle vs. ligand was significant for all pairs (TC CD28-TMD with PC-CD28 TMD $p = 0.03$, all other $p \ll 0.01$) except for TC FGFR4-TMD with PC CD28-TMD ($p = 0.17$), TC Valine-TMD with PC FGFR1-TMD ($p = 1$), TC CD28-TMD with PC Valine-TMD ($p = 0.73$), TC FGFR4-TMD with PC Valine-TMD ($p = 0.96$), TC Valine-TMD with PC Valine-TMD ($p = 1$).
- Interaction between TC TMD and PC TMD $p = 2.0 \times 10^{-13}$
- Interaction between TC TMD and ligand treatment $p = 3.3 \times 10^{-41}$
- Interaction between PC TMD and ligand treatment $p = 2.0 \times 10^{-23}$
- Interaction between TC TMD, PC TMD, and ligand treatment $p = 2.0 \times 10^{-12}$

Functional signaling assays in **Fig. 3.6b**

- TC TMD $p = 9.5 \times 10^{-99}$
 - All differences were significant (all $p \ll 0.01$).
- PC TMD $p = 6.5 \times 10^{-81}$
 - All differences were significant (all $p \ll 0.01$).
- Ligand treatment $p = 3.8 \times 10^{-66}$
 - Vehicle vs. ligand was significant for all pairs (all $p \ll 0.01$) except for TC FGFR4-TMD with PC CD28-TMD ($p = 0.97$), TC Valine-TMD with PC CD28-TMD ($p = 1$), TC CD28-TMD with PC FGFR1-TMD ($p = 0.84$), TC FGFR4-TMD with PC FGFR1-TMD ($p = 1$), TC GpA-TMD with PC FGFR1-TMD ($p = 0.68$), TC Valine-TMD with PC FGFR1-TMD ($p = 1$), TC Valine-TMD with PC FGFR4-TMD ($p = 1$), TC CD28-TMD with PC Valine-TMD ($p = 1$), TC FGFR4-TMD with PC Valine-TMD ($p = 1$), TC Valine-TMD with PC Valine-TMD ($p = 1$).
- Interaction between TC TMD and PC TMD $p = 4.3 \times 10^{-53}$
- Interaction between TC TMD and ligand treatment $p = 5.4 \times 10^{-50}$
- Interaction between PC TMD and ligand treatment $p = 1.3 \times 10^{-25}$
- Interaction between TC TMD, PC TMD, and ligand treatment $p = 2.3 \times 10^{-27}$

Functional signaling assays in **Fig. 3.6c**

- TC TMD $p = 1.5 \times 10^{-102}$
 - All differences were significant (all $p \ll 0.01$) except for Val vs. FGFR4 ($p = 0.99$).
- PC TMD $p = 1.6 \times 10^{-97}$
 - All differences were significant (all $p \ll 0.01$) except for Val vs. FGFR1 ($p = 0.16$).
- Ligand treatment $p = 7.2 \times 10^{-79}$
 - Vehicle vs. ligand was significant for all pairs (all $p \ll 0.01$) except for TC FGFR4-TMD with PC CD28-TMD ($p = 1$), TC Valine-TMD with PC CD28-TMD ($p = 0.75$), TC CD28-TMD with PC FGFR1-TMD ($p = 0.10$), TC FGFR4-TMD with PC FGFR1-TMD ($p = 1$), TC GpA-TMD with PC FGFR1-TMD ($p = 0.98$), TC Valine-TMD with PC FGFR1-TMD ($p = 1$), TC Valine-TMD with PC FGFR4-TMD ($p = 0.08$), TC CD28-TMD with PC Valine-TMD ($p = 0.95$), TC FGFR4-TMD with PC Valine-TMD ($p = 1$), TC Valine-TMD with PC Valine-TMD ($p = 1$).
- Interaction between TC TMD and PC TMD $p = 4.7 \times 10^{-56}$
- Interaction between TC TMD and ligand treatment $p = 3.3 \times 10^{-41}$
- Interaction between PC TMD and ligand treatment $p = 2.4 \times 10^{-59}$
- Interaction between TC TMD, PC TMD, and ligand treatment $p = 9.2 \times 10^{-27}$

Functional signaling assays in **Fig. 3.6d**

- TC TMD $p = 4.2 \times 10^{-72}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR1 vs. CD28 ($p = 0.12$).

- PC TMD $p = 1.0 \times 10^{-75}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR4 vs. CD28 ($p = 0.49$) and GpA vs. FGFR4 ($p = 0.10$).
- Ligand treatment $p = 2.2 \times 10^{-98}$
 - Vehicle vs. ligand was significant for all pairs (all $p \ll 0.01$) except for TC CD28-TMD with PC Valine-TMD ($p = 0.31$), TC FGFR4-TMD with PC Valine-TMD ($p = 1$), TC Valine-TMD with PC Valine-TMD ($p = 1$)
- Interaction between TC TMD and PC TMD $p = 1.7 \times 10^{-39}$
- Interaction between TC TMD and ligand treatment $p = 7.3 \times 10^{-49}$
- Interaction between PC TMD and ligand treatment $p = 1.7 \times 10^{-53}$
- Interaction between TC TMD, PC TMD, and ligand treatment $p = 9.1 \times 10^{-23}$

Below are the outcomes from three-way ANOVAs and Tukey's HSD tests. Null hypotheses were that there existed no effects of NTEVp TMD, CTEVp TMD, ligand treatment, or their pairwise interactions on the measured reporter expression.

Functional signaling assays in **Fig. 3.6e**

- CTEVp TMD $p = 3.0 \times 10^{-10}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR4 vs. CD28 ($p = 0.20$), GpA vs. CD28 ($p = 1$), and GpA vs. FGFR4 ($p = 0.20$).
- NTEVp TMD $p = 1.2 \times 10^{-5}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR4 vs. FGFR1 ($p = 0.91$), GpA vs. FGFR1 ($p = 1$), GpA vs. FGFR4 ($p = 0.97$).
- Ligand treatment $p = 1.2 \times 10^{-66}$
 - Vehicle vs. ligand was significant for all TMD pairs (all $p \ll 0.01$).
- Interaction between CTEVp TMD and NTEVp TMD $p = 2.1 \times 10^{-7}$
- Interaction between CTEVp TMD and ligand treatment $p = 4.2 \times 10^{-9}$
- Interaction between NTEVp TMD and ligand treatment $p = 5.7 \times 10^{-9}$
- Interaction between CTEVp TMD, NTEVp TMD, and ligand treatment $p = 1.0 \times 10^{-2}$

Functional signaling assays in **Fig. 3.6f**

- CTEVp TMD $p = 1.8 \times 10^{-44}$
 - All differences were significant (all $p \ll 0.01$) except for GpA vs. FGFR4 ($p = 0.27$).
- NTEVp TMD $p = 3.3 \times 10^{-17}$
 - All differences were significant (all $p \ll 0.01$) except for FGFR4 vs. FGFR1 ($p = 0.31$).
- Ligand treatment $p = 3.9 \times 10^{-83}$
 - Vehicle vs. ligand was significant for all TMD pairs (all $p \ll 0.01$).
- Interaction between CTEVp TMD and NTEVp TMD $p = 7.9 \times 10^{-40}$
- Interaction between CTEVp TMD and ligand treatment $p = 1.0 \times 10^{-17}$
- Interaction between NTEVp TMD and ligand treatment $p = 9.8 \times 10^{-35}$
- Interaction between CTEVp TMD, NTEVp TMD, and ligand treatment $p = 2.6 \times 10^{-17}$

Appendix 4. Supplementary information for Chapter 4

Supplementary Table A4.1 Filters for microscopy

Name	Product#	Filter 1	Filter 2	Filter3
ET-EBFP2/Coumarin/Attenuated DAPI	49021	ET405/20x	ET460/50m	T425lpxr-UF1
ET - EYFP	49003	ET500/20x	ET535/30m	T515lp-UF1
ET - DSRed (TRITC/Cy3)	49005	ET545/30x	ET620/60m	T570lp-UF1
ET Cy5.5	49022	ET650/45x	ET720/60m	T685lpxr-UF1

Supplementary Table A4.2 Filters for flow cytometry analysis

Instrument	Fluorescent protein	Parameter/ Channel name	Excitation laser	Filter set
BD LSR Fortessa	EBFP2, mTagBFP2	Pacific Blue	Violet, 405 nm	450/50
	EYFP, mNeonGreen	FITC	Blue, 488 nm	505LP, 530/30
	DsRed-Express2, LumiScarlet	PE-Texas Red	Light Green, 552 nm	600LP, 610/20

Supplementary Table A4.3 Filters for flow cytometry sorting

Instrument	Fluorescent protein	Parameter	Excitation laser	Filter set
BD FACS Aria	EBFP2, mTagBFP2	Pacific Blue	Violet 407 nm	450/50
	EYFP	FITC	Blue 488 nm	505LP, 525/30

Appendix 5. Supplementary information for Chapter 5

Contents:

Supplementary Tables A5.1-5.2

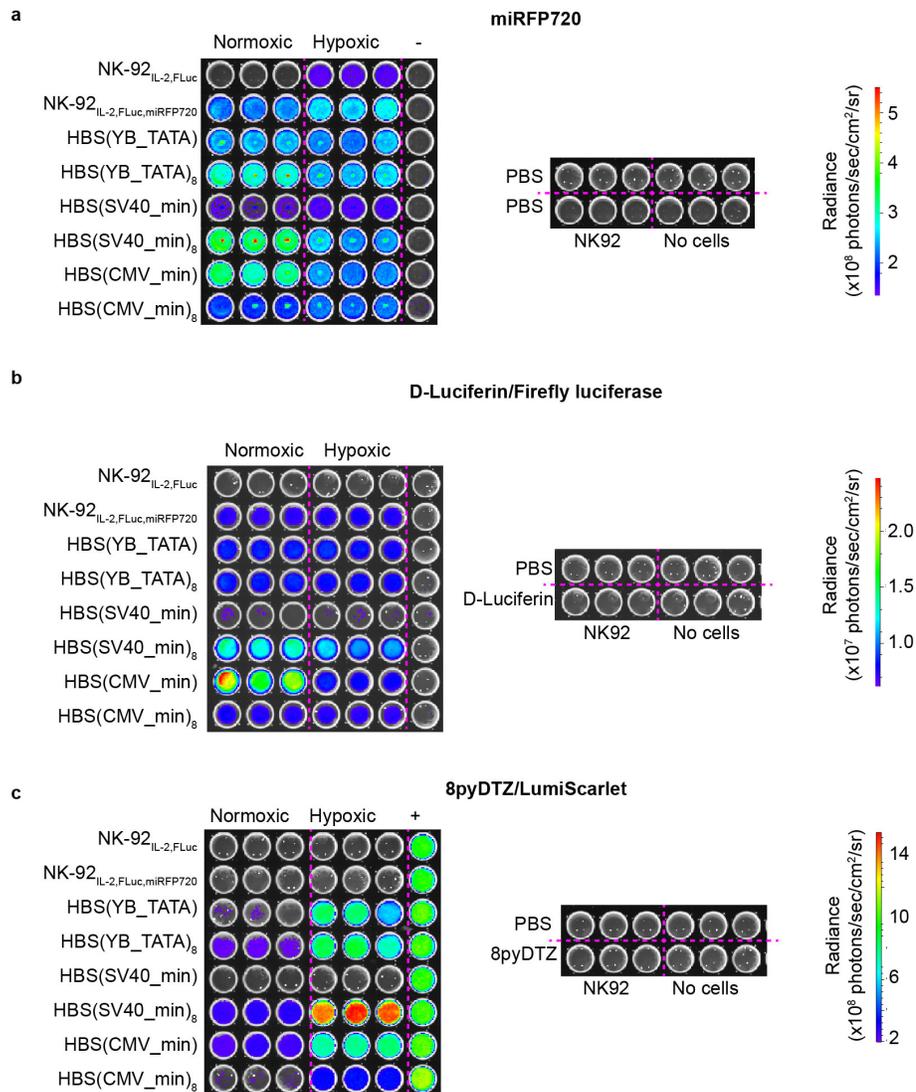
Supplementary Figure A5.1

Supplementary Table A5.1 Filters for flow cytometry sorting

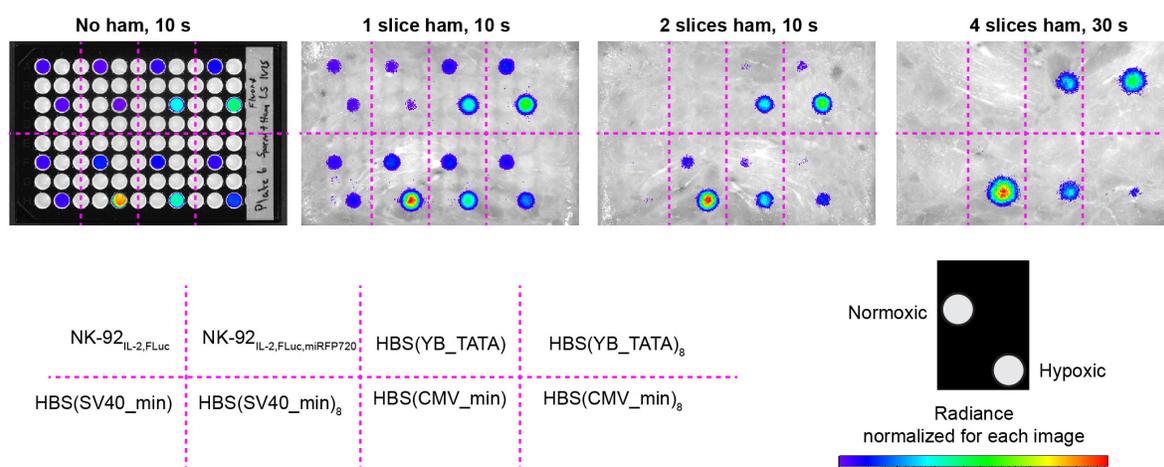
Instrument	Fluorescent protein or dye	Parameter	Excitation laser	Filter set
BD FACS Aria	miRFP720	ACP-Cy7	Red 633 nm	690 LP, 730/45
	DAPI	DAPI	UV 355 nm	450/50

Supplementary Table A5.2 Filters for flow cytometry analysis

Fluorophore	Parameter name	Excitation laser	Emission filters
DAPI	DAPI	UV 355 nm	450/50
mTagBFP2	Pacific Blue	Violet, 405 nm	450/50
mNeonGreen, EYFP	FITC	Blue, 488 nm	505LP, 530/30
cyOFP1 (in Antares2)	PerCP	Blue, 488 nm	550LP, 585/42
mScarlet-I (in LumiScarlet), DsRed-Express2	PE-Texas Red	Light green, 552 nm	600LP, 610/20
miRFP670	APC	Red, 640 nm	670/30
miRFP720	Alexa 750	Far Red, 685 nm	690LP, 730/45
Tested, not used	Alexa 700	Red, 640 nm	685,LP 740/35
7-AAD	PE-Cy7	Light green, 552 nm	750LP, 760/80
Propidium Iodide	PE-Texas Red	Light green, 552 nm	600LP, 610/20
Cell Trace Violet	Pacific Blue	Violet, 405 nm	450/50



Supplementary Fig. A5.1 Comparison of performance of NK-92 HBS lines with various minimal promoters by imaging on IVIS Spectrum. Cells were imaged after 3-4 days of hypoxic or normoxic culture for a) fluorescence signal from miRFP720 b) signal from firefly luciferase after D-luciferin addition c) signal from LumiScarlet BRET reporter after 8pyDTZ addition. The right column in a-c contains a K562 line transduced with LumiScarlet to control for slight difference in the timing of substrate addition to each well.



Supplementary Fig. A5.2 Imaging of HBS activation through biologic tissue. Samples from Fig. A.5.1 were separated into wells on a new 96-well plate and substrate added. Plate was imaged under an increasing number of ham slices. Under 4 slices of ham, the response of the NK-92 cells to hypoxia was visible with a 30s exposure time. Radiance was normalized within each image and is not comparable across images.