Fine tuning gene expression levels in mammalian cells with engineered microRNA target sites

Yale S. Michaels
New College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Fine tuning gene expression levels in mammalian cells with engineered microRNA target sites

Yale S. Michaels, New College

A thesis submitted for the degree of Doctor of Philosophy

Trinity term, 2018

Abstract

Precise, analogue regulation of gene expression is critical for development, homeostasis and regeneration in mammals. In contrast, widely employed experimental and therapeutic approaches such as knock-in/out strategies are more suitable for binary control of gene activity, while RNA interference (RNAi) can lead to pervasive off-target effects and unpredictable levels of repression. In this thesis, I report on a method for precise control of gene expression levels in mammalian cells based on engineered, synthetic microRNA response elements (MREs). To develop this system, we first tested a small panel of rationally designed MREs with various mismatches to miR-17 or miR-21. These MRE variants exerted predictable, stepwise control over reporter gene expression in a human cell line. Next, to improve the precision of MRE-based gene tuning, we established a high-throughput sequencing approach for measuring the efficacy of thousands of miR-17 MRE variants in parallel. This allowed us to create a library of microRNA silencing-mediated fine-tuners (miSFITs) of varying strength that can be employed to control the expression of user specified genes. A selection of miSFITs enabled precise, predictable tuning of PD-1, a co-inhibitory receptor expressed on T-cells and an important target for anti-tumour immunotherapy. In addition to tuning a PD-1 transgene in Jurkat T-cells, we successfully introduced miSFITs into the endogenous PD-1 3'UTR in embryonic stem cells using the CRISPR/Cas9 system. To further demonstrate the value of this technology, we also used a panel of miSFITs to tune the expression of a tumour-associated antigen in a mouse melanoma model. This analysis revealed that antigen expression level is a key determinant of the anti-tumour immune response *in vitro* and *in vivo*. miSFITs are a powerful tool for modulating gene expression levels with applications in research and cellular engineering.

Acknowledgments

I would like to thank all current and former members of the Fulga lab for their support and guidance over the course of my DPhil. I acknowledge Qianxin Wu for embodying perseverance and determination, Markus Toegel who always puts others before himself, Aron Szabo whose humour and spirit are contagious and David Knapp who is immune to failed experiments and physical pain. I thank Tim Rajamkumar, for teaching me the techniques of molecular biology and FIFA, Bruno Steinkraus for showing me how to overcome the pressures of a DPhil in style and Quentin Ferry for sharing his many talents. A special mention goes to Toni Baeumler who taught me about work ethic and patience.

I am grateful to Alice Lightowlers and Max Jamilly who, between mastering every sport, craft and language, found time to provide astute feedback on this thesis. I thank Mike Barnkob for teaching me immunology and for all of his technical and intellectual assistance. I also acknowledge the newest group of helpful and enthusiastic DPhil students to join the lab: Andrew Ramos, Clara Pavillet, Muhammad Hanifi and Oana Pelea. I thank covert biologist Nina Yancy for her useful suggestions.

I would like to thank the students who I have had the pleasure of supervising. I am grateful to Hector Barbosa who exceeded all expectations and made instrumental contributions during his time in the lab. I also acknowledge Natalie Beadle and Elena Zanchini di Castiglionchio who have recently joined Tudor's group and have already become master cloners.

I am thankful for my friends and family, especially my parents and brothers who have supported me throughout this degree and my grandmother Joan Michaels who swiftly and skilfully proof-read the introduction to this thesis. I thank Allison

Marmel for her feedback on the thesis and for sharing the ups and downs of the entire process.

I am grateful to the scientific mentors who have supported me throughout my career. Prof. Gunnar Valdimarsson introduced me to the world of scientific research. Prof. Geoff Hicks has generously shared his time, advice and wisdom with me for more than a decade. Noam Prywes was an outstanding supervisor and has continued to be a valued mentor and friend.

I am grateful for thoughtful feedback and advice over the past four years from Prof. Tatjana Sauka-Spengler and Prof. Ahmed Ahmed. My thanks also go to Prof. Enzo Cerundolo for generously sharing his expertise and resources and for acting as a one-man thesis committee.

Finally, I thank my DPhil supervisors for this challenging and fulfilling experience. I thank Prof. Tom Milne for stimulating scientific discussions, words of encouragement when they were needed and for insightful commentary on the world of science and beyond. I am grateful to Prof. Tudor Fulga for bringing me to Oxford, for sharing his creativity and enthusiasm for science, for giving me the freedom to explore my interests and for motivating me to do my best.

The research presented here was made possible by financial support from the Christopher Welch Scholarship, the WIMM Prize Studentship and the Clarendon Scholarship.

List of abbreviations

ACT	Adoptive T-cell therapy
AGO	Argonaut
APC	Antigen presenting cell
CAR	Chimeric antigen receptor
Cas9	CRISPR associated protein 9
CD	Cluster of differentiation
cDNA	Complementary DNA
CHX	Cycloheximide
CMV	Cytomegalovirus
CRIPSRa	CRISPR activation
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR interference
Ct	Cycle of threshold
CTCF	CCCTC-binding factor
CTLA-4	Cytotoxic T-lymphocyte associated protein 4
dCas9	Dead Cas9
ddPCR	Digital droplet PCR
DNA	Deoxyribonucleic acid
ECFP	Enhanced cyan fluorescent protein
EGFP	Enhanced green fluorescent protein
EYFP	Enhanced yellow fluorescent protein
FACS	Fluorescence activated cell sorting
HDR	Homology directed repair
HR	Homologous recombination
HTS	High-throughput sequencing
Indel	Insertions and/or deletions
IVT	In vitro transcribed
Km	Substrate concentration at half maximal reaction velocity
mESC	Mouse embryonic stem cell
MFI	Mean fluorescent intensity
MHC	Major histocompatibility complex
miRISC	miRNA induced RNA silencing complex
miRNA	microRNA
miSFIT	miRNA silencing-mediated fine-tuner
MOI	Multiplicity of infection
MRE	miRNA response element
mRNA	Messenger RNA
NHEJ	Non-homologous end joining
ORF	Open reading frame
OVA	Ovalbumin
PCR	Polymerase chain reaction

PD-1	Programmed cell death 1
PD-L1	Programmed death-ligand 1
pDNA	Plasmid DNA
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
PSC	Pluripotent stem cell
Puro	Puromycin resistance cassette
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNP	Ribonucleoprotein
rtTA	Reverse tetracycline-controlled transactivator
sgRNA	Synthetic guide RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
ssODN	Single stranded oligonucleotide homology donor
TCR	T-cell receptor
TIL	Tumour infiltrating lymphocyte
tRNA	Transfer RNA
uORF	Upstream open reading frame
UTR	Untranslated region
Vmax	Maximum reaction velocity

Declaration of authorship

This thesis, and the work presented herein is my own unless otherwise specified. Some of the text and figures used in the thesis are adapted from a manuscript posted on the preprint server Biorxiv, of which I am the lead author[1]. This work was carried out primarily at the Weatherall Institute of Molecular Medicine, University of Oxford. I am submitting this thesis for the degree of DPhil in Medical Sciences. This thesis has not been submitted for any other degree.

The following experiments were conducted in collaboration with others:

1.) I cloned the MRE variant libraries used to generate Fig. 4.2 and Fig. 5.2 with help from Hector Barbosa (Weatherall Institute of Molecular Medicine, University of Oxford).

2.) I performed the polysome profiling experiments presented in Fig. 4.4 with help from Mary K. Thompson (Dept. of Biochemistry, University of Oxford).

3.) Hector Barbosa also assisted with the HTS validation experiments summarised in Fig. 4.5.

4.) Ben Davies and Daniel Biggs (Wellcome Trust Centre for Human Genetics, University of Oxford) provided guidance and assistance with mESC genome engineering. This data is shown in Fig. 5.3.

5.) *In vivo* tumour growth experiments presented in Fig. 6.3 and Fig. 6.4 were performed in collaboration with Mike Barnkob (MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford).

6.) Aron Szabo (Weatherall Institute of Molecular Medicine, University of Oxford) assited with confocal microscopy shown in Fig. 3.2 and Fig. 3.4.

Title and abstract of works published/patented during this thesis

1. **Michaels, Y.S**., Barnkob, M.B., Barbosa, H., Baeumler, T.A., Thompson, M.K., Andre, V., Colin-York, H., Fritzsche, M., Gileadi, U., Sheppard, H.M., et al. (2018). Precise tuning of gene expression output levels in mammalian cells. bioRxiv.

The work presented in this manuscript is described in detail in this thesis[1].

2. **Michaels, Y.S**., Wu, Q., and Fulga, T.A. (2017). Interrogation of Functional miRNA-Target Interactions by CRISPR/Cas9 Genome Engineering. Methods Mol Biol 1580, 79-97.

"Post-transcriptional silencing by microRNAs (miRNAs) is a critical constituent of eukaryotic gene regulation. miRNAs are short (~22nt) noncoding RNAs capable of specifically targeting the miRNA-induced-silencing-complex (miRISC) to transcripts bearing a complementary miRNA response element (MRE). Although recent methodological advances have greatly improved our understanding of miRNA biogenesis and the mechanisms by which miRNAs repress their cognate targets, exploring the physiological relevance of direct miRNA-target interactions in vivo has remained an outstanding challenge. Here we describe the experimental protocol underlying a novel approach, which allows direct interrogation of specific miRNA-MRE interactions by CRISPR/Cas9-mediated genome engineering. In this instance, the CRISPR/Cas9 system is first used to catalyze homology-directed replacement of candidate MREs with molecular barcodes at endogenous loci. Subsequently, the effect of MRE mutation on transcript abundance (i.e., MRE activity) can be rapidly evaluated by routine quantitative PCR. This strategy enables functional investigation of a putative miRNA-target pair in a pool of transiently transfected cells, obviating the need for generation of clonal cell lines or transgenic animals. This protocol can be implemented in any cell line in less than 2 weeks, and can readily be scaled up for multiplex studies. To facilitate the conceptual workflow underlying this strategy, we also describe a genome-wide resource for automated design and computational evaluation of CRISPR/Cas9 guide RNAs targeting all predicted MREs in various species (miR-CRISPR).[2]"

3. Fulga, T.A., **Michaels, Y.S**., and Milne, T.A. (2018). Method for Modifying Genes. In WO/2018/011590 (Great Britain: Oxford University Innovation Limited).

"The present invention relates to method of modulating the level of expression of an endogenous gene in a cell, the method comprising inserting a heterologous microRNA (miRNA) response element (MRE) into the 3'-untranslated region (3'-UTR) of the gene. The binding of endogenous miRNAs to the MRE results in or leads to a repression of the level of expression of the gene. The invention also relates to cells and transgenic animals whose endogenous genes comprise heterologous MRE in their 3'-UTRs.[3]"

4. Knapp, D.J., **Michaels, Y.S**., Jamilly, M., Ferry, Q.R., Barbosa, H., Milne, T.A., and Fulga, T.A. (2018). Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. bioRxiv.

“Spatial/temporal control of Cas9 guide RNA expression could considerably expand the utility of CRISPR-based technologies. Current approaches based on tRNA processing offer a promising strategy but suffer from high background. Here we developed a variant screening platform to identify differential sequence determinants of human tRNA promoter and processing activities. Rational design based on the ensuing principles allowed us to engineer an improved tRNA scaffold that enabled highly specific guide RNA production from a Pol-II promoter.[4]”

5. Wu, Q., Ferry, Q.R.V., Baeumler, T.A., **Michaels, Y.S**., Vitsios, D.M., Habib, O., Arnold, R., Jiang, X., Maio, S., Steinkraus, B.R., et al. (2017). In situ functional dissection of RNA cis-regulatory elements by multiplex CRISPR-Cas9 genome engineering. Nat Commun 8, 2109.

“RNA regulatory elements (RREs) are an important yet relatively under-explored facet of gene regulation. Deciphering the prevalence and functional impact of this post-transcriptional control layer requires technologies for disrupting RREs without perturbing cellular homeostasis. Here we describe genome-engineering based evaluation of RNA regulatory element activity (GenERA), a clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 platform for in situ high-content functional analysis of RREs. We use GenERA to survey the entire regulatory landscape of a 3′UTR, and apply it in a multiplex fashion to analyse combinatorial interactions between sets of miRNA response elements (MREs), providing strong evidence for cooperative activity. We also employ this technology to probe the functionality of an entire MRE network under cellular homeostasis, and show that high-resolution analysis of the GenERA dataset can be used to extract functional features of MREs. This study provides a genome editing-based multiplex strategy for direct functional interrogation of RNA cis-regulatory elements in a native cellular environment.[5]”

Table of Contents

List of Figures and tables

Chapter 1: Introduction

In this thesis, I document my efforts to develop a tool that will allow scientists and clinicians to precisely tune gene expression levels in mammalian cells. This introductory chapter begins by describing key mechanisms that have evolved in multicellular organisms to control gene expression. Next, I illustrate the importance of gene dosage in homeostasis and disease. In doing so, I justify why a method for finely manipulating gene expression levels is valuable to the biomedical research community. I make a general case for the potential utility of a synthetic gene tuning platform, and then I discuss specific applications of such a method in the field of tumour immunology. I then address the advantages and limitations of existing strategies for manipulating gene expression, focusing on the fact that none of these methods is suitable for fine-tuning. This introduction ends with a description of microRNAs (miRNAs), a class of naturally occurring post-transcriptional tuners of gene expression. I explain how miRNAs function and how I propose to co-opt them to develop a synthetic gene expression tuning method suitable for use in mammalian cells.

1.1 Precise control of gene expression in health and disease

Two lines of evidence demonstrate the importance of gene expression levels in complex biology. First, eukaryotic organisms allocate a large proportion of their resources towards precisely controlling gene expression through time and space, suggesting that gene regulation is under strong selective pressure[6-8]. Second, perturbations in gene expression levels can result in developmental defects and disease[6-8]. In this sub-chapter, I provide support for these two lines of evidence,

which together justify the need for new methods that can precisely tune gene expression levels.

1.1.1 Layers of gene regulation

Gene regulation is a robust process that allows cells to perform routine functions and to respond to changes in their environment. In multi-cellular organisms, carefully orchestrated gene control enables development, differentiation and regeneration[7]. Several layers of gene regulation co-operate within the cell to achieve this control, beginning with the physical distribution of DNA within the nucleus[9]. Recent progress in studying chromatin conformation has revealed that chromosomes are spatially oriented in the nucleus in a manner that physically partitions portions of DNA into interaction domains[10-12] (Fig. 1.1.a).

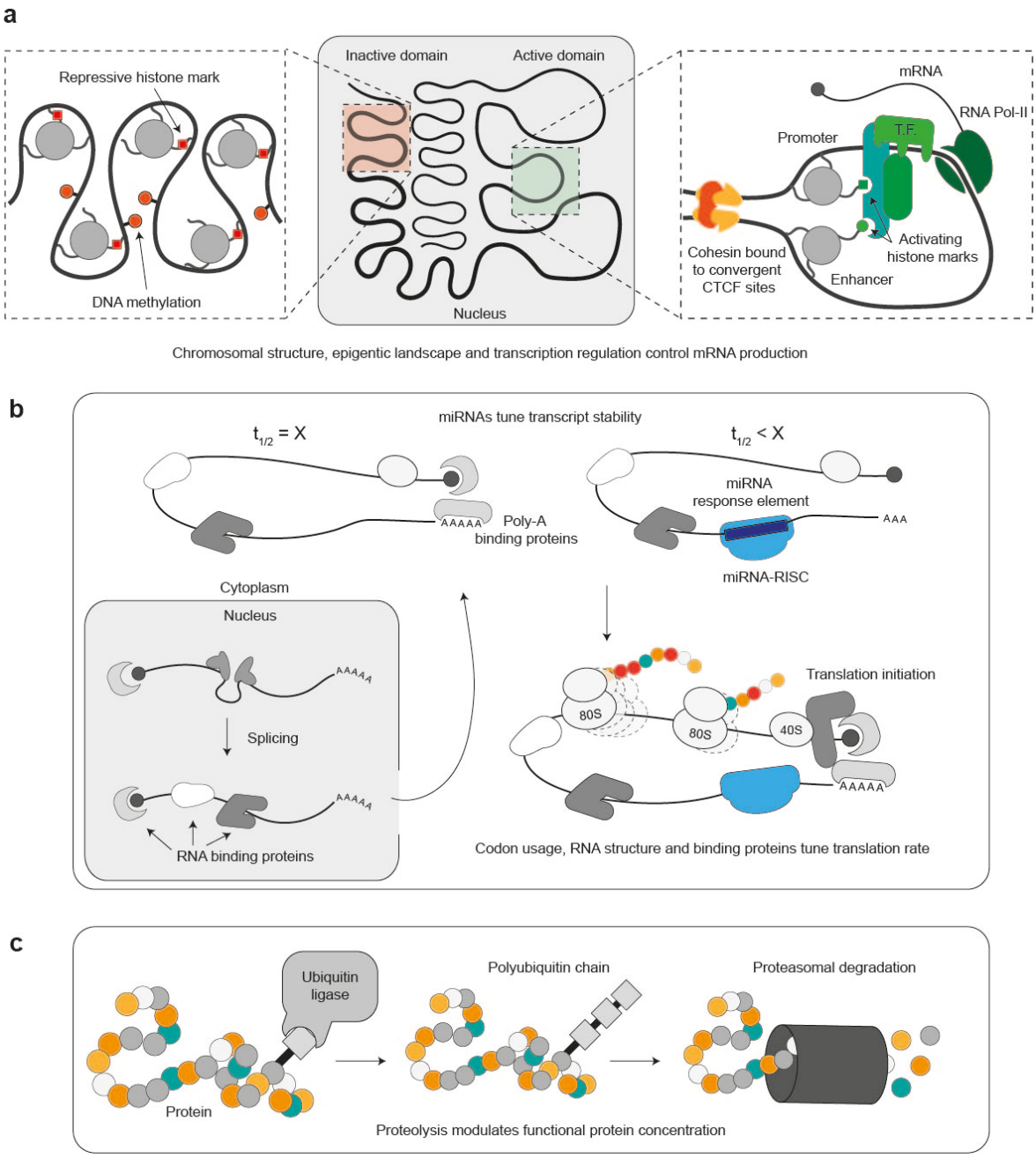

Figure 1.1. Multiple regulatory layers control gene expression levels in eukaryotes.
(a) Examples of transcriptional control of gene expression. **(b)** Post-transcriptional mechanisms regulate mRNA stability and translation state. **(c)** Protein modification and degradation regulate the concentration of active gene-products.

Spatial partitioning likely allows the cell to efficiently allocate transcriptional machinery to the promotors of genes that need to be expressed in a given cell type. Cis-regulatory sequences known as enhancers help bring transcription factors in close proximity to gene promotors and can substantially increase

transcription rates[13,14]. Enhancers are capable of participating in tissue-specific, long-distance looping interactions with promotors megabases away[15]. Controlling when and where enhancers can interact with their cognate promotors is one way that eukaryotic cells co-opt three-dimensional DNA partitioning to achieve cell and tissue specific gene regulation[15].

Covalent modifications to histone proteins and DNA itself, known collectively as epigenetic changes, provide a second layer of gene expression control[16,17] (Fig. 1.1.a). A group of conserved proteins called chromatin writers are able to add and remove acetyl, methyl and phosphate functional groups to and from histone tails[16,17]. These modifications can either influence the affinity of histones for the DNA wrapped around them, thus influencing DNA accessibility, or else provide recognition elements for chromatin readers and transcription factors[16,17].

Interestingly, epigenetic marks and the expression state of associated genes seem to be spatially co-ordinated in the nucleus. Regions of topologically associated DNA often share similar epigenetic landscapes and transcription states[12]. It remains unknown whether epigenetic modifications and chromatin binding proteins are a primary driver of spatial localisation or vice versa, but it is possible that these two layers of regulation work reciprocally, one reinforcing the other. DNA looping interactions mediated by cohesin proteins binding to convergent CTCF sequences appear to sequester loops of DNA into isolated domains of co-ordinated expression status[11](Fig. 1.1.a).

Spatial organisation of DNA and epigenetic modifications mediate their effects by controlling when and where transcription factors can bind to their cognate DNA sequences[18,19] (Fig. 1.1.a). Transcription factors are capable of directly or indirectly regulating transcription by recruiting or prohibiting RNA polymerase II from binding to gene promoters[18]. Changes in chromosome conformation, epigenetic modifications or transcription factor concentration can all lead to concomitant changes in gene transcription levels.

Once a gene has been transcribed into mRNA, post-transcriptional regulation provides an additional layer of control. RNA binding proteins can associate with mRNAs, influencing their stability, localisation and accessibility[20] (Fig. 1.1.b). Additionally, short non-coding RNAs such as miRNAs are able to repress partially complementary transcripts, either by initiating their degradation or by suppressing their translation[21,22] (Fig. 1.1.b, see section 1.4 for more details on miRNAs).

In addition, the rate of translation of an mRNA transcript into a poly-peptide is influenced by mRNA codon usage[23], RNA secondary structure[24] and alternative splicing isoforms[25] (Fig. 1.1.b). A final layer of regulation operates at the protein level. Post-translational modifications, such as ubiquitination profoundly influence protein half-life[26] (Fig. 1.1.c).

At first glance, this multifaceted orchestration of gene expression—which combines chromosome conformation, epigenetic state, transcription factors, posttranscriptional control, differential translational efficiency and post-translational modification—may seem redundant. However, multi-layered gene

regulation allows eukaryotic cells to respond to changes in their environment or developmental cues on widely different time scales. For example, spatially situating a chromosome domain in an accessible conformation near other active genes can allow sustained transcription of a required housekeeping gene in a manner that is insensitive to stochastic changes in the cellular environment[12]. In contrast, ubiquitination of a protein can lead to rapid proteasomal degradation, allowing the cell to quickly respond to acute stress[26]. In fact, as I discuss below, precise, multifaceted control of gene product levels is necessary for proper cellular function.

1.1.2 Gene expression and disease

Although gene regulation superficially appears robust, even partial perturbations to gene expression levels can result in disease. The ever-growing list of diseases that are caused by or associated with mutations affecting chromatin regulators and transcription factors provides evidence for the consequences of misregulation of gene expression. Lee and Young have compiled a list of over 200 gene-regulatory proteins that have been implicated in cancer and diabetes as well immunological, neurological, developmental and cardiovascular disorders[7]. Providing further support for the importance of gene expression regulation, a large-scale analysis of genome-wide association studies determined that 88% of disease-associated polymorphisms are in non-coding regions of DNA[6]. Authors of a recent pre-print article concluded that changes in gene expression account for over 20% of the heritability of 30 common diseases[8]. Taken together, these findings provide compelling evidence that maintaining precise control of gene expression levels is critical for eukaryotic homeostasis and human health.

These lines of evidence suggest that if scientists and clinicians could precisely program expression levels of any gene of interest, we could significantly increase our understanding of the molecular mechanisms of disease. It may also be possible to therapeutically modulate expression levels to restore or even augment normal cellular function.

1.2 The role of gene expression in the anti-tumour immune response

Gene-regulation has been strongly implicated in numerous diseases, including immune-related disorders. This thesis specifically investigates how gene expression levels influence the anti-tumour immune response. Therapies that promote an anti-tumour adaptive immune response (known collectively as immunotherapy) have recently shown great clinical promise[27]. Furthermore, lymphocytes such as T-cells can be removed from human patients, modified and reintroduced by transplantation[27]. Consequently, the ability to control gene expression in the immune system provides a tractable means of understanding and improving cell-based cancer therapy[28,29]. Because of these considerations, tumour immunology is an ideal paradigm for developing and testing a gene expression tuning system.

1.2.1 The adaptive immune system is capable of recognising tumours

In mammals, the adaptive immune system has evolved to defend against foreign pathogens by identifying and reacting to non-self peptide fragments. Genomic instability is a hallmark of cancer and results in high mutational load[30], leading to the presence of non-self antigens in most tumours. It was, therefore, a long-

standing mystery how highly mutated tumours escaped the adaptive immune system to exert their lethal effects. Conversely, could the immune system ever be coerced into responding to cancer?

In the late 19th century, William Coley hypothesized that local stimulation might provide the adaptive immune system with the boost needed to curb tumour growth[31,32]. By injecting inoperable sarcomas with heat-killed bacteria, Coley and others observed potent tumour reduction, providing the first evidence for the efficacy of tumour immunotherapy[31,32]. More than a century later, scientists have begun to understand why the adaptive immune system fails to reject tumours in the absence of such a boost. Under strong selective pressure, tumours evolve to evade immune surveillance[30,33]. In animal models, tumours employ three main strategies to circumvent T-cell cytotoxicity: i) loss-of-function mutations in major histocompatibility (MHC) genes, which encode the proteins necessary for extracellular presentation of peptide antigens to T-cells; ii.) elimination of clones expressing immunogenic mutant antigens; or iii.) upregulation of immunosuppressive molecules within tumours or the tumour microenvironment[33]. Recent clinical studies are beginning to confirm that strategies i.) and ii.) act in humans as well as mouse models[34,35]. Strategy iii.) has been reported extensively in patients with dozens of different sub-types of cancers and is the basis of a highly successful new class of cancer immunotherapy known as checkpoint blockade[27].

1.2.2 Co-inhibitory receptor expression controls the anti-tumour immune response

Exemplifying strategy iii.), several types of cancers express high levels of T-cell inhibitory ligands capable of engaging endogenous co-inhibitory receptors on CD4+ and CD8+ T-cells[36]. Programmed cell death 1 (PD-1) is a well-characterised co-inhibitory receptor expressed on activated T-cells[36-39] (Fig. 1.2.a). When PD-1 engages with its cognate ligand PD-L1, it recruits the phosphatase Shp2 to the T-cell receptor (TCR) and its co-receptor, CD28[39]. Dephosphorylating these two activating receptors impedes T-cell stimulation, proliferation and effector function[39]. PD-1/PD-L1 signalling likely serves an auto-regulatory role during normal immune surveillance, preventing self-reactivity[40]. PD-1 knockout mice display a lupus-like phenotype and protracted lifespan[41]. In humans, polymorphisms in PD-1 have been linked to lupus and arthritis, supporting the protective role of this signalling modality[40,42]. In numerous cancer types, tumours and tumour-resident antigen-presenting cells (APCs) frequently overexpress PD-L1, co-opting this naturally protective mechanism to supress the anti-tumour immune response[36] (Fig. 1.2.a).

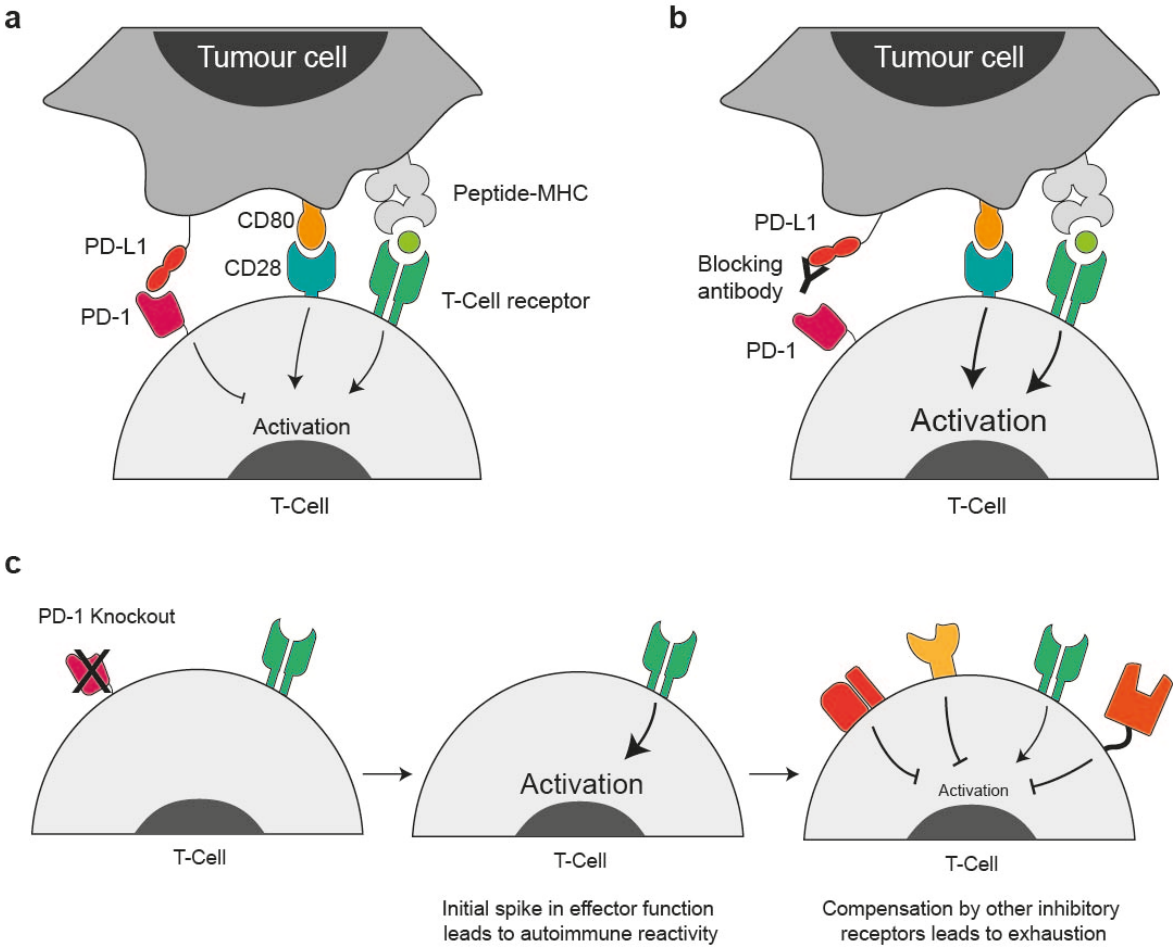

Figure 1.2. Stimulation and inhibition control T-cell activation state and anti-tumour immunity.
(a) Recognition of a cognate peptide presented on a target cell's MHC-I complex as well as co-stimulatory signally through CD28 promote T-cell activation. Signalling through inhibitory receptors such as PD-1 supress activation. Inhibitory signals can come from the tumour itself or antigen presenting cells in the tumour microenvironment. **(b)** Antibodies that block PD-1's ability to interact with its cognate ligand PD-L1 promote T-cell activation and tumour killing. **(c)** The impact of complete PD-1 knockout based on a mouse model of chronic viral infection. Co-inhibitory receptors are coloured red, yellow and orange.

Motivated by the prevalence of this immunosuppressive mechanism in human cancer, researchers developed therapeutic antibodies that block PD-L1, or its cognate receptor PD-1[27] (Fig. 1.2.b). These so-called checkpoint inhibitors showed impressive clinical efficacy in trials for melanoma and lung cancer, yielding durable responses in patients with advanced, chemo-refractory

disease[27,43,44]. Checkpoint blockade antibodies targeting PD-1/PD-L1 or a second inhibitory receptor (CTLA-4) have now been approved for use against multiple cancer types[27].

Although checkpoint inhibitors have produced impressive clinical results and generated mainstream media praise, they are not without limitations. The majority of patients do not respond to these drugs, and early responders can develop resistance[45]. Interestingly, Koyama *et al*. observed that resistance to PD-1 inhibition is often mediated by upregulation of other T-cell co-inhibitory receptors such as TIM-3[46]. Furthermore, up to 70% of patients treated with PD-1 or PD-L1 blocking antibodies suffer from immune-related adverse events (IRAEs)[45]. For CTLA-4 blockade, up to 90% of patients suffer from IRAEs [45]. Common immune-related side effects of checkpoint inhibitors include vitiligo, diarrhoea, colitis and thyroid dysfunction[45]. Given that PD-1-null mice and humans with PD-1 polymorphisms exhibit autoimmunity, it is not surprising that systemic checkpoint inhibition leads to IRAEs in so many patients.

Intravenously delivering checkpoint inhibitors likely lowers the threshold of activation for all T-cells in the body, irrespective of their specificity[45]. An alternative approach that has been proposed to reduce off-target immunity is to eliminate PD-1 specifically on tumour reactive T-cells via *ex vivo* genome engineering[47,48]. By isolating and modifying only tumour infiltrating T-cells or peripheral blood monocytes that demonstrate reactivity to tumour specific peptide-MHC tetramers, it seems possible to improve specificity towards tumours rather than healthy tissues. Unfortunately, a recent study undertaken in a mouse model of

cytomegalovirus (CMV) infection challenges this possibility[49]. When PD-1 knockout T-cells specific for CMV were adoptively transferred into an infected host, there was an initial burst in T-cell proliferation and effector function, as expected by the authors[49] (Fig. 1.2.c). Surprisingly, after this acute burst of activity, other inhibitory receptors such as CTLA-4 were upregulated in a compensatory fashion, ultimately resulting in highly dysfunctional T-cells and a lack of viral clearance[49] (Fig. 1.2.c). Although this study was performed in a viral infection model, these findings suggest that complete PD-1 deletion will likely fail to generate durable anti-tumour immunity.

Taking these considerations into account, it is likely that fine-tuning PD-1 expression levels on effector T-cells will be more effective than completely abrogating its activity. In a recent *ex vivo* study, researchers transiently modulated PD-1 levels on CD-8 T-cells specific for an HIV peptide antigen by transfecting different doses of PD-1 mRNA[38]. Increasing PD-1 expression suppressed T-cell activation, proliferation and cytokine production in a dose-dependent fashion[38]. Although this work was carried out with HIV-specific T-cells, these findings suggest that PD-1 expression levels might also dictate the strength of the anti-tumour immune response.

A method of stably tuning PD-1 on tumour reactive T-cells could be used to identify an immunotherapy "Goldilocks zone". T-cells that are sensitive to immunogenic tumours but maintain enough auto-regulatory signalling to avoid reactivity against healthy tissues would provide a safer and more effective treatment modality. Furthermore, tuning PD-1, rather than ablating it entirely,

could prevent compensatory co-inhibitory receptor expression such as that observed by Odorizzi *et al.* in an animal model[49]. The aforementioned effort to titrate PD-1 relied on transiently overexpressing the receptor by transfection[49]. In the context of anti-cancer immunotherapy, a method of stably down-regulating endogenous PD-1 is required. The specific case of PD-1 exemplifies the need for a stable and predictable gene expression tuning system.

1.2.3 Targeting tumour associated antigens with adoptive T-cell therapy

In addition to checkpoint blockade, adoptive T-cell therapy (ACT) is a second class of immunotherapy where gene expression levels may play a crucial role. The goal of ACT is to generate an anti-tumour immune response by expanding tumour-specific T-cells *ex vivo* and infusing them into a patient[50]. Although there are several variations on this approach, one effective ACT method is to expand tumour-infiltrating T-cells derived from biopsies or resected tumours[50] and re-infuse them intravenously. This approach has been shown to yield an objective response in 49% to 72% of metastatic melanoma patients[50]. The success of ACT depends on tumours expressing and presenting immunogenic antigens[51]. Tumour associated antigens fall into three classes: i.) overexpressed antigens that are present on healthy tissues but present at higher levels on tumours; ii.) neoantigens that occur when a cancer mutation generates a non-self peptide; or iii.) cancer-germline antigens which are expressed on tumours but otherwise restricted to the male germline[51]. Testes are an immune-supressed organ likely to be shielded from cancer-germline specific T-cells. Consequently, both cancer-germline and neoantigen reactive T-cells are thought to be more specific than T-

cells reactive to overexpressed antigens[51]. However, this hypothesis has not been rigorously tested in controlled experiments.

An additional unknown factor is how the level of expression of tumour-associated antigens influences the efficacy of ACT. I hypothesize that finely controlling tumour antigen expression could elucidate this relationship and help predict which tumours are likely to respond to ACT on the basis of antigen expression. Furthermore, experimentally controlling antigen expression levels could help determine if overexpressed antigens are tractable therapeutic targets. Antigen expression level may also be an important predictor of which tumours will respond to checkpoint blockade. Blocking co-inhibitory receptor signalling can only lead to tumour clearance if T-cells can detect cancer cells. Even in the absence of treatment, antigen expression levels likely shape the evolving interactions between a tumour and effector T-cells. In this thesis, I provide evidence that experimentally tuning antigen expression has the potential to improve our understanding of the relationship between tumours and the immune system.

1.3 Uses and limitations of existing tools for experimental manipulation of gene expression

A large body of existing research demonstrates that adding or removing a gene product from a biological context can elucidate the causative role that a specific gene plays in a given process or pathology. Several technologies have been developed to either increase or decrease the expression of user-specified genes. However, these methods were not designed for, nor are they effective at, precisely tuning gene expression levels in mammalian cells. Knockouts[52], RNA interference

(RNAi)[53] and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) interference (CRISPRi)[54] are methods for eliminating or repressing gene expression. Knockins, expression vectors and CRISPR activation (CRISPRa) enable expression of transgenes or over-expression of endogenous genes[55]. These methods have all provided biological insight into the role that gene products play in biological processes, but none are suitable for precise, predictable control of gene expression levels. Other researchers have previously developed fine tuning systems, largely based on variable promoters[56-60] or ribosome binding sites[61-63], to overcome this limitation in bacteria and yeast. In this sub-chapter, I outline a selection of available genetic manipulation strategies, highlighting their key advantages and limitations. In assessing these limitations, I justify my decision to develop a new gene tuning technology based on libraries of synthetic miRNA response elements (MREs).

1.3.1 Methods for reducing or eliminating gene expression

Gene knockouts in human cells and animal models have helped scientists decipher the function of thousands of proteins but this strategy has significant limitations. Historically, gene knockouts have been achieved either by random mutagenesis[64], or in a targeted fashion by homologous recombination (HR)[52] (Fig. 1.3.a). Unfortunately, many gene knockouts are lethal at the embryonic stage, precluding the study of their function in intact adult tissues[65]. Furthermore, as discussed in detail in sub-chapters 1.1 and 1.2, the effect of a protein in a given cellular context can be highly dose-dependent, with complete knockouts failing to recapitulate relevant physiological changes. Mono-allelic knockouts can partially address this problem by offering an intermediate level of expression, but this falls

short of enabling quantitative changes other than a potential 50% reduction in expression. Finally, conventional mutagenesis methods such as HR and random mutagenesis are extremely inefficient, technically challenging, expensive and time consuming[52,64].

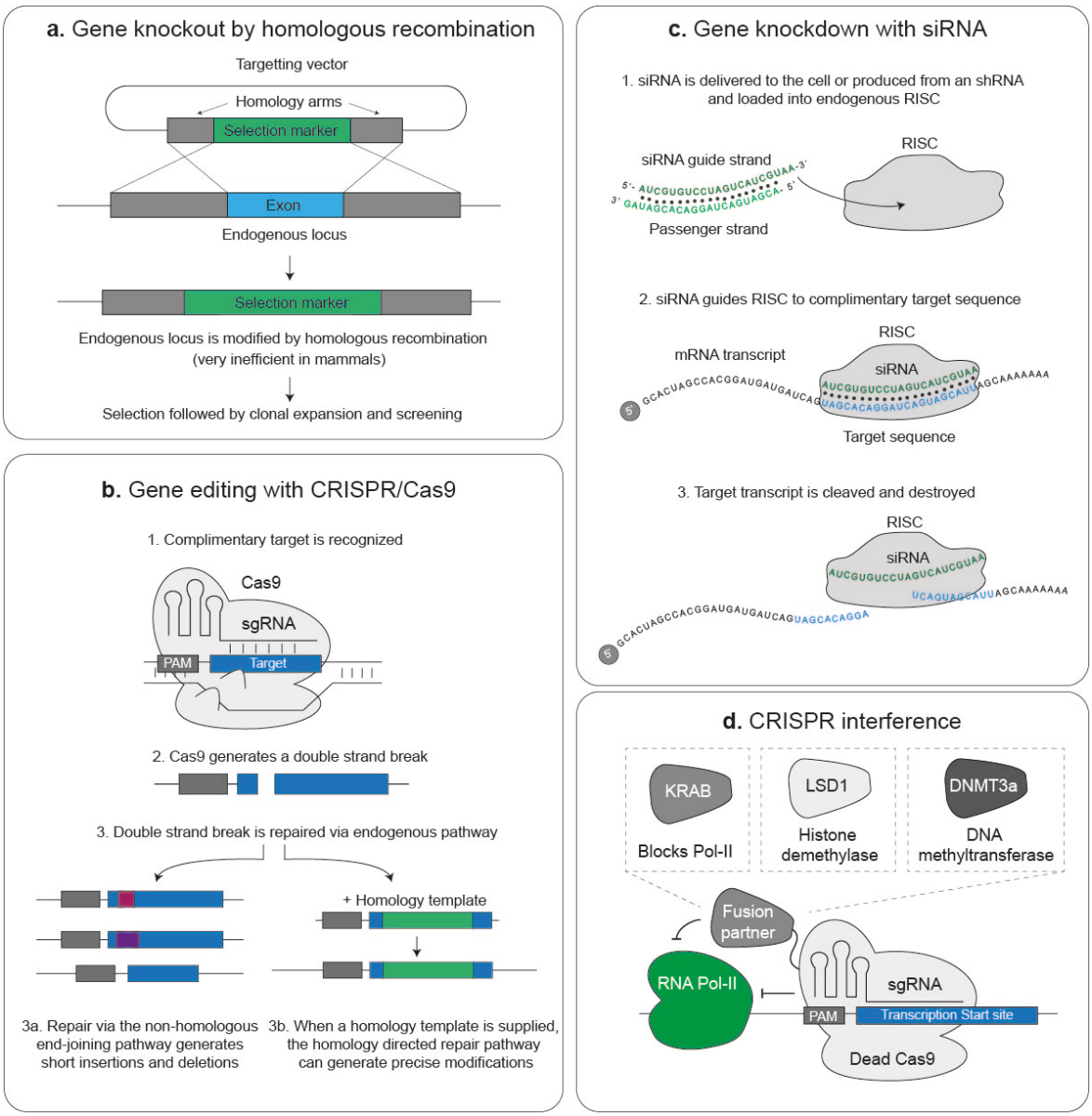

Figure 1.3. Methods of reducing or eliminating gene activity in eukaryotic cells.
(a) Targeted gene-knockouts can be achieved by homologous recombination using a DNA donor with homology to the locus of interest. **(b)** The CRISPR/Cas9 system can improve the efficiency of homologous recombination or insertion/deletion mutations (PAM = protospacer adjacent motif). **(c)** siRNA co-opts endogenous RNA interference machinery to post-transcriptionally repress gene expression in a sequence specific fashion (RISC = RNA induced silencing

complex). **(d)** CRISPR/Cas9 can be repurposed to repress transcription. Dead Cas9 carries loss of function mutations in both of its nuclease domains.

Recent advances in genome engineering technologies have overcome some of the limitations of conventional knockout approaches but still do not provide a solution for precise control of gene expression levels. Prior to genome engineering, knockouts were generated by supplying cells with a repair template, sharing homology to the gene of interest and incorporating a mutation[52] (Fig. 1.3.a). At low frequency, the endogenous homologous recombination pathway can incorporate the mutation encoded by the repair template into the host genome. The goal of genome engineering is to increase the frequency of repair by introducing double strand breaks at the locus of interest to. The CRISPR/Cas9 system—the latest and simplest genome editing methodology—enables efficient, targeted DNA cleavage[66,67]. Derived from a highly conserved bacterial/archaeal adaptive immune mechanism, Cas9 is an enzyme capable of cleaving double stranded DNA in a targeted sequence-specific fashion[66,67] (Fig. 1.3.b).

By coupling the Cas9 nuclease with a synthetic guide RNA (sgRNA), it is possible to direct Cas9 to any sequence with an appropriate proto-spacer adjacent motif (PAM)[66,67] (Fig. 1.3.b). Cas9-generated double strand breaks can be repaired by one of two endogenous repair pathways, non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ can erroneously repair a Cas9-generated double strand break, introducing short insertions/deletions (indels) or polymorphisms[67] (Fig. 1.3.b). Frame-shifting indels in an open reading frame (ORF) can lead to a premature stop-codon and a loss of protein function[67]. If researchers supply a repair template with homology to the locus targeted by Cas9,

it is possible to co-opt the endogenous HDR pathway[67] (Fig. 1.3.b). Depending on the design of the repair template, a researcher can introduce nucleotide substitutions, short DNA tags or even large insertions at the targeted locus by HDR[68]. Consequently, the CRISPR/Cas9 system has substantially improved the ease and efficiency of genetic manipulation[68]. That being said, similar to previous mutagenesis methods, Cas9-generated knockouts render genes non-functional, completely eliminating expression of their wildtype protein products.

By delivering exogenous small interfering RNA (siRNA), it is possible to repress the expression of a gene rather than rendering it entirely non-functional[53] (Fig 1.3.c). siRNAs, or short-hairpin RNAs (shRNAs), designed to target a gene of interest, exert their function by co-opting the endogenous RNA interference machinery[69] (Fig. 1.3.c). siRNAs guide the endogenous RNA-induced silencing complex (RISC) to degrade complementary target mRNAs (Fig. 1.3.c).

Unfortunately, the extent of downregulation elicited by si/shRNAs is variable and unpredictable[70,71]. Eukaryotic mRNAs adopt complex higher-order structures and are decorated with a myriad of RNA binding proteins[20]. The accessibility of a given target sequence[71] and the half-life of the mRNA[72] can both influence siRNA efficacy in a manner that is difficult to predict. Furthermore, RISC can degrade mRNAs that display only partial complementarity to an exogenously delivered siRNA. Consequently, another major drawback of siRNAs is the pervasive off-target effects that they instigate[73]. siRNAs are easy to use but do not offer a specific and predictable means of tuning gene expression levels.

Recent efforts to repurpose Cas9 as a transcriptional repressor have provided an alternative method of downregulating endogenous gene expression (Fig. 1.3.d). By mutating Cas9's two nuclease domains, Gilbert *et al.* created a catalytically inactive version of the protein dubbed dead-Cas9 (dCas9)[55]. When coupled with an sgRNA targeting a transcription start site, dCas9 sterically interferes with transcription initiation in a method known as CRISPRi[55] (Fig. 1.3.d). dCas9 has also been fused to repressive domains like KRAB to further increase the efficiency of CRISPRi[55] (Fig. 1.3.d). Although it is a useful research tool, CRISPRi requires expression of Cas9, a protein that is immunogenic in mammals[74]. Stable expression of Cas9 is consequently unsuitable for some *in vivo* and therapeutic indications. Thus, CRISPRi displays advantages and disadvantages compared to RNAi, but neither method provides fine control over gene expression levels.

1.3.2 Methods for introducing or increasing gene expression

In sum, the aforementioned methods of knockouts, RNAi and CRISPRi are suitable for turning gene expression down or off. However, it can also be informative to experimentally increase the expression of an endogenous gene or to introduce a new gene product. Homologous recombination can be used to knock a gene into the genome at a defined locus. Knockins can be achieved in mice and human cells using the CRISPR/Cas9 system and a repair template harbouring the gene of interest[68] (Fig. 1.4.a). NHEJ is a rapid repair mechanism that competes with HDR in most cell types[75]. This makes precise knockins an inefficient process, as Cas9 mediated double strand breaks generally result in indels[67].

Another more efficient method of introducing a transgene into mammalian cells is to use an expression vector such as a plasmid or virus[76] (Fig. 1.4.a). Plasmids can be delivered to cultured cells at high copy numbers using lipid-based transfection reagents or electroporation[76]. Viral vectors such as lentiviruses can infect most dividing cell types in culture and *in vivo*[77]. Importantly, lentiviruses integrate into the host genome, enabling transgene expression that is stable over time[77]. By controlling the viral titre used to infect the cells of interest, it is possible to coarsely control the number of integrations of the viral genome and the level of expression of the transgene[78]. Commonly used viral promoters drive very high levels of gene expression[78]. High expression levels can be useful when large quantities of a protein are required to observe a phenotype, but it also makes these vectors unsuitable for fine-tuning expression. Even with just a single integration, protein levels can exceed a physiologically relevant dose.

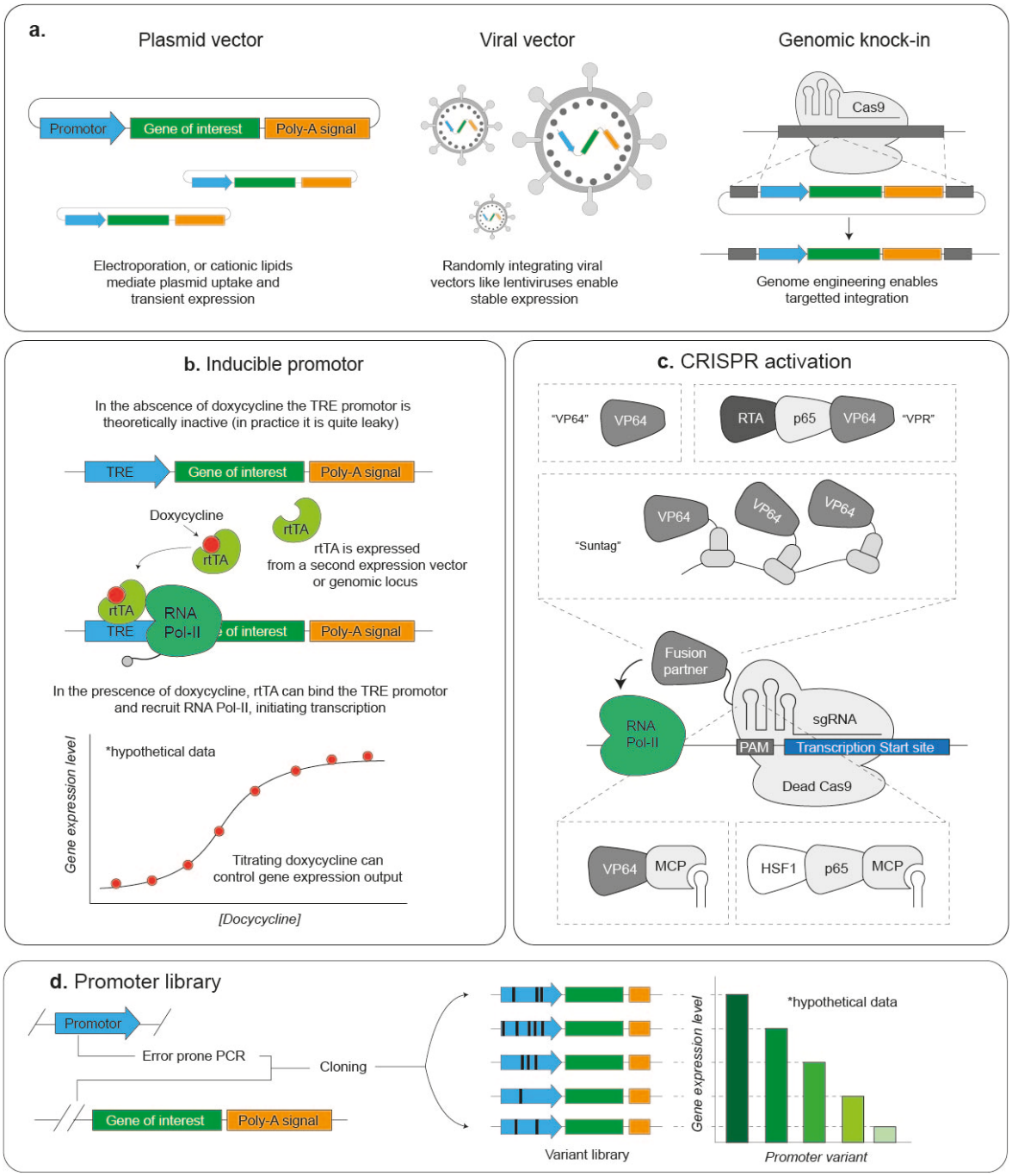

Figure 1.4. Methods of introducing or overexpressing genes in eukaryotic cells.
(a) Methods of delivering transgenes to eukaryotic cells. **(b)** An example of a small-molecule-inducible promotor (TRE = tetracycline responsive element, rtTA = reverse tetracycline-controlled transactivator). **(c)** CRISPR/Cas9 can be repurposed as a transcriptional activator. Dead Cas9 is fused to a general transcription factor such as VP64, RTA, HSF1 or p65. MCP recognizes MS2 loops inserted into the sgRNA scaffold and can be used to recruit additional activator proteins. **(d)** In yeast, libraries of promotors generated by error prone PCR (or through rational engineering) can drive different levels of gene expression. A similar method was first demonstrated in bacteria.

In order to achieve conditional and titratable transgene expression, researchers have developed inducible promoters[79,80] (Fig. 1.4.b). Depending on the configuration, delivery of a small molecule can either permit[79] or prevent[80] a transcriptional activator from binding to the promoter. Altering the concentration of the small molecule, such as doxycycline, induces a concomitant change in transcriptional output[79,80] (Fig. 1.4.b). Inducible promoters are a powerful tool for titrating gene expression, but the small molecules that trigger them have confounding effects on cellular function[81] and are difficult to dose consistently *in vivo*.

In addition to titrating transgene expression, it is also possible to upregulate expression of endogenous genes using a technique called CRISPR activation[82] (Fig. 1.4.c). CRISPRa is effectively the inverse of CRISPRi. dCas9 is fused to a transcriptional activator domain such as VP64 and directed to a transcription start site by a sgRNA[82] (Fig. 1.4.c). CRISPRa can drive strong upregulation of endogenous gene expression, but its efficacy is highly variable between genes[83]. The factors that determine how readily a gene can be upregulated with CRISPRa have yet to be fully detailed, but the baseline transcriptional activity of the target gene and the local epigenetic environment likely play a role [84]. As with CRISPRi, CRISPRa offers a powerful resource for manipulating endogenous gene expression, but it is not yet suitable for quantitative tuning.

1.3.3 Existing methods for fine-tuning gene expression levels

To address the need for precise, predictable, small molecule-free expression modulation systems, researchers devised genetically encoded fine-tuning systems[56-63,85-91]. Alper *et al.* provided an early demonstration of this concept by subjecting a bacterial promoter to error-prone PCR and monitoring the expression output of the resulting mutants[57]. This pioneering study enabled discrete, stepwise control of gene expression spanning a 325-fold range[57]. The same approach was also effective for tuning gene expression in yeast[57] (Fig. 1.4.d).

The resolution and predictability of genetically-encoded gene tuning has subsequently been improved through a myriad of approaches. Modulating translation initiation rates by engineering the ribosome binding site and surrounding sequence has enabled improved fine-tuning in bacteria[61-63]. In yeast, rationally engineering transcription-factor binding motifs[56,58,59], controlling nucleosome architecture[90] and modulating splicing efficiency[91] have all been employed for gene expression tuning. Fine-tuning systems have provided new insight into bacterial and yeast biology. One recent high-throughput study used a library of 130-engineered promoters with a 500-fold range in expression output to tune 81 genes in yeast[60]. The authors assayed the effect of gene expression level on *Saccharomyces cerevisiae* fitness in multiple growth conditions[60]. They concluded that tuning gene dose has disparate effects on growth that depend on the carbon source supplied[60]. Surprisingly, wildtype expression levels for approximately 20% of genes were suboptimal for yeast fitness[60].

In addition to learning about endogenous biological processes, promotor libraries have also been used to optimise overexpression of recombinant proteins. Pasini *et al*. showed that tuning expression of a repressor protein, and consequently a downstream transgene, reduced the metabolic burden associated with heterologous protein production in bacteria[92].

Genetically encoded methods for tuning gene expression in bacteria and yeast have provided new biological insight and improved synthetic biology applications. In my PhD, I set out to develop a technology for modulating mammalian gene expression levels that is similarly free of confounding exogenous triggers. In multicellular organisms, epigenetic landscape, transcription factor expression profiles and translational control systems vary with cell identity[18,25,26,93]. Consequently, the order of expression exerted by promoter libraries is likely to be difficult to predict.

A tuning system based on artificial upstream open reading frames (uORFs) showed promise for controlling transgene translation rates in mammalian cells[89]. The premise of this technology is that uORFs suppress translation of downstream genes in a sequence dependent manner. Altering the uORF sequence proved to be an effective tool for controlling transgene expression levels. However, this approach will be difficult to implement for endogenous genes and is incompatible with tissue specific gene tuning[89]. Systems that can be programmed for either ubiquitous or cell-type specific tuning would be of great benefit to researchers studying complex multicellular organisms.

I reasoned that precisely modulating RNA stability at the post-transcriptional level would enable portable and predictable gene expression tuning that could be programmed to function in diverse cell types. miRNAs are naturally occurring fine-tuners of gene expression[21,94]. These well-studied regulatory RNAs are capable of repressing gene expression in nearly all mammalian cell types[95]. Furthermore, microRNAs exert their function by interacting with short, well-defined miRNA Response Elements (MREs) [5,94,96]. I reasoned that by designing a library of synthetic MREs with varying complementarity to an endogenous microRNA, it would be possible to predictably tune gene expression levels in mammalian cells. Engineered MREs could be introduced into an expression vector to tune a transgene or they can be knocked-in to defined genomic loci to tune endogenous gene expression.

1.4 miRNAs as fine-tuners of gene expression

In this sub-chapter, I will provide pertinent background information on miRNAs. I will briefly describe miRNA biogenesis and mechanism of action, focusing on the characteristics that make miRNA effective modulators of endogenous gene expression. I will also discuss current gaps in our knowledge of miRNA function that need to be addressed in order to co-opt these regulatory molecules for developing a synthetic fine-tuning platform. Finally, I will propose a strategy for re-directing miRNAs to precisely modulate expression of transgenes or user-specified endogenous genes.

miRNAs are short non-coding RNAs, 20 to 23nt in length, that are capable of repressing targets bearing partially complementary sequences in the

transcriptome[21,94]. At least 400 miRNAs are expressed in human cells, many of which are conserved across metazoans[21,95]. More than half of all mammalian mRNAs contain predicted or validated miRNA target sites (also known as MREs)[97]. The evolutionary conservation and pervasive nature of miRNA targeting provide evidence for the importance of miRNA's regulatory function. Mutations that impair normal miRNA-mediated gene regulation have been linked to cancer, diabetes, immune disorders and cardiovascular disease in humans and mouse models[94]. Mutations that perturb global miRNA biogenesis and function and even mutations that disrupt individual miRNAs or MREs can cause disease, affirming the importance of post-transcriptional gene repression[94].

1.4.1 miRNA biogenesis

miRNAs are often expressed within hairpin structures in RNA Polymerase-II generated transcripts known as primary miRNAs (pri-miRNAs)[98] (Fig. 1.5.a). These miRNA-containing hairpins are recognized and liberated from the primary transcript by a complex of proteins known as Microprocessor[99] (Fig. 1.5.a). Microprocessor comprises two nucleases, Drosha and DGCR8, that cleave the pri-miRNA, leaving behind a characteristic 2nt overhang[99]. The resulting hairpin structure, known as a precursor miRNA (pre-miRNA), is exported into the cytosol by Exportin-5 and RAN-GTP[100] (Fig. 1.5.a). Next, the nuclease Dicer cleaves the loop away from the pre-miRNA, leaving behind a short miRNA duplex[101] (Fig. 1.5.a). The final step in miRNA biogenesis occurs when the miRNA duplex is loaded into an Argonaut (AGO) protein, and one of the strands of the duplex, known as the passenger strand, is released[94] (Fig. 1.5.a). The resulting

ribonucleoprotein peptide (RNP) complex, comprising the miRNA, AGO and other accessory proteins, is collectively known as RISC.

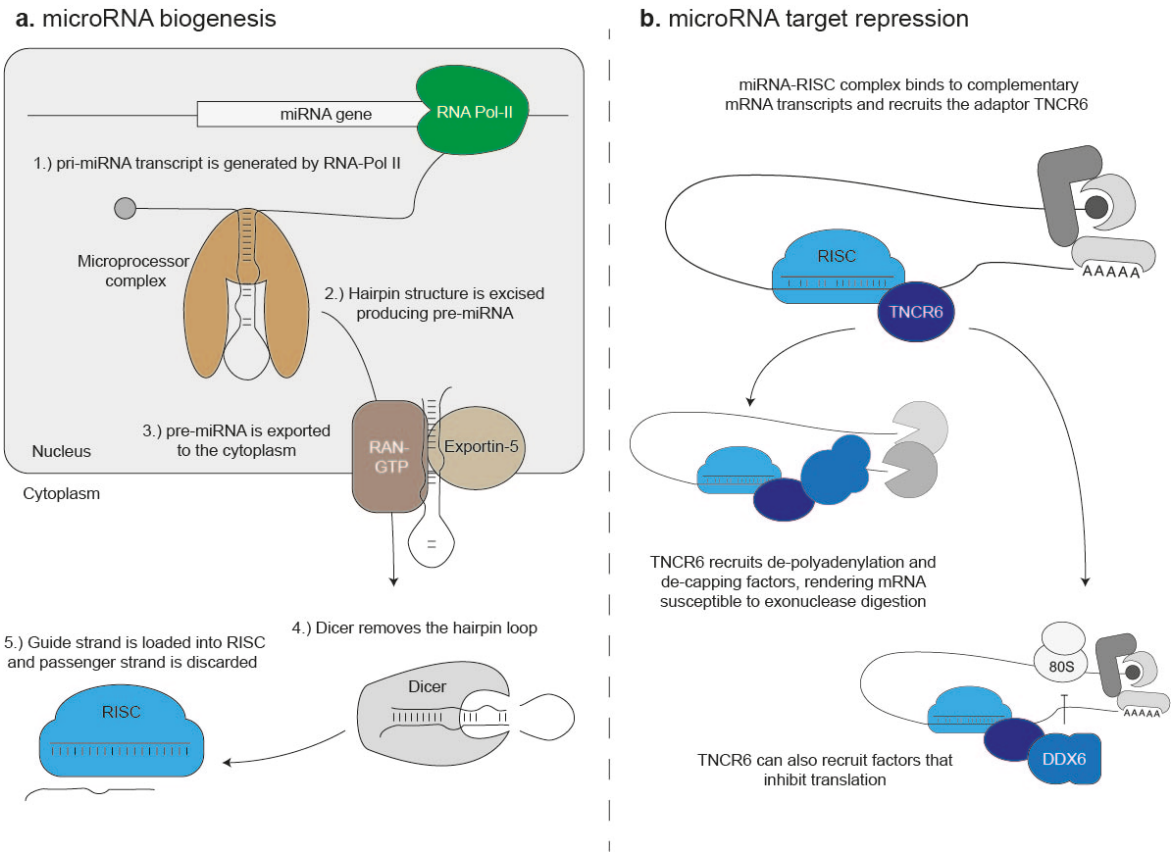

Figure 1.5. miRNA biogenesis and mechanism of action.
(a) Key steps in miRNA expression and maturation in mammalian cells. **(b)** General mechanisms of miRNA-mediated gene repression include transcript destabilisation and translational inhibition.

1.4.2 miRNA target identification and repression

RISC scans cytoplasmic mRNAs by lateral diffusion until it encounters sequences with complementarity to nucleotides 2-8 from the 5' end of the miRNA, a region known as the seed[102]. Upon identifying a potential target, RISC ceases to diffuse along the mRNA, and target repression can begin via multiple mechanisms[102]. Certain AGO family members such as AGO2 can directly cleave a transcript, but this is thought to require extensive complementarity between the miRNA guide and the mRNA target[103]. When miRNA:mRNA complementarity is less extensive,

RISC can initiate mRNA destabilization[104]. This latter mechanism is far more prevalent than direct cleavage in mammals[94]. In cases of lesser complementarity, RISC recruits the adaptor protein TNRC6[94,104] (Fig. 1.5.b). TNCR6 subsequently recruits de-adenlyases to the poly-A tail (Fig. 1.5.b)[105]. Hydrolysis of the poly-A tail is followed by removal of the 5'-7-methyl-guanosine cap and, ultimately, exonuclease-mediated digestion of the transcript[105] (Fig. 1.5.b).

In addition to instigating transcript degradation, miRNAs can also repress translation of their target mRNAs. The adaptor protein TNCR6 can recruit DDX6, a helicase that suppresses translation by competitively inhibiting translation initiation factors (Fig. 1.5.b)[106]. Transcript degradation and translational repression have both been documented in humans and mice but there is an ongoing debate as to which is the dominant mechanism of miRNA-mediated repression[107-110]. Furthermore, the extent to which these two mechanisms are correlated across MREs is unknown. A recent study demonstrated that miRNA-mediated mRNA degradation can occur co-translationally[111]. The authors showed that exonucleases hydrolyse mRNA from 5' to 3' following the last translating ribosome[111]. This work supports a mechanism whereby translational repression and transcript degradation are physically coupled processes.

1.4.3 Re-purposing endogenous miRNAs

In addition to repressing endogenous gene expression, miRNAs have been exploited to control transgene expression[78,112]. Much like coding transcripts, many miRNAs display differential expression patterns across cell types[95]. Synthetic MREs have been used as reporters for miRNA expression in live cells[113]. Brown *et*

al. ingeniously demonstrated that synthetic MREs with complementarity to tissue specific miRNAs can be used to restrict transgene expression to a tissue or cell type of interest[78]. Expanding on this approach, Xie *et al.* developed a cell-classifier gene circuit capable of specifically inducing apoptosis in HeLa cells on the basis of their miRNA expression profile[112]. Synthetic miRNA target sites are an effective tool for regulating transgene expression[78,112] but have hitherto been used exclusively to elicit binary expression control. Given that miRNAs tune, rather than completely silence, endogenous targets[94], I reasoned that synthetic MREs could also be engineered to fine-tune transgene expression in a user-specified manner.

To engineer synthetic MREs capable of precisely and predictably tuning gene expression levels, I considered the factors that influence the strength of miRNA mediated repression. Several parameters are known to influence the magnitude of miRNA-target interactions, such as miRNA abundance and the relative stoichiometry between a miRNA and its putative target pool [113-115]. Contextual features such as mRNA secondary structure and target accessibility have also been reported to impact the strength of miRNA regulation[116,117]. However, miRNA abundance and contextual RNA features are intrinsic to the cell type and gene of interest and cannot be readily manipulated to control gene output levels. The extent of complementary base-pairing between a miRNA and its cognitive MRE is also known to impact the strength of repression[96,102,117,118]. I reasoned that altering the sequence of a synthetic MRE, and consequently its complementarity to an endogenous miRNA, could provide a minimally disruptive mechanism for tuning gene expression levels.

In a seminal study, Brennecke *et al*. used a small panel of MRE reporter constructs to determine how miRNA:target complementarity influences the strength of miRNA mediated repression[96]. They concluded that complementarity in the seed region (nucleotides 2-8 from the 5' end of the miRNA) is the minimal requirement for potent downregulation[96]. They also reported that complementarity with the 3' end of the miRNA had little impact on repression. More recently, Wee and colleagues systematically introduced di-nucleotide mismatches between an siRNA and target mRNA[118]. They loaded each mismatched siRNA into AGO2 and quantified target slicing kinetics using an *ex vivo* cleavage assay[118]. In addition to reaffirming the importance of the seed sequence, Meng-Wee *et al*. also uncovered an important role for nucleotides 13 to 16, a domain they dubbed the supplementary region[118]. Although more comprehensive than previous reports, this study has several limitations. In this work, mismatches were introduced in the siRNA rather than the target site, which could influence RISC assembly and stability in addition to target selection. Additionally, miRNA functionality was quantified using *ex vivo* binding and slicing assays that do not capture the effects of de-polyadenylation or translational repression[118]. Finally, only around 30 distinct mismatches were introduced into the siRNA[118]. This is a relatively small sampling (<2%) of the 2346 possible single or di-nucleotide variants of a 23nt siRNA. Although complementarity is clearly an important determinant of the strength of miRNA repression, the details of this relationship have yet to be fully elucidated.

Bases on these considerations, I reasoned that if we could improve our understanding of how the complementarity between a miRNA and MRE influences gene repression, we could use this information to engineer MREs of known

potency. By introducing these engineered MREs downstream of a gene of interest, it would be possible to achieve predictable tuning of expression levels. In the work presented herein, we comprehensively evaluated the impact of MRE sequence variation on the magnitude of miRNA mediated gene repression in live cells and use that information to predictably tune gene expression.

In chapter 3 of this thesis, I describe a fluorescent reporter system we developed for understanding miRNA-target interactions in live cells. We used this approach to test a small panel of rationally designed MREs with varying complementarity to endogenous or exogenous miRNAs. This small set of MREs allowed us to control reporter expression levels in a stepwise fashion.

In chapter 4, I describe a high-throughput assay to simultaneously measure the impact of all possible single and di-nucleotide MRE variants on transcript degradation and translational repression. This high-throughput method allowed us to understand how each nucleotide in an MRE contributes to miRNA repression.

In chapter 5, I describe how we applied this empirical understanding of miRNA targeting preferences to tune a fluorescent reporter protein with high precision. We also used MRE variants to tune PD-1, a co-inhibitory receptor that regulates the anti-tumour immune response. We tuned transgenic PD-1 overexpression in Jurkat T-cells and used the CRISPR/Cas9 system to introduce synthetic MREs into the endogenous PD-1 3'UTR.

Finally, in chapter 6, I show that precise gene tuning can provide new insight into the anti-tumour immune response. To that end, we tuned expression of a tumour-associated antigen in a mouse melanoma model. Our MRE-based tuning technology allowed us to determine that T-cell activation and tumour killing are functions of antigen expression levels in our model system.

1.5 Summary of thesis aims

The primary aim of this thesis is to develop a technology that will enable precise tuning of gene expression levels in mammalian cells. I have sub-divided this primary aim into four milestones, each of which is addressed in a different results chapter of this thesis. These milestones are stated below.

1. Demonstrate that altering the complementarity between a synthetic MRE for a miRNA alters the magnitude of repression exerted by that MRE (Chapter 3).

2. Determine how each nucleotide in a synthetic miR-17 MRE contributes to miRNA mediated gene repression (Chapter 4).

3.) Apply this single-nucleotide resolution understanding of miR-17 targeting preferences towards predictably tuning transgene expression in mammalian cells (Chapter 5).

4.) Use this new gene tuning technology to understand how tumour-associated antigen expression influences the anti-tumour immune response (Chapter 6).

Chapter 2: Materials and Methods

For oligonucleotide sequences, flow cytometry reagents (including antibody clone names) and maps of key plasmids used in this study, see the appendix of this thesis. All oligonucleotides were purchased from Integrated DNA Technologies unless otherwise specified.

2.1 MRE reporter and miR-21 expression vector cloning

To monitor MRE functionality, we cloned a reporter plasmid comprising a bi-directional CMV promotor driving expression of iBlue fused to a degradation signal derived from Ornithine Decarboxylase and ECFP fused to the same degradation signal. We began with the parent construct, pll3.7_bi-directional (a generous gift from Tim Rajakumar) which contains a bi-directional CMV promotor and an ORF that contains iBlue coupled to the degradation signal. To introduce an ORF with ECFP fused to the degradation signal, we PCR amplified ECFP from 8xCTS2-mCMVp-ECFP[119], a gift from Quentin Ferry (using the primers EXFP_Cloning_Fwd and EXFP_Cloning_Rev) and the Ornithine Decarboxylase degradation signal from pll3.7_bi-directional (using the primers Deg_tag_Cloning_Fwd and Deg_tag_Cloning_Rev). We used Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB) to perform all PCRs required to produce the MRE reporter constructs. These PCRs were purified using the QIAquick PCR Purification Kit (Qiagen). We linearized the parent plasmid, pll3.7_bi-directional by digesting with the restriction enzymes SbfI (NEB) and NotI (NEB). We combined the purified ECFP and degradation signal PCR products with the parent plasmid using the In-Fusion HD Cloning System (Takara Clontech), following the manufacturer's instructions. We transformed the In-Fusion

product into Subcloning Efficiency DH5α Competent Cells (ThermoFisher). We named the resulting construct pbi-dir-diBlue-dECFP. To create pbi-dir-diBlue-dEYFP, we repeated this process with a EYFP template in place of ECFP.

The MRE reporter plasmid pbi-dir-diBlue-dECFP contains BsmBI sites in the 3'UTR that create unique overhangs. To introduce MRE variants into pbi-dir-diBlue-dECFP, we digested the vector with BsmBI (NEB). MRE inserts were generated by synthesizing and annealing synthetic, overlapping oligonucleotides comprising the MRE and appropriate sticky overhangs. We ligated the inserts to the linearized parent vector for 2 hours at room temperature using T4 DNA Ligase (NEB). We transformed the ligation product into Subcloning Efficiency DH5α Competent Cells (ThermoFisher).

We also generated a miR-21 expression plasmid that co-expressed destabilised EYFP called p-Titre-miR-21. To generate p-Titre-miR-21, we amplified the a 320 bp region of the miR-21 primary transcript centred around the miR-21 guide strand from pEM_RFP_miR-21, a gift from Tim Rajamkumar, (using primers pre_miR21_BiDir_Cloning_Fwd and pre_miR21_ BiDir_Cloning_Rev). To remove diBlue and generate suitable overhangs, we digested pbi-dir-diBlue-dEYFP with AgeI (NEB) and EcoRI (NEB). We digested the miR-21 PCR product with the same enzymes and ligated the insert to the destination vector with T4 DNA ligase, transforming the construct as described above.

2.2 HEK 293T cell culture and transfection

HEK-293T cells were grown in Dulbecco's modified Eagle's medium (DMEM,

Gibco) supplemented with 15% FBS (GIBCO) and 1% Penicillin-Streptomycin (P/S, 10,000 U/mL, Gibco). We screened cells for mycoplasma at the outset of the project. Cells were seeded in 12 well plates, 24 hours prior to transfection, allowing them to reach 80-90% confluency on transfection day. On the day of transfection, we replaced complete growth media with DMEM, 2% FBS (no P/S). For each condition, we prepared three independent transfection mixtures, each containing 1 μg of plasmid DNA and 3 μl Polyethylenimine (PEI, 1 mg/ml, Sigma-Aldrich) in 100 μl Opti-MEM (Gibco). Each mixture was applied dropwise to one well of a 12 well plate and incubated for 24 hours.

2.3 Flow cytometry and fluorescence-activated cell sorting.

All flow-cytometry experiments were performed on the BD LSR Fortessa Analyzer or the FACSymphony (BD Biosciences) and data were analysed using FlowJo (Version 10.3.0). We harvested adherent cells (B16-F10 or HEK 293T) using 0.05% Trypsin with EDTA (Thermo Fisher Scientific). For experiments requiring antibody staining we washed cells with FACS buffer (PBS with 5% FBS) before and after staining. For sorting experiments, we used the SH800S cell sorter (SONY) with a 100μm sorting chip. We used untransfected or unstained cells to set sorting and analysis gates. For experiments involving transient transfection of MRE reporter constructs, we gated cells on iBlue, the internal, unsilenced control reporter.

2.4 Confocal microscopy

To image reporter expression, we seeded cells on glass cover slips in 12 well

plates and transfected as described above. 24 hours after transfection, coverslips were harvested for analysis. Confocal images were acquired using the LSM 780 Confocal Microscope (Zeiss) with the LD LCI PA 25x objective. Imaging was performed with ZEN Black SP7 LSM software (Zeiss) with assistance from Aron Szabo. We used Fiji (Version 1.0, ImageJ) for post-acquisition image processing.

2.5 miRNA quantification

To quantify miRNA expression in HEK 293T cells, we first extracted total RNA using the miRNeasy Mini kit (Qiagen). Next, we reverse transcribed the relevant miRNA using a target-specific structured RT primer (TaqMan Small RNA Assays, ThermoFisher) from 20ng total RNA input, following the manufacturer's instructions. Next we performed absolute quantification of the target miRNA by digital-droplet PCR (ddPCR). We mixed the cDNA with a miRNA-specific Taqman probe (ThermoFisher) and ddPCR SuperMix for probes (Bio-Rad) and partitioned the mixture into droplets using the QX200 AutoDG (Bio-Rad). We PCR amplified the droplet mixture according to the ddPCR SuperMix manufacturer's instructions. Finally, we quantified miRNA abundance with the QX200 Droplet Reader (Bio-Rad) and QuantaSoft analysis program (Bio-Rad).

2.6 MRE kinetic analysis

We evaluated the influence of miRNA and mRNA levels on MRE repression by co-transfecting p-Titre-miR-21 and pbi-dir-diBlue-dECFP into HEK 293T cells. We cultured and transfected cells as described above with 10ng of p-Titre-miR-21 and 990ng of pbi-dir-diBlue-dECFP. We harvested cells after 24 hours and performed flow cytometry as described above. We analysed the resulting FCS files in R

(Version 3.4.1). Cells gated on FSC-A and SSC-A were computationally binned on either EYFP (miRNA reporter) or iBlue (internal transfection control) fluorescence. For each bin we calculated normalized ECFP (MRE reporter) expression using the equation, MFI_{ECFP}/MFI_{iBlue}.

2.7 MRE variant library construction.

We purchased hand-mixed, partially degenerate oligonucleotides from Integrated DNA Technologies (IDT) comprising a constant flanking region and a variable region with partial complementarity to either hsa-miR-17 or hsa-miR-21. In this study, the term synthetic MRE refers to a sequence of equal length, and largely complementary to a given miRNA.

To choose the degree of degeneracy in our MRE library we estimated the probability, P, of recovering 0,1,2,3,4 or more mismatches in a 23 nucleotide MRE for various percentages of complementarity using the binomial formula,

$$P(x) = \frac{23!}{(23-x)!\,x!} p^x q^{23-x}$$

where p is the percentage of complementary base at each position, q is the percentage of mismatched bases at each position and X is the number of matching bases in the entire MRE. Based on this calculation, we decided to incorporate 91% of the complementary base and 3% of each non-complementary base at each position in the MRE libraries to achieve maximum coverage of single and double nucleotide variants. Each degenerate oligo was PCR amplified in triplicate using Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB), using primers miR17_Lib_Gen_F and miR17_Lib_Gen_R, which append BsmBI recognition sites on both sides of the MRE. The resulting PCR products were

pooled and purified using the MinElute PCR Purification Kit (Qiagen).

We performed a large-scale restriction cloning reaction to ligate the degenerate MRE PCR product into the bi-directional MRE reporter plasmid. We linearized 10.5 μg of the reporter plasmid downstream of ECFP by digesting with BsmBI. The degenerate MRE PCR product (300ng) was cut with BsmBI and ligated to the linearized, dephosphorylated (Antarctic Phosphatase, NEB) and gel purified (QIAquick Gel Extraction Kit, Qiagen) reporter plasmid using T4 DNA Ligase (NEB) at 16°C overnight. We purified the ligations using the QIAquick PCR Purification Kit (Qiagen) and transformed approximately 3.6 μg of the purified product into 10-beta Electrocompetent E.coli (NEB) following the manufacturer's instructions. Transformants were plated overnight at 32°C on 24.5 cm^2 ampicillin-treated LB agar plates. We recovered the resulting plasmid library using the QIAfilter Plasmid Midi Kit (Qiagen).

2.8 MRE variant library transfection.

HEK-293T cells were grown and seeded as previously described. We prepared three independent transfection mixtures, each containing 4 μg of the degenerate MRE reporter library, 4 ng of Cel-miR-67 MRE control plasmid and 12 μl Polyethylenimine (PEI, 1 mg/ml, Sigma-Aldrich) in 400 μl Opti-MEM (Gibco). Each mixture was applied dropwise to 4 wells of a 12 well plate and incubated for 24 hours.

2.9 Polysome profiling.

To generate enough cell lysate for polysome profiling, we seeded HEK-293T cells in two independent 15 cm^2 culture dishes, allowing them to reach 70-80% confluency by the day of transfection. For each dish, we combined 25 μl each Lipofectamine 3000 and P3000 Reagent (Thermo Fisher) with 12.5 μg of the degenerate MRE reporter library and 100 ng of Cel-miR-67 MRE control plasmid, transfected according to the manufacturer's instructions and incubated for 24 hours. To arrest translation, cycloheximide (CHX, Merck) was added to the culture dishes at 100 μg/mL for 10 minutes at 37°C. Next, dishes were placed on ice and washed with cold PBS (Life Technologies) supplemented with CHX (100μg/mL). We scraped the dishes in PBS + CHX (100μg/mL), centrifuged the harvested cells at 1,000xg for 3 minutes at 4°C and discarded the supernatant. Cell pellets were re-suspended in 200 μl of hypotonic lysis buffer (10mM HEPES pH 7.8, 1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT, 1% Triton X-100 and 100mg/mL CHX) and incubated for 5 minutes on ice. Next, we lysed the cells with 10 strokes through a 26 gauge needle and pelleted the nuclei by centrifuging at 1,500xg for 5 min at 4 °C. The supernatant was flash frozen in liquid nitrogen and stored at -80 °C.

10-50% (W/V) sucrose gradients were generated using a Gradient Master (Biocomp Instruments) from 10% and 50% sucrose solutions in gradient buffer (100mM KCl, 5mM $MgCl_2$, 20mM HEPES-KOH pH7.5, 1mM DTT, 100μg/mL CHX). We thawed the cell lysates and layered them on top of the chilled sucrose gradients before centrifuging at 4°C for 2 hours at 36,000 RPM in a SW-41 rotor. Gradients were fractionated from the top using a Gradient Fractionator (Biocomp Instruments). To recover RNA from the resulting fractions we added 2.25 volumes

of 8M Guanidine HCl (Sigma-Aldrich) and vortexed the samples. Next, 3.25 volumes of isopropanol were added and samples were incubated overnight at -20°C. Reactions were centrifuged at > 12,000 rpm for 30 minutes at 4°C and the supernatant was aspirated. RNA pellets were re-suspended in a mixture of 90 μl nuclease free H_2O (Invitrogen), 10 μl 3M Sodium Acetate (Invitrogen) and 1 μl 5mg/ml glycogen (Ambion) and precipitated with 250 μl of cold 100% ethanol (VWR). After 30 minutes incubation on ice, samples were centrifuged at >12,000 rpm for 30 minutes at 4°C. Next, pellets were washed with 500μl of 70% ethanol and re-suspended in nuclease free H_2O.

2.10 pDNA and cDNA library preparation and high-throughput sequencing.

We used the All Prep DNA/RNA Mini kit (Qiagen) to simultaneously extract plasmid DNA (pDNA) and mRNA from HEK-293T cells transfected with the degenerate MRE reporter library. After performing a genomic DNA wipe-out step, cDNA was generated from mRNA and polysome-associated RNA using the QuantiTect Reverse Transcription kit (Qiagen) following the manufacturer's instructions. To create amplicon libraries for high-throughput sequencing, the degenerate MRE and a short flanking region were PCR amplified using the primers bi-dir-Miseq-F and bi-dir-Miseq-R. For cDNA and pDNA we used Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB) and the following cycling conditions: initial denaturation (98°C for 30 s), 23 amplification cycles (98°C for 10 s, 65°C for 10 s, 72°C for 10 s) and final extension (72°C for 5 min). For cDNA from RNA recovered from polysome fractions we used KAPA HiFi HotStart

ReadyMix (Fisher Scientific) and the following cycling conditions: initial denaturation (98°C for 30s), 21 amplification cycles (98°C for 10 s, 65°C for 10 s, 72°C for 10 s) and final extension (72°C for 5 min). These initial PCR products were gel-purified using the QIAquick Gel Extraction Kit (Qiagen). We diluted the recovered product between 10 and 30 fold depending on band intensity.

To make amplicon libraries compatible with Illumina machines, we performed a second PCR to append TruSeq index sequences and p5/p7 adapters to each amplicon. We used a dual barcoding strategy where a unique combination of forward and reverse index primers were assigned to each biological sample. We performed the PCRs with Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB) and the following cycling conditions: initial denaturation (98 °C for 30 s), 13 amplification cycles (98°C for 10 s, 62°C for 10 s, 72°C for 10 s) and final extension (72°C for 5 min). We used Agencourt AMPure XP beads (0.75X, Beckman Coulter) at 0.75X to purify the amplicon libraries which we subsequently quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher). The samples were sequenced (150bp PE sequencing) on either the HiSeq4000 (Illumina) or the MiSeq v2 (Illumina).

2.11 High-throughput sequencing data analysis.

High-throughput sequencing data were analysed using R (Version 3.4.1) and all scripts are available upon request. After inspecting the quality of sequencing data with FastQC, we used the Biostrings package (version 2.44.2) to trim reads down to the MRE and subsequently count the occurrence of each type of variant of interest in all amplicon libraries. We calculated variant frequency by normalizing

read counts of each variant of interest to total library read counts in the respective library. We calculated transcript abundance for each variant by dividing its read frequency in the cDNA library to its read frequency in the respective pDNA library. We calculated translation efficiency for variants present in polysome profiles by dividing their read frequency in the heavy-polysome-bound library by read frequency in the respective monosome-bound library.

2.12 Validation of high-throughput sequencing results by RT-qPCR.

To validate our high-throughput sequencing assay we randomly selected miR-17-MRE variants by screening colonies from a 1/30,000 dilution of the variant library by Sanger sequencing using primer bi-dir-MRE-seq-1. Colonies were screened until we identified 15 unique single and double nucleotide variants. These constructs were individually transfected into HEK-293-T cells in triplicate in addition to control reporters encoding a Cel-miR-67-MRE and a perfectly complementary miR-17 MRE using the PEI transfection method described above. 24 hours after transfection we extracted RNA using the RNeasy Mini Kit (Qiagen). For each replicate, cDNA was generated from 100 ng of total RNA using the QuantiTect Reverse Transcription kit (Qiagen). We performed RT-qPCR using the SsoAdvanced Universal SYBR Green Supermix kit (Bio-Rad) on a CFX384 real-time system (Bio-Rad) with primer pairs spanning the MRE (MRE_qPCR-F and MRE_qPCR-F) or within the iBlue transcript (iBlue_qPCR-F and iBlue_qPCR-F) which serves as an internal control. The $\Delta\Delta$Ct method was used to compare expression of all MRE variants to a Cel-67-MRE control reporter by comparing the Ct of ECFP to that of iBlue for each sample replicate.

2.12 Lentiviral vector cloning and virus production.

We generated PD-1 lentiviral expression vectors using standard restriction cloning methods. The parent vector AB.pCCL.sin.cPPT.GFP.miR-17-3p.sensor.PGK.dNGFR.WPRE was a gift from Brian Brown (Addgene plasmid #85866). To simplify subsequent cloning steps a SbfI recognition site was introduced downstream of the minimal CMV promotor. The human PD-1 ORF was amplified from the PD-1 BRET vector (a generous gift from Simon Davis) using the primers PD1_Lenti_Shuttle_F and PDL1_Lenti_Shuttle_R. We digested this PCR product, as well as the destination vector with SbfI and NheI and ligated them using T4 DNA ligase (NEB).

The In-Fusion HD Cloning System (Takara Clontech) was used to replace PD-1 in the lentiviral expression vector with cytoplasmic-localized ovalbumin (OVA) coupled to EGFP by a T2A peptide cleavage signal to create a OVA-T2A-EGFP vector. We PCR amplified T2A-EGFP from pX458 (A gift from Feng Zhang, Addgene plasmid #48138) with the primers GFP_in_fusion_F2 and EGFP-in-fusion-R. OVA (without the first 47 amino acids) was amplified from the OVACyt vector[120] using the primers Ova_In_Fusion_R2 and Ova-in-fusion-F. We fused the parent vector (linearized with SbfI and NheI) with the two inserts following the In-Fusion manufacturer's instructions. To create miSFIT-tuning vectors we generated MRE inserts from short oligonucleotides (IDT). MRE inserts were annealed and phosphorylated (T4 PNK, NEB) and introduced downstream of PD1 or OVA-T2A-EGFP by restriction cloning between NheI and AgeI. MRE insertion was confirmed

by Sanger sequencing using primer BBBdir-seq-2.

To produce lentiviral particles in HEK-293T cells, we co-transfected each lentiviral transfer vector with pCMV-dR8.91 and pMD2.G at a ratio of 1.5:1:1 using Polyethylenimine (PEI, 1 mg/ml, Sigma-Aldrich) as described above. After 24 hours we exchanged the transfection media (DMEM, 2% FBS, no P/S) with full media (DMEM, 15% FBS). We collected and filtered (0.22 μm filter, Millipore) viral supernatant 24 hours later and stored it at -80°C until transduction. We transduced B16-F10 cells and Jurkat T-cells using un-concentrated viral supernatant. Jurkat T-cells (clone 1.G4) were maintained in RPMI-1640 media (Gibco) supplemented with 10mM HEPES (Life Technologies), 1mM Sodium Pyruvate (Life Technologies), and 15% Fetal Bovine Serum (FBS, GIBCO). B16-F10 melanoma cells were grown in DMEM (Gibco) supplemented with 15% FBS (Gibco). To generate miSFIT cell lines, we transduced cells at low multiplicity of infection (For Jurkat T-cells < 15% transduced, for B16-F10s < 3% transduced), waited 5 to 7 days and selected stably transduced cells by FACS using SH800S cell sorter (SONY) with a 100μm sorting chip.

2.13 mESC culture and CRISPR/Cas9 genome editing

OTI mESCs were provided by Jackie Sloane-Stanley. We maintained cells in KnockOut DMEM (Gibco) supplemented with 10% FBS (GIBCO), 1% Penicillin-Streptomycin (P/S, 10,000 U/mL, Gibco), 1% L – Glutamine (200mM, Gibco), 1% 2-Mercaptoethanol (50μM, Gibco) and 1% MEM Non-Essential Amino Acids (Gibco) and 1x LIF(Sigma-Aldrich). To edit PD-1, we designed sgRNA targeting the mouse PD-1 3'UTR in regions devoid of existing MREs (as predicted by

miRanda). We selected 5 target sequences with low predicted off-target probabilities (crispr.mit.edu). We cloned the sgRNAs into the Cas9-2A-Puro plasmid, pX459(A gift from Feng Zhang, Addgene plasmid #62988) following the Zhang lab protocol[121]. We designed single stranded oligonucleotide HDR donors (ssODNs) comprising an MRE and BamHI recognition site flanked by 45nt arms homologous to to the region adjacent to the predicted Cas9 cleavage position in the PD-1 3'UTR (IDT). Each pX459 plasmid (5μg) was co-electroporated with 2μl of the associated ssODN (100μM) into 1 x 10^6 mESCs using the Neon Transfection System (1400V, 10ms, 3 pulses, Thermo Fisher). After 48 hours, electroporated cells were selected with 600ng/mL puromycin (GIBCO) for an additional 48 hours.

2.14 Analysing genome editing efficiency

We incubated transfected mESCs overnight in lysis buffer (0.1% SDS(Sigma), 40μg/mL RNase A (Thermofisher), 400μg/mL Proteinase K (Bioline)) at 37°C to extract gDNA for PCR. The resulting crude lysates were either used directly for PCR or subjected to ethanol precipitation. For primary OT-I T-cells, we incubated cells in 50mM NaOH (Sigma) for 10 minutes at 95°C to extract gDNA. We neutralised the NaOH with 1/10^{th} volume of 1M TRIS-HCl, pH 8.0 (Sigma) and used the neutralised lysate as input for PCR. To assess genome editing, we PCR amplified a 100 to 500bp region around the sgRNA target sequence using Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB). For PD-1 sgRNAs A and B we used primers MMU_PD1_Seq1 and MMU_PD1_Seq4. For PD-1 sgRNAs D, E and F we used primers MMU_PD1_Seq10 and MMU_PD1_Seq11. To check for HDR, we digested the PCR products with BamHI (NEB). To check for

NHEJ, we re-annealed the PCR products and digested any resulting heteroduplexes with T7 Endonuclease I (NEB) according to the manufacturer's instructions. We visualised the digestion products by gel electrophoresis in the presence of GelRed Nucleic Acid Gel Stain (Biotium).

2.15 CRISPR/Cas9 genome editing in OT-I T-cells

Primary splenocytes were harvested from C57BL/6, OT-I mice and stimulated with SIINFEKL peptide (20μg/mL Cambridge Peptides) and IL-2 (10 units/mL BioLegend) in RPMI-1640 (Gibco) supplemented with 10% FBS (Gibco), 1% P/S, (10,000 U/mL, Gibco), 10mM HEPES (Life Technologies), 1mM Sodium Pyruvate (Life Technologies), 50 μM 2-Mercaptoethanol (Gibco) and 1% MEM Non-Essential Amino Acids (Gibco). After 48 hours, $CD8^+$ T-cells were isolated using the mouse $CD8a^+$ T-cell Isolation Kit (Miltenyi Biotec).

We used Cas9 RNPs to edit primary mouse OT-I T-cells. To produce sgRNA, we first used Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB) and the primers sgIVT-MMU-PD1-D and sgIVT_R to PCR amplify the sgRNA scaffold from pX459-sgPD1-D. We used the resulting PCR product as an IVT template for the MEGAscript T7 Transcription Kit (Invitrogen), in accordance with the manufacturer's instructions. We purified the IVT reactions using a 10% polyacrylamide TBE-Urea gel (Thermo Fisher). The RNA was recovered from the polyacrylamide gel by overnight extraction in 500mM ammonium acetate (Thermo Fisher) followed by ethanol precipitation. We electroporated $1x10^6$ T-cells with 30μg recombinant Cas9 (NEB), 40μg purified sgRNA and 1μl ssODN (100μM)

using the Neon Transfection System (1600V, 10ms, 3 pulses, Thermo Fisher). T-cells were returned to warm media for 48 hours prior to analysis.

2.16 B16-F10 melanoma / T-cell co-cultures

To study how antigen levels influence T-cell activation and cellular fitness *in vitro* we co-cultured our B16-F10 OVA-miSFIT cell lines with OT-I T-cells. T-cells were harvested and stimulated as described above. For melanoma fitness experiments, 20,000 B16-F10 cells (approx. 50/50 mixture of B16-F10 OVA- cells and OVA^+, miSFIT cells) were seeded per well in a 96 well plate (n = 3 per cell line). After allowing B16-F10 cells to adhere for 3 hours, OT-I T-cells were added to each well at different T-cell : B16-F10 ratios. Mixed cultures were incubated overnight and analysed by flow cytometry. Relative fitness was calculated by dividing the frequency of $NGFR^+$ cells in the +T-cell condition by the frequency of $NGFR^+$ cells in the no-T-cell condition. For T-cell activation experiments, stimulated T-cells were rested for 72 hours prior to being co-cultured with individual OVA-miSFIT cell lines at an 8:1 T-cell to B16-F10 ratio (n = 5 per cell line). After 24 hours, we analysed T-cells by flow-cytometry. Antibody clones and suppliers are listed in the appendix of this thesis.

2.17 In vivo tumour growth assays

For *in vivo* tumour growth assays, 150,000 B16-F10 OVA-miSFIT cells were intradermally injected into WT C57BL/6 recipient mice (n = 6 recipient mice per cell line) on day 0. We isolated OT-I T-cells and stimulated them for 48 hours (see above and Fig. 4a) and intravenously injected 500,000 $CD8^+$, OT-I T-cells per

recipient mouse on day 7. Following the T-cell infusion we measured tumours every second day using Digital Callipers (Fisher Scientific). The experimenter performing the measurement was blinded to tumour identity. We culled mice when tumours exceeded 95mm^2 using approved methods. Mice that did not have detectable tumours by day 17 were excluded from the study.

For TIL analysis, tumours were injected as described above but OT-I T-cells were adoptively transferred on day 8 to reduce the likelihood of complete tumour clearance. On day 13 all mice were culled and spleens and tumours were harvested by dissection. Spleens were processed as described above and tumours were dissociated using the Tumor Dissociation Kit, mouse (Miltenyi Biotec). Cells were washed, blocked with TruStain fcX (Biolegend) and stained with antibodies listed in the appendix. Animal experiments were conducted under the constraints of a project licence approved by an internal Oxford review board and the UK home office.

Chapter 3: Tuning gene expression levels with miRNA response elements

New methods for finely-tuning gene expression levels in mammalian cells have the potential to improve our understanding of biological processes. In this chapter I describe our efforts to develop and test a gene tuning system based on redirecting miRNAs to a user-specified gene of interest. I show that altering the sequence of a MRE and consequently, its complementarity with an effector miRNA, provides an effective method of tuning gene expression output levels.

3.1 A fluorescent reporter system for evaluating MRE kinetics

Altering the complementarity of an MRE to an endogenous miRNA could enable control over the strength of ensuing gene-repression. A comprehensive study conducted by Wee *et al*. explored how di-nucleotide mismatches spanning the length of an siRNA influence the kinetics of Ago2-catalyzed mRNA cleavage in cell lysates[118]. To extend these findings to live cells, and to analyse the impact of base changes in an MRE rather than an siRNA or miRNA, we developed a fluorescent reporter-base cellular assay. We reasoned that introducing nucleotide substitutions in the MRE would allow us to control the strength of repression instigated by an endogenous miRNA, obviating the need for delivering exogenous siRNAs which are known to cause broad off-target effects[73].

To compare the strength of MRE variants, we generated a mammalian expression plasmid comprising a bi-directional CMV promotor that drives expression of two fluorescent proteins, iBlue[122] and ECFP (Fig. 3.1a). A unique MRE is introduced into the 3'UTR of ECFP while iBlue serves as a non-targeted internal control (Fig. 3.1a). We also appended degradation tags derived from ornithine

decarboxylase[123] to both fluorescent proteins to promote rapid protein turnover, making steady-state protein levels dependent on transcript abundance and translational efficiency.

In the first implementation of this assay, we co-transfected a miRNA expression vector alongside the MRE reporter plasmid. This allowed us to verify that any observed repression is miRNA dependent. We chose miR-21 because it is a well characterized miRNA that plays an important role in differentiation and oncogenesis[124], and because it is expressed at very low levels in HEK 293T cells[95]. We designed a second bi-directional expression vector that drives divergent expression of miR-21 and destabilized EYFP (Fig. 3.1.a). We cloned 320nt of the miR-21 primary transcript centred around the miR-21 hairpin to reproduce endogenous microRNA processing. In this vector, EYFP serves as a transfection control and allows us to monitor level of miR-21 expression in any given cell. We named the resulting plasmid p-Titre-miR-21.

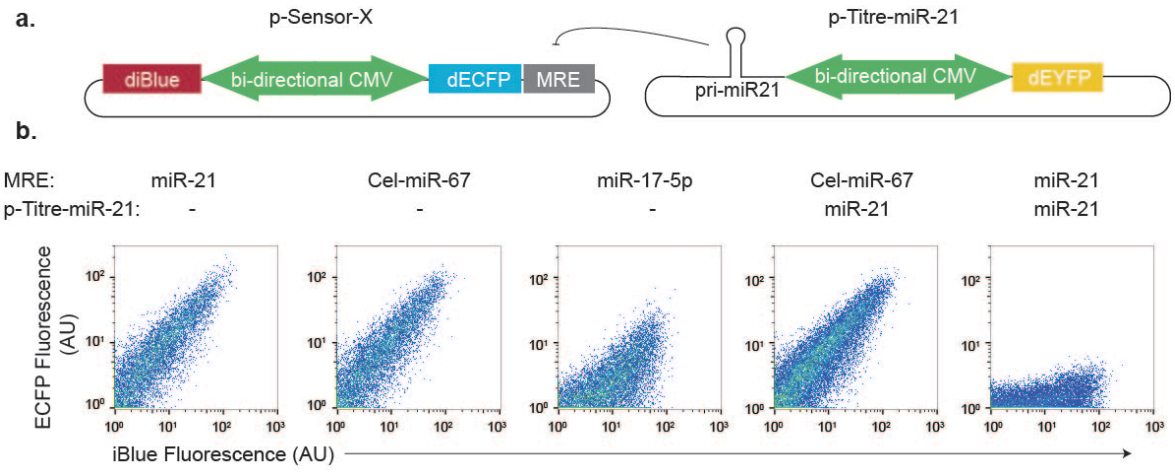

Figure 3.1. A dual plasmid system for measuring MRE functionality.
(a) Schematic representation of plasmids used to assess MRE activity (left) and to exogenously deliver miR-21(right). pri-miR-21 comprises a 320nt window of the miR-21 primary transcript centred around the mature miRNA. diBlue, dECFP and dEYFP denote versions of each fluorophore fused to a degradation tag **(b)** Flow cytometry plots of HEK-293T cells transfected with the indicated constructs. Plots are representative of three independent transfection replicates. Data reflects singlet events gated on FSC-A and SSC-A.

To validate this system, we generated a panel of control MRE reporters. We cloned versions of the reporter with MREs that have perfect complementarity to either miR-21, miR-17 (which is endogenously expressed at high levels in HEK-293T cells[95]) or the *C. elegans* miR-67 (Cel-miR-67) (which is not expressed in human cells[125]). We transiently transfected each reporter into HEK-293T cells alongside p-Titre-miR-21 or an empty control plasmid that does not express miR-21. After 24 hours, we monitored reporter expression by flow-cytometry (Fig. 3.1.b). The results of this analysis validated the specificity of the two-plasmid reporter system. In the absence of their cognate miRNAs, MRE reporters expressed comparable levels of ECFP and iBlue. An MRE with perfect complementarity to miR-17 reduced ECFP fluorescence without affecting iBlue, the internal control. Co-transfection of p-Titre-miR-21 repressed ECFP expression for the construct comprising a miR-21 MRE but not a construct bearing a Cel-miR-67 control MRE. Taken together, these observations support the validity of this reporter strategy for detecting the activity of endogenous or delivered miRNAs and confirm that expression of iBlue, our internal control, is independent of miRNA-MRE interactions in these experimental conditions.

3.2 Tuning reporter expression using synthetic MREs

Next, we asked whether introducing various di-nucleotide mismatches in the miR-21 MRE would predictably modulate the strength of miR-21-mediated target repression. In their 2012 study, Wee *et al*. assayed the effect of adjacent, di-nucleotide mismatches of the kinetics of miR-RISC cleavage activity[118]. The authors loaded siRNAs with distinct mismatches to let-7, a conserved miRNA, into Ago2 and incubated this complex with varying concentrations of radio-labelled

target mRNA[118]. The maximum rate of cleavage (Vmax) and the mRNA concentration required to achieve half-maximal cleavage (Km) were determined for each siRNA variant[118]. This study revealed that di-nucleotide mismatches between an siRNA and its cognate target can substantially alter the kinetics and magnitude of target slicing in cell lysates[118]. While this study exclusively analysed catalytic siRNA-mediated target cleavage, in the cellular environment, miRNA-mediated gene repression can also occur via transcript destabilization and translational repression[94]. Despite these additional complexities, we reasoned that the *ex vivo* kinetic parameters determined by Wee *et al*. might provide a starting point for controlling the strength miRNA repression in live cells.

In an effort to rationally tune reporter expression in live cells, we ranked all di-nucleotide mismatches from strongest to weakest based on their reported Vmax/Km in the Wee *et al*. study[118]. This value, known as the specificity constant, normalizes catalytic efficiency by differences in affinity across mRNA targets. We then selected 4 distinct di-nucleotide mismatches that spanned a broad range of specificity constants and designed miR-21 MREs with mismatches at the corresponding positions. Next, we cloned these 4 mismatched MREs into the bi-directional reporter plasmid (Fig 3.2.a). We hypothesized that these MREs would provide a range of ECFP expression output and labelled them MRE A through D based on their expected potency. Each di-nucleotide mismatched MRE, as well as a perfectly complementary MRE (which we labelled MRE E) and a Cel-miR-67 control MRE, were co-transfected into HEK-293T cells along with p-Titre-miR-21.

We assayed repression across three transfection replicates by flow cytometry (Fig 3.2.b) and fluorescent microscopy (Fig 3.2.c). Indeed, the panel of di-nucleotide miR-21 MRE variants that we selected conferred stepwise control over ECFP expression with a 25-fold dynamic range in mean fluorescent intensity (MFI) between the weakest and strongest level of repression. The weakest MRE, which contained mismatches at positions 8+9 with respect to the 5'end of miR-21, caused a 3.0% (SD = 0.7%) reduction in normalized reporter expression (MFI_{ECFP}/ MFI_{iBlue}) compared to the Cel-miR-67 control MRE. In comparison, a perfectly complementary MRE elicited a 74.4% (SD = 0.2%) reduction in expression. Mismatches at positions 5+6, 17+18 and 1+2 repressed reporter expression by 8.4% (SD = 3.1%), 13.6% (SD = 1.4%) and 48.8% (SD = 1.9%) respectively. These data suggest that introducing mismatches within a MRE sequence provides an effective method of modulating the strength of miRNA mediated gene silencing.

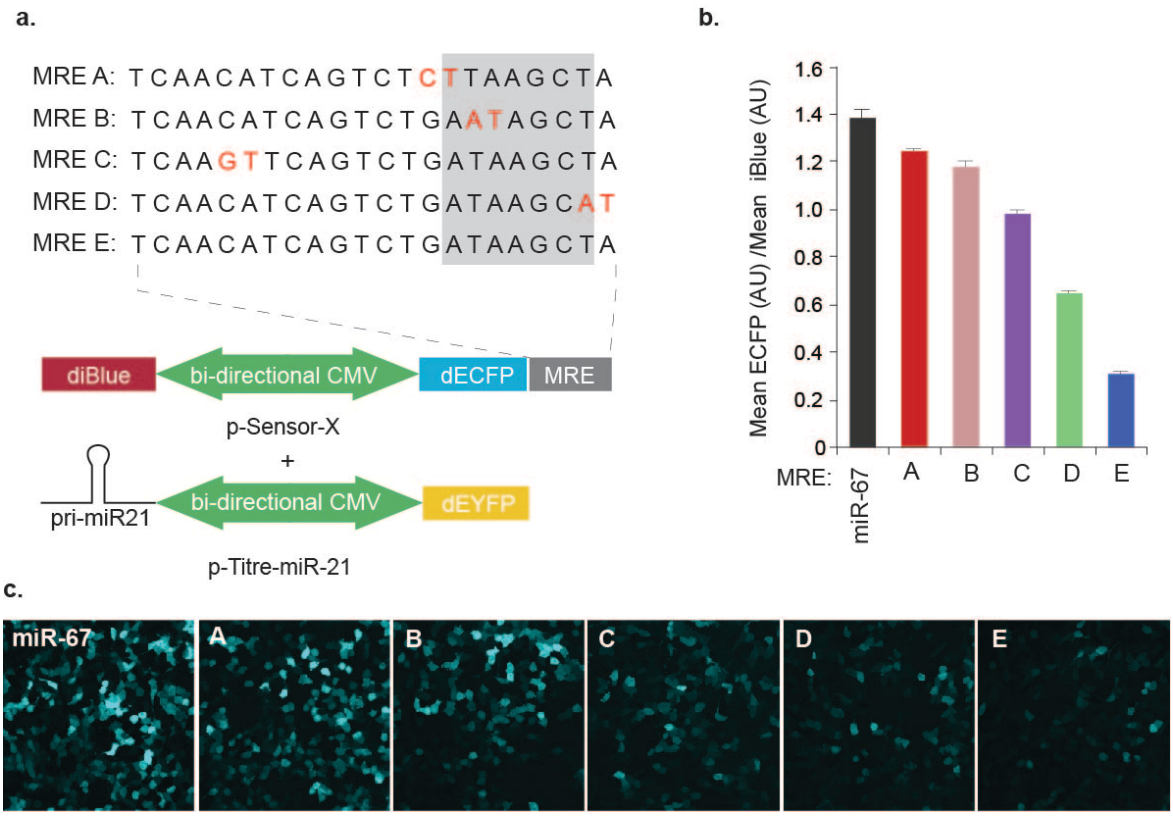

Figure 3.2. Tuning reporter expression with a panel of miR-21 MREs
(a) A panel of MREs with varying complementarity to miR-21. Mismatched di-nucleotides are shown in red. The region in the MRE corresponding to the miR-21 seed sequence is boxed in grey. p-Titre-miR-21 schematic adapted from Fig. 3.1. for convenience **(b)** Flow cytometry quantification of expression levels of each MRE sensor following co-transfection with p-Titre-miR-21 in HEK-293T cells (error bars indicate standard deviation of independent transfections, n=3). **(c)** Confocal fluorescent microscopy of HEK-293T cells co-transfected with p-Titre-miR-21 the indicated MRE sensor plasmid (25x objective, 405nm laser).

3.3 Dissecting the kinetics of miRNA silencing in live cells

We initially set out to develop a generalizable method of tuning gene expression that can be applied to a broad set of genes in multiple cell types. In chapter 3.2, we demonstrated that mismatched MREs can modulate ECFP expression in HEK-293T cells (Fig 3.2.b,c). These experiments were conducted using a fixed concentration of reporter plasmid and miRNA expression plasmid. In contrast, in a native cellular context, miRNA and mRNA expression levels vary by several orders of magnitude between genes and across cell types/states. For our MRE-

based tuning system to function broadly, it must be robust across a wide range of mRNA and miRNA expression levels[95,126]. Since our reporter and miRNA expression plasmids both comprise internal control fluorophores (iBlue and EYFP respectively), we reasoned that we could distinguish between cells that received high or low levels of each plasmid by flow cytometry. In previous studies, a miRNA reporter was combined with a non-silenced control to determine how mRNA levels influence the strength of endogenous miRNA mediated silencing[127,128]. We extended this approach to simultaneously measure the impact of miRNA and mRNA levels on the strength of a panel of mismatched MREs.

First, we asked whether EYFP levels correlate with miRNA concentration in HEK 293-T cells transfected with p-Titre-miR-21 (Fig. 3.3.a,b). We performed fluorescence activated cell sorting (FACS) to isolate cells expressing four different levels of EYFP. We extracted total RNA from each of these four cell pools and used a targeted RT-primer to reverse transcribe miR-21. We quantified miR-21 abundance using digital-droplet PCR (ddPCR) (Fig. 3.3.c). The strong correlation between EYFP fluorescence and miR-21 abundance (R^2 = 0.99, linear regression) confirmed that EYFP is an accurate proxy for miRNA concentration for future experiments.

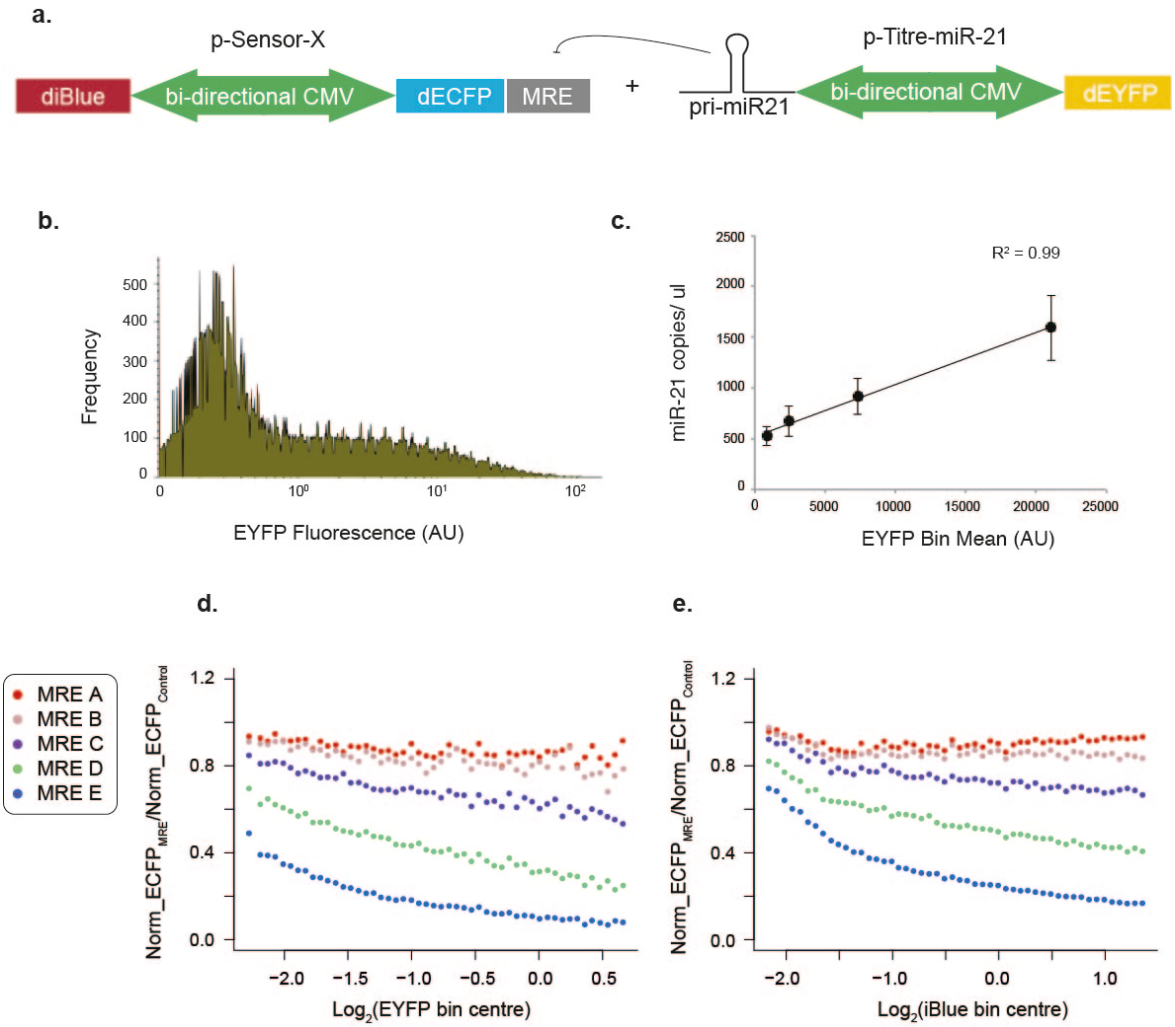

Figure 3.3. Dissecting the kinetics of miR-21 target repression.
(a) Schematic of miR-21 expression plasmid and MRE sensor reproduced from Fig. 3.1 for convenience. **(b)** Flow cytometry analysis of HEK-293T cells transfected with p-Titre-miR-21. Data is representative of n=3 independent transfections. **(c)** ddPCR quantification of miR-21 expression on cells pools sorted on the basis of EYFP expression (error bars indicate standard deviation, n=3, trend line and R^2 calculated from a linear regression model). **(d, e)** Computational binning analysis of flow cytometry data displays the effect of **(d)** transcript abundance and **(e)** miRNA levels on MRE reporter expression. Points are coloured by MRE and reflect the mean of all cells in the indicated bin. Norm_ECFP is the MFI of ECFP divided by the MFI of iBlue in each bin. Note that (b) reflects data acquired on a different flow-cytometer than the one used to sort cells for analysis in (c). Consequently, the X-axis values should not be compared directly.

Given this robust correlation, we re-analysed the flow cytometry data previously acquired for each of five mismatched MRE variants. We computationally divided cells into 50 bins based on their EYFP fluorescence and quantified ECFP

fluorescence in each bin (Fig. 3.3.d). We normalized these values to ECFP expression for a non-targeted MRE in each corresponding bin. This allowed us to determine how miR-21 concentration affected gene repression for each mismatched MRE. As expected, the strength of repression increased with miRNA concentration for all MREs in the panel. Encouragingly, the relative order of gene expression output for the MREs we tested remained constant over the broad range of miR-21 concentrations that we sampled. This finding suggests that MRE-based gene tuning is robust to changes in miR-21 levels.

Next, we asked how target mRNA abundance influences miRNA-mediated gene repression. In cell lysates, Ago2 catalysed mRNA cleavage obeys Michaelis-Mentin kinetics[118]. If miRNA repression exhibits Michaelis-Mentin kinetics in live cells, we would expect the rate of silencing to rise with increasing mRNA substrate levels. To test this hypothesis, we binned cells according to iBlue expression which serves as an indicator of how efficiently the reporter was delivered to each cell. In the unsilenced control, iBlue strongly correlates with ECFP (Fig. 3.1.b) supporting the validity of this dual reporter strategy. In the presence of di-nucleotide MRE variants, the magnitude of ECFP repression increased with increasing iBlue levels (Fig. 3.3.e). This observation supports the hypothesis that higher target abundance increases the rate of miRNA-mediated target repression. Furthermore, the relative order of repression elicited by the MREs in the panel was maintained over four orders of magnitude. This finding suggests that MRE-based tuning can be effective for target genes spanning a wide range of native expression levels.

This three-colour reporter system provides a useful method for assaying post-transcriptional gene repression at varying miRNA and target concentrations. This method provides insight into the kinetics and substrate dependency of miRNA silencing in live cells. Importantly, the data generated using this reporter system suggests that MREs with varying complementarity to miR-21 can tune reporter expression in an order that is constant over a broad range of miRNA and target levels in HEK-293T cells.

3.4 Validating the MRE reporter assay with an endogenous miRNA

The method we employed to simultaneously express a miRNA and a MRE reporter demonstrated the flexibility of MRE-based tuning across varying miRNA and target levels. However, this approach necessitates exogenous miRNA overexpression. A tuning system that responds to an endogenous miRNA would be advantageous for two key reasons. i.) although exogenous miRNAs are relatively easy to transfect in cell culture, they are more difficult to deliver efficiently *in vivo*. ii.) administering foreign miRNAs or siRNAs results in widespread off-target perturbations to endogenous gene expression programs[73]. By designing MREs with varying complementarity to an endogenous miRNA, we reasoned that it would be possible to predictably tune gene expression levels without delivering exogenous miRNAs.

miR-17 is a well-studied miRNA expressed in numerous cell types in humans and mice[95]. Importantly, this miRNA is expressed at high levels in HEK-293T cells[95]. We designed a panel of MREs with di-nucleotide mismatches to miR-17 at the same positions that we previously used to created mismatches to miR-21 (Fig.

3.4.a). We cloned these MREs downstream of destabilized EGFP in a bi-directional reporter construct that also comprises iBlue as an un-silenced internal control. We also cloned a perfectly complementary miR-17 MRE and two concatenated, perfect MREs (Fig. 3.4.a). We transiently transfected each of these constructs into HEK-293T cells in triplicate. After allowing endogenous miR-17 to act on the reporter transcripts for 24 hours, we analysed expression of EGFP under the control of each MRE variant by flow cytometry and fluorescent microscopy (Fig. 3.4.b,c). This panel of MREs conferred discrete, stepwise control of EGFP expression, confirming that it is possible to tune expression output using an endogenously expressed miRNA.

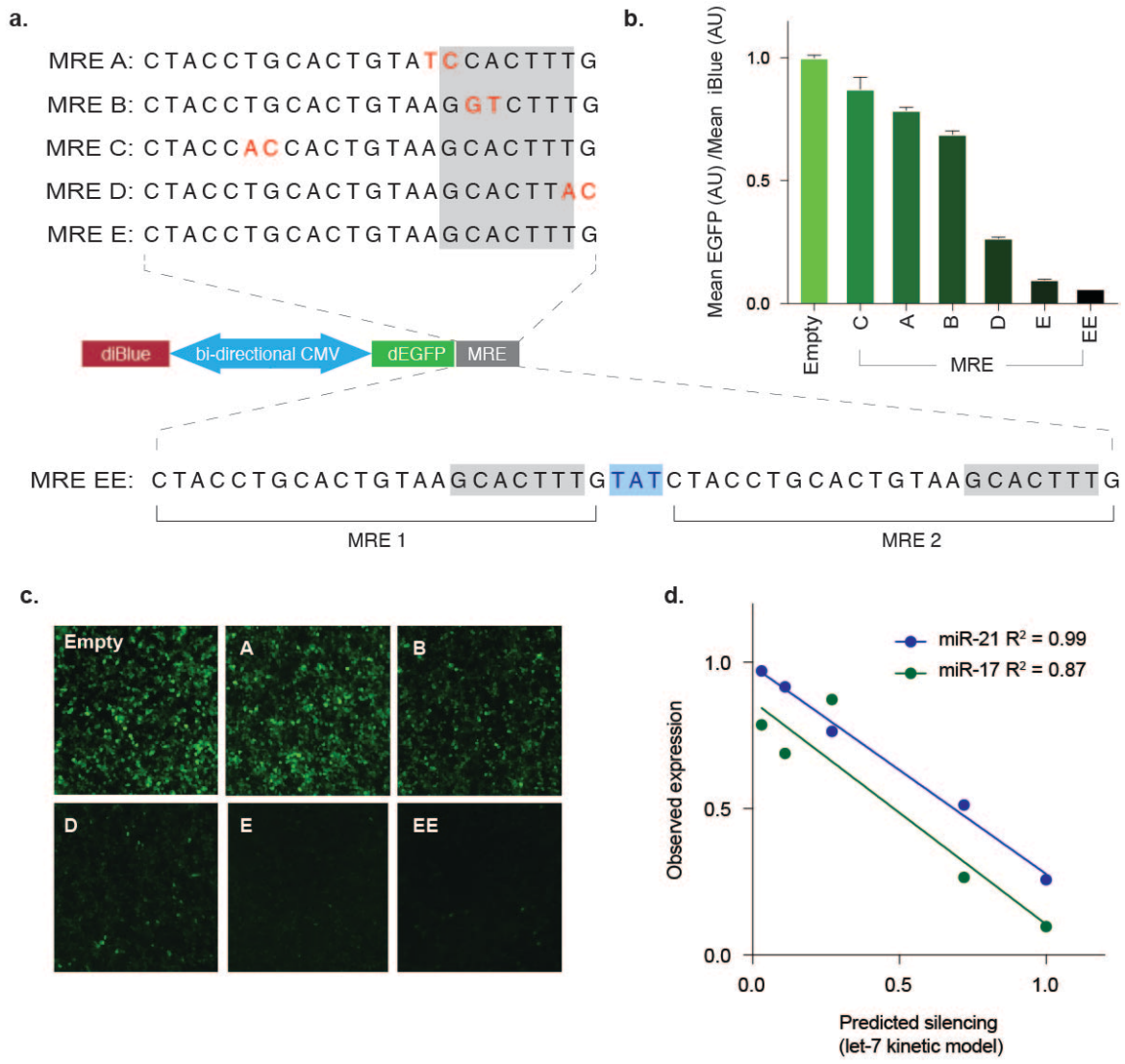

Figure 3.4. Tuning gene expression using MREs with varying complementarity to an endogenous miRNA.
(a) A panel of MREs with varying complementarity to miR-17. Complementary nucleotides are indicated in black. Mismatched di-nucleotides are shown in red. The region in the MRE corresponding to the miR-17 seed sequence is boxed in grey. MRE EE comprises two perfectly complementary miR-17 MREs separated by a spacer sequence highlighted in blue. **(b)** Flow cytometry quantification of reporter expression levels in HEK-293T transfected with each MRE (error bars indicate standard deviation of independent transfections, n=3). **(c)** Confocal fluorescent microscopy of HEK-293T cells co-transfected with p-Titre-miR-21 the indicated MRE sensor plasmid (25x objective, Argon multiline 458/488/514 laser). **(d)** Linear regression analysis comparing predicted and observed MRE repression. Observed expression was calculated by flow cytometry data for each MRE variant normalized to a no-MRE or Cel-67 MRE control. Predicted expression was calculated from the specificity constants measured by Wee *et al*. for let-7 siRNAs with mismatches to a target RNA at the same positions as mismatches between MREs A,B,C,D and E and their cognate miRNAs.

Interestingly, the order of repression instigated by mismatches in miR-17 MREs did not match the order invoked by mismatches to miR-21. Specifically, a miR-17 MRE with a dinucleotide mismatch at positions 17+18 caused two-fold weaker silencing than a miR-21 MRE with mismatches at the same position (P = 0.02, unpaired T-test, 2 tailed). This observation led us to question how well MRE-mediated silencing can be predicted on the basis of mismatch identity. We originally chose our panel of di-nucleotide MRE mismatches based on the kinetics of *ex vivo* miRNA mediated target cleavage as reported by Zamore and colleagues[118]. We returned to this original kinetic model and asked how well it forecasted miR-21 and miR-17 mediated gene repression in live cells. We performed a linear regression analysis comparing the reported specificity constant (Vmax/Km) to the observed reporter expression for each mismatched MRE. Although *ex vivo* kinetics were determined for a let-7 MRE, they efficiently predicted miR-21 mediated repression (R^2 = 0.99, linear regression) and, to a lesser extent, miR-17 mediated repression (R^2 = 0.97, linear regression) in live cells (Fig. 3.4.d).

3.5 Discussion

In summary, we first established a robust fluorescent reporter-based assay for monitoring the repressive capacity of MREs in live human cells (Fig 3.1). We next demonstrated that introducing di-nucleotide mismatches between a miRNA and MRE is an effective method of controlling the magnitude of gene repression (Fig. 3.2). The order of gene expression levels elicited by these mismatched MREs is robust across a range of transcript and miRNA abundance (Fig. 3.3). This tuning approach is also effective for MREs with varying complementarity to an

endogenous (rather than transfected) miRNA (Fig. 3.4). Finally, we showed that the strength of miRNA-mediated repression in live cells correlates with the rate of *ex vivo* miRNA-mediated transcript slicing (Fig 3.4.d).

In these experiments, we used flow-cytometry to quantify the extent to which each MRE represses a reporter gene. I chose flow cytometry because it displays several advantages over RT-qPCR or western blotting. Flow cytometry is a quick and simple method of measuring expression that is compatible with large sample sizes. Additionally, flow cytometry provides a single-cell level readout of gene expression, a property that allowed us to determine how MREs function across varying transcript and miRNA levels (Fig. 3.3). There are however limitations of using flow cytometry to measure miRNA activity. miRNAs have been shown to repress both the stability and the translational efficiency of targeted mRNAs[108,110,111]. We used flow cytometry as a measure of steady state protein levels. Consequently, our data reflects the cumulative effect of all repressive mechanisms and cannot distinguish between transcript degradation and translational repression. This topic is revisited in chapter 4, where we describe the development of a method capable of measuring each mechanism of miRNA-mediated repression independently, thus overcoming this limitation of flow cytometry.

In this chapter, we used siRNA slicing data as a predictor of MRE strength (Fig. 3.4). Although *ex vivo* kinetic data accurately predicted the strength of five MREs in our panel, these data only exist for the limited set of MRE mismatches that were analysed in the Wee *et al*. study[118]. Furthermore, even in this small test set, we

observed differences in repressive strength when comparing mismatches at the same position but for different miRNAs. On one hand, rationally designing and testing small sets of mismatched MREs using *ex vivo kinetics* is a simple and effective method for controlling the strength of gene-expression. On the other hand, this approach is limited in scope and may not generalise across miRNAs. To study a larger set of MREs and to improve the precision that can be achieved using MRE-based gene tuning, we decided to develop a screening platform for unbiased high-throughput analysis of thousands of miR-17 MRE variants.

Chapter 4: Dissecting the regulatory landscape of a microRNA target site

In an effort to precisely tune gene expression levels, we rationally designed a panel of four MREs with different mismatches to miR-17. In chapter 3, I demonstrated that these MREs conferred stepwise control over reporter expression. Unfortunately, not all of the MREs behaved as predicted by previous *ex vivo* mRNA cleavage data. Furthermore, the small panel that we tested does not provide adequate tuning resolution for sensitive applications. To improve our understanding of miRNA target selection in live cells, and to increase the precision and predictability of MRE-based gene tuning, we developed a high-throughout, unbiased MRE screening platform. In this chapter I outline the High-throughput Sequencing (HTS)-based method that we established for dissecting the regulatory landscape of MREs at single-nucleotide resolution. I will describe the results we obtained by applying this method to MREs with complementarity to miR-17 and miR-21 and explain how we adapted this assay to study miRNA-mediated translational repression. Finally, I describe the conventional methods we used to validate the accuracy of this high-throughput approach. By performing an unbiased screen, we were able to gain new insight into miRNA targeting and substantially improve the precision of MRE based gene tuning.

4.1 A high-throughput analysis of MRE regulation of transcript abundance

By evaluating the regulatory capacity of a library of MREs with varying complementarity to miR-17, we reasoned that we could dissect the targeting landscape of this miRNA. The resulting dataset could be used to select MREs of a

desired strength, providing a predictable approach for fine-tuning gene expression. One way to assess the effect of nucleotide variation on a functional element is to synthesize an oligonucleotide pool comprising a small percentage of the incorrect nucleobases at each position. Such “degenerate” sequence pools have been previously used to understand the fitness landscape of a tRNA gene[129].

We designed and synthesized a 23nt degenerate oligonucleotide pool with 91% complementarity to miR-17 and 3% of each alternative nucleotide at every position (Fig. 4.1.a). Using the binomial probability equation, we predicted that incorporating 9% total degeneracy at each position would maximize the probability of recovering single, and di-nucleotide mismatched MREs (Fig. 4.1.c, see materials and methods for more information). This degenerate oligo pool was cloned downstream of a fluorescent reporter (ECFP) in a mammalian expression plasmid and the ensuing MRE variant library was transfected into HEK-293T cells that endogenously express miR-17 (Fig. 4.1.b). We also co-transfected a control reporter bearing an MRE complementary to C. elegans Cel-miR-67, which is not expressed in human cells[125]. After waiting 24 hours for endogenous miR-17 to act on the transcripts templated by the variant, we harvested mRNA and plasmid DNA (pDNA) and subjected them to HTS (Fig. 4.1.b,d,e).

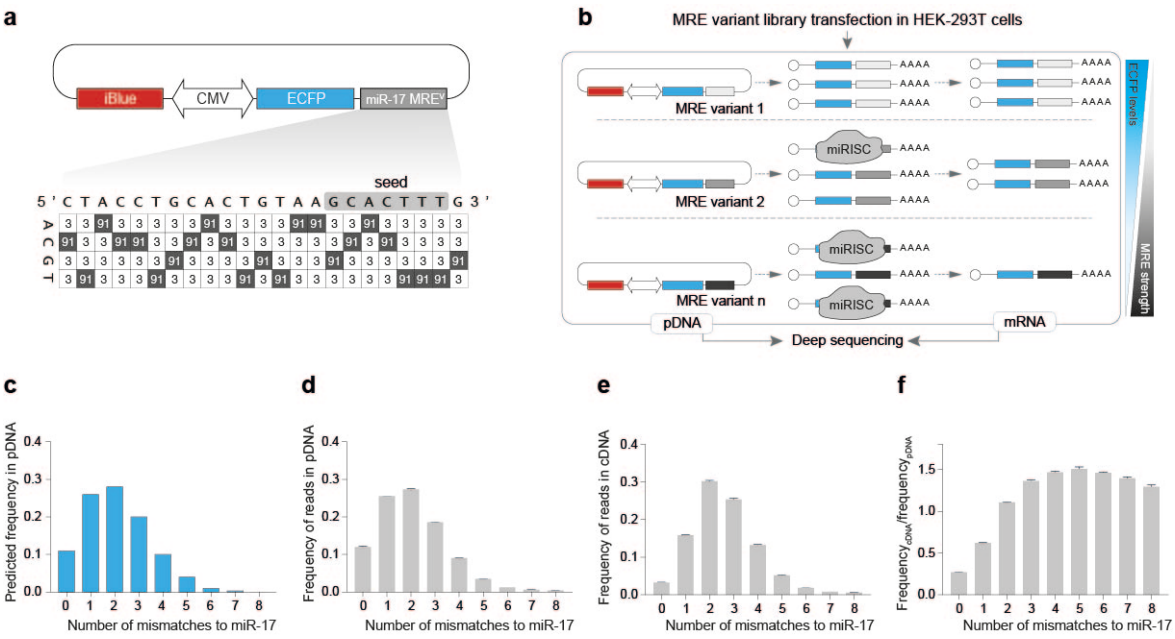

Figure 4.1. High-throughput analysis of a MRE regulatory landscape.
(a) MRE reporter library diagram. Values indicate the proportion of nucleotides at each position in the MRE (shaded squares = nucleotides complementary to miR-17). **(b)** Schematic of the MRE regulatory landscape analysis pipeline. **(c)** Expected probability distribution for the number of mismatches per MRE in the miR-17 MRE variant library. **(d)** Frequency of sequencing reads containing between 1 and 8 mismatches to miR-17 in plasmid DNA (pDNA) from HEK 293T cells transfected with the MRE variant library (n = 3 biological replicates, mean +/- standard deviation). **(e)** Frequency of reads in cDNA libraries from the same pools of cells as in (d). **(f)** Average enrichment for MREs with varying numbers of mismatches calculated by normalizing (e) to (d).

Sequencing the plasmid DNA (pDNA) recovered from the HEK-293T cells transfected with the MRE library confirmed that the actual distribution of mismatches in the plasmid library agreed well with our intended binomial distribution (compare Fig. 4.1.c and d). The degeneracy in the MRE oligonucleotide pool was neither bottle-necked nor skewed during cloning and transfection. When we sequenced the cDNA recovered from the same cells, we observed an enrichment in MREs with a greater number of mismatches as compared to pDNA (compare Fig. 4.1.d and e). This is consistent with miR-17 preferentially targeting highly complementary MREs. To quantify this apparent enrichment and to estimate the strength of the MRE variants present in the library,

we divided the frequency of each variant in the mRNA pool by its frequency in the pDNA pool (Fig. 4.1.f). As expected, MREs with higher complementarity to miR-17 were silenced more effectively (Fig. 4.1.f). Even single nucleotide mismatches diminished silencing by 2.30-fold on average (+/- 0.03, 95% CI) compared to a perfectly matched target (Supplementary Fig. 1). On average, MREs with 4 or more mismatches were almost completely non-functional. These data confirm that this high-throughput assay is capable of capturing miRNA-mediated gene repression at the mRNA level. Furthermore, this method has the sensitivity to discern MRE variants of different strengths.

4.2 The functional landscape of miR-17 target selection

After broadly assessing the impact of increasing numbers of mismatches on MRE strength, we sought to explore miR-17 targeting preferences in more detail. We focused our subsequent analysis on single nucleotide variants and asked how each position within the MRE contributes to miRNA-mediated repression (Fig. 4.2.a). We calculated the average enrichment ($frequency_{cDNA}/frequency_{pDNA}$) caused by mismatches at each position in the MRE. Should complementarity at a given position be important for repression, then introducing a mismatch at that position should diminish repression and increase mRNA abundance.

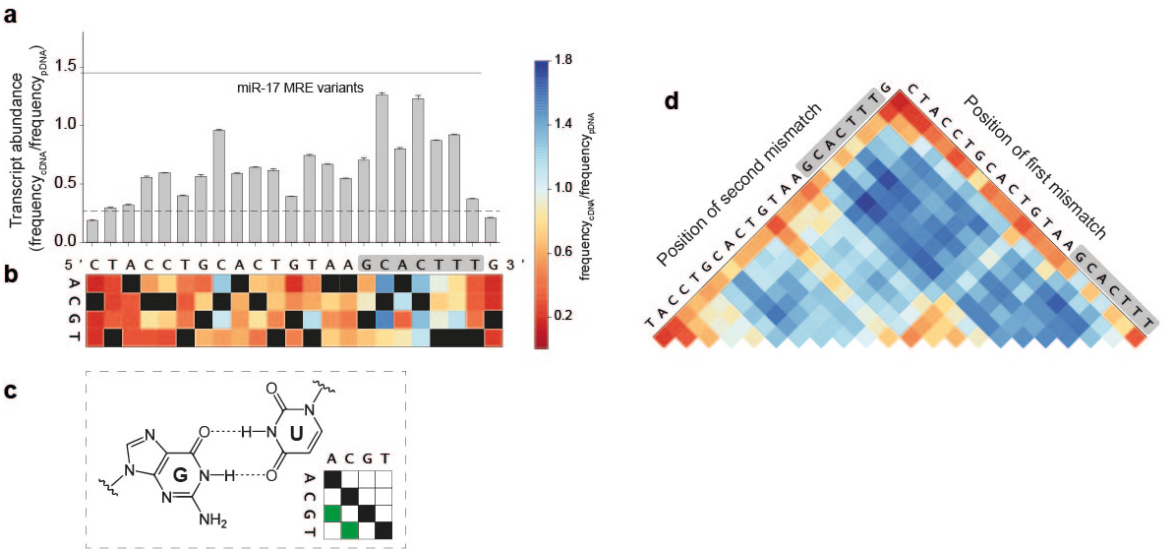

Figure 4.2. The single-nucleotide regulatory landscape of a miR-17 MRE. **(a)** Impact of MRE variants on transcript abundance. Bar graph shows relative contribution of each nucleotide to MRE function, as determined by high-throughput sequencing (n = 3 biological replicates, mean +/- standard deviation; dashed line = expression of a perfectly complementary MRE, solid line = expression of a non-targeted MRE- Cel-miR-67). **(b)** Heat-map displays the effect of each possible mismatch by position and reflects the mean of three replicates (complementary bases are displayed in black). **(c)** Schematic of G:U wobble pairing. The possible base changes that result in G:U wobble pairs are depicted in the inset. **(d)** The impact of di-nucleotide substitutions on reporter expression (mean of 3 biological replicates; colour scale is the same as in (c); grey box = seed region).

As anticipated[5,96,118], certain seed mismatches strongly abrogated silencing, confirming the important role of this region in target selection (Fig. 4.2.a). Intriguingly however, non-seed nucleotides also had a significant impact on the degree of repression. One position even had a greater impact on silencing than most seed nucleotides (Fig. 4.2.a). Complementarity with the 3'-most proximal base of the MRE did not improve silencing, consistent with the observation that the 5'-nucleotide in a miRNA is embedded within RISC and unavailable for base-pairing interactions[130,131] (Fig. 4.2.a). Complementarity with the first three bases from 5'end of the MRE did not play an important role in repression. This finding is also in agreement with previous studies[117,118].

We further hypothesised that distinct base substitutions at the same position could have differing effects on gene silencing. To test this hypothesis, we computationally divided all single-nucleotide MRE variants by the position and base identity of their mismatch (Fig. 4.2.b.). Indeed, not all base changes have the same impact on MRE functionality. In addition to the canonical Watson-Crick base pairs, the RNA bases uracil and guanine can also form a thermodynamically stable pair known as a G:U wobble[132] (Fig 4.2.c.). Mutations that introduce G:U wobbles were always less deleterious to silencing than non-pairing base changes, highlighting the importance of thermodynamic stability for miRNA-mediated repression (Fig. 4.2.b).

Next, we examined the double-nucleotide variants in the MRE library (Fig 4.2.d). Pairs of mismatches were generally more deleterious to silencing than either individual mismatch alone. Specifically, our analysis revealed that pairs of mismatches within the seed or combinations of seed mismatches with mismatches in positions 14 to 20 strongly impaired miRNA activity (Fig. 4.2.d).

The high-throughput approach employed here provides an empirical view of the miR-17 targeting landscape, revealing how each nucleotide, and pair of nucleotides, contribute to repression. Both seed and non-seed bases make important contributions to silencing and G:U wobble pairs are tolerated better than non-pairing base substitutions.

4.3 Different miRNAs display distinct targeting preferences

The data that I have described thus far reflects the targeting preferences of one endogenous miRNA, miR-17. Because miR-17 belongs to a family of miRNAs that share the same seed-sequence[133], it is possible that these results may be confounded by the activity of other miRNAs in HEK-293T cells. To overcome this limitation, and to determine whether the targeting preferences that we observed for miR-17 generalize to other miRNAs, we decided to analyse the targeting preferences of an exogenously delivered miRNA, miR-21.

Using the same approach that we previously applied to miR-17, we designed an oligonucleotide pool that is largely complementary to miR-21 but comprises 9% total degeneracy at each position (Fig 4.3.a). We cloned this oligonucleotide pool into a reporter construct to generate a miR-21 MRE variant library. Because HEK-293T cells basally express very low levels of miR-21[95], we co-transfected a miR-21 expression vector (p-Titre-miR-21) along with the MRE variant library (Fig 4.3.a). We allowed the exogenously delivered miR-21 to act on the transcripts templated by the plasmid library for 24 hours. Next, we harvested mRNA and pDNA from the transfected cells and subjected them to targeted deep sequencing. We calculated a normalized expression value for all single-nucleotide miR-21 MRE variants by dividing their frequency in the cDNA pool by their frequency in the pDNA pool.

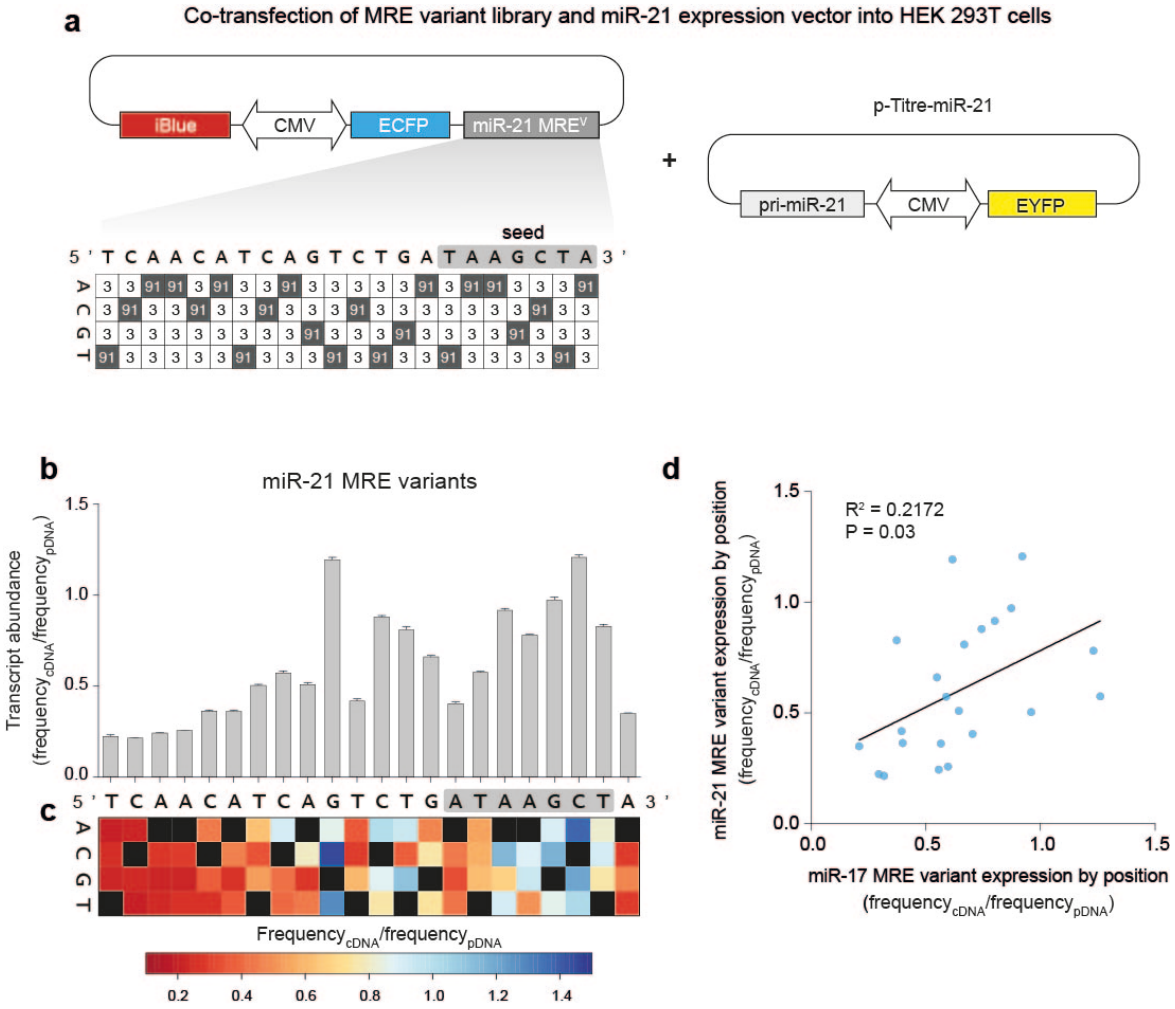

Figure 4.3. The single-nucleotide regulatory landscape of a miR-21 MRE.
(a) Schematic representation of the miR-21 MRE reporter library design. Values indicate the proportion of that nucleotide at each position in the MRE (shaded squares = nucleotides complementary to miR-21). Since miR-21 is expressed at low levels in HEK-293T cells, p-Titre-miR-21 was used to transgenically overexpress miR-21. **(b)** Impact of MRE variants on transcript abundance. Bar graph shows the effect of single-nucleotide mismatches at each position as determined by dividing cDNA/pDNA sequencing reads for each type of variant (n=3 biological replicates, mean +/- standard deviation). **(c)** Heat-map displays the effect of each possible mismatch by position and reflects the mean of three replicates (complementary bases are filled in black). **(d)** Linear regression comparing the impact of mismatches at the same position in miR-17 vs. miR-21 MREs.

We focused our analysis on single nucleotide MRE variants and asked how each position within the MRE contributes to miR-21 mediated mRNA degradation (Fig 4.3.b). Interestingly, miR-21 displayed distinct targeting preferences from miR-17. Mismatches at nucleotides 17 through 20, with respect to the 5'-end of miR-21, did

not strongly impair silencing, indicating that these bases are not very important for miR-21 mediated repression (Fig 4.3.b). In contrast, mismatches at nucleotides 9,10 and 11 increased reporter expression substantially, suggesting an importance for these bases in target recognition (Fig 4.3.b). In addition to these notable differences, there were some similarities in the regulatory landscapes of miR-21 and miR-17. In both cases, seed complementarity was important for regulation. But what was also true for both miRNAs, is that mismatches at certain non-seed positions could strongly impair silencing. In fact, for miR-21, base pairing at position 13 was more important for repression than all-but-one seed position (Fig 4.3.b). Notably, the current dogma in the miRNA field states that seed complementarity is the most important determinant of repression[94,96,117,118]. This assumption also lies at the foundation of most miRNA target prediction algorithms. Therefore, the strong consequences of non-seed mismatches that we observed in this study were largely unexpected.

Another similarity between miR-17 and miR-21 is the apparent impact of G:U wobble pairs on repression. We assessed the impact of different base changes at the same position in the miR-21 MRE (Fig 4.3.c). As we previously observed for miR-17, base changes that introduced G:U wobble pairs to miR-21 were always less detrimental to mRNA repression than changes to non-pairing bases (Fig 4.3.c).

Finally, we compared the relative importance of each position between the two MREs. We performed a linear regression analysis comparing the average expression for single nucleotide variants at the same position for miR-17 vs miR-

21. We observed a weak but statistically significant correlation between these two datasets (R^2 = 0.2172, P = 0.03). This analysis revealed miRNA specific targeting preferences that may not be predicted by generic algorithms. These results suggest that tuning gene expression with MRE libraries may require further high-throughput analysis if users would like to use trigger miRNAs other than the ones we have already studied.

To summarise this sub-chapter, we analysed a library of MRE variants with differing complementarity to miR-21. We were able to confirm that non-seed nucleotides are important for strong miRNA mediated repression. This data also showed that G:U wobble pairing improves silencing compared to non-pairing bases. Finally, we demonstrated that miR-17 and miR-21 rely on complementarity at distinct positions to impart strong repression.

4.4 Translational repression and transcript degradation are correlated across MREs

A HTS-based assay allowed us to assess the effect of miRNA / MRE mismatches on mRNA stability. In addition to promoting transcript degradation, miRNAs have also been shown to repress translation[25,107,108,110,111] (Fig. 4.4.a). However, the extent to which these two processes are correlated across different MREs remains unclear. We hypothesized that miRNA-mediated translational inhibition may be a function of MRE sequence identity. To determine the degree to which the MRE variants in the library mediate translational repression, we optimized a polysome profiling approach to isolate transcripts on the basis of their translation status. If

certain MREs are being translationally repressed, we expect their transcripts to be associated with fewer ribosomes than unrepressed MREs.

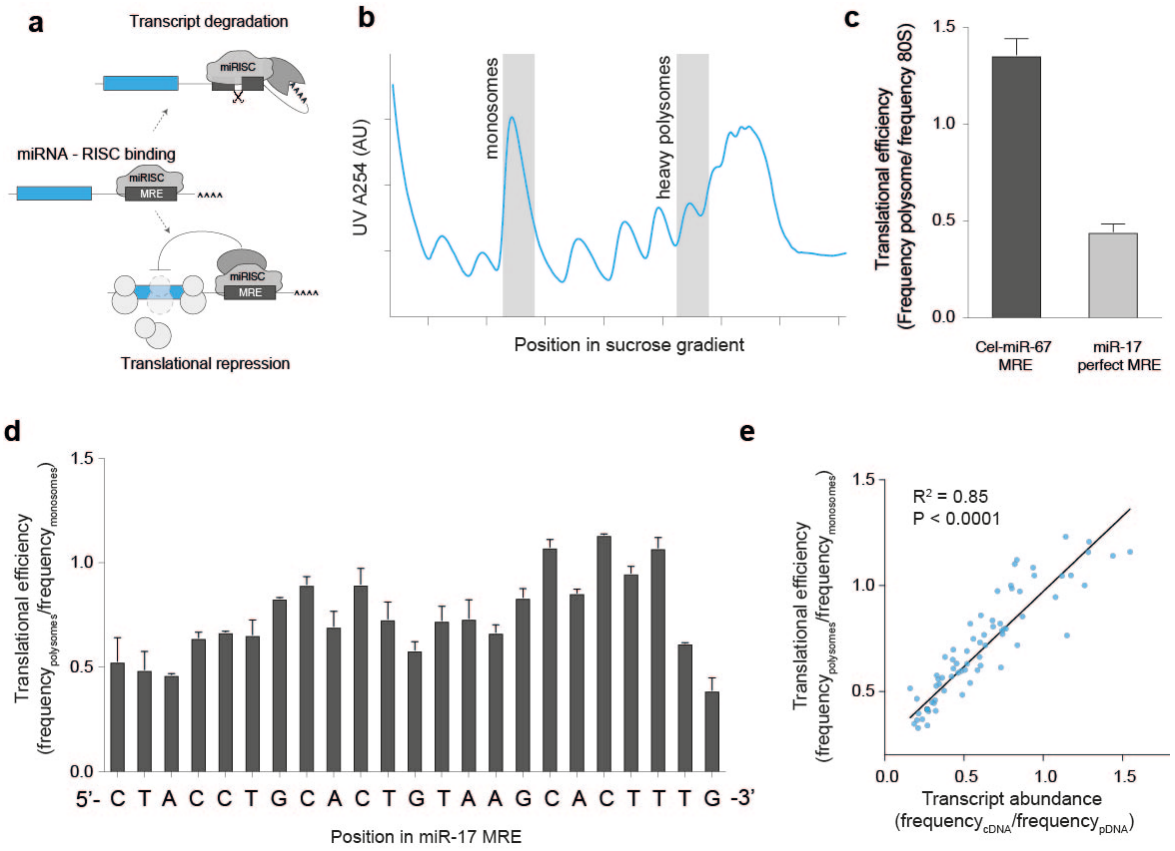

Figure 4.4. The impact of MRE sequence on mRNA stability and translational efficiency are correlated.
(a) Schematic representation of the two major pathways underlying miRNA-mediated repression. **(b)** Polysome profiles generated by sucrose gradient fractionation. Blue trace denotes the spatial distribution of RNA across the gradient as monitored by UV absorbance. Analysed fractions (monosomes and heavy polysomes) are shaded in grey (data is representative of two biological replicates). **(c)** Comparison of translational efficiency between targeted (miR-17) and untargeted (Cel-miR-67) MREs (error bars denote the range of two biological replicates). **(d)** Impact of MRE variants on translational efficiency. Bar graph shows the effect of single-nucleotide mismatches at each position as determined by dividing polysome/monosome sequencing reads for each type of variant (error bars denote the range of two biological replicates). **(e)** Correlation between translational efficiency and transcript stability for all single nucleotide miR-17 MRE variants (P value reflects test that the slope of a linear regression model significantly differs from 0).

By treating cells with cycloheximide (CHX) it is possible to fix ribosomes to the mRNAs that they are translating[25]. We transfected the miR-17 MRE plasmid

library and a control Cel-miR-67 MRE reporter into HEK-293T cells. After 24 hours, we arrested translation with CHX and lysed the cells in hypotonic buffer. In order to separate RNAs by ribosome density, we subjected the lysates to sucrose gradient ultracentrifugation. We fractionated the resulting gradients and monitored RNA distribution by measuring UV absorbance (Fig 4.4.b). The polysome profiles that we generated allowed us to clearly resolve the small and large ribosomal subunits, as well as monosomes, di-somes and heavier polysomes (Fig 4.4.b).

Previous experiments have shown that translationally repressed mRNAs are depleted from heavy polysome fractions[25]. To validate this finding for miR-17 MREs, we isolated monosome-bound and heavy polysome-bound mRNAs from the library-transfected cells (Fig. 4.4.b). cDNA was generated from both fractions and amplicon libraries containing the variable MRE were subjected to high-throughput sequencing. We used the ratio of reads in the heavy polysome-bound fraction to reads in the monosome-bound fraction as a measure of translational efficiency for each MRE variant. First, we compared translational efficiency between transcripts comprising a Cel-miR-67 MRE vs an MRE that has perfect complementarity to miR-17 (Fig 4.4.c). The miR-17 MRE was substantially depleted from the heavy polysome fraction compared to the monosome fraction, demonstrating that this MRE is translationally inhibited.

Next, we asked how different positions within the miR-17 MRE contribute to miRNA-mediated translational repression (Fig 4.4.d). We observed a pattern that reflected the previous transcript degradation experiment (compare Fig 4.4.d to Fig 4.2.a). Interestingly, positions that were important for miR-mediated mRNA

destabilisation also appeared to be important for translational inhibition. To quantify the extent of this relationship, we performed a linear regression comparing the effect of all possible single-nucleotide MRE variants on transcript stability vs translational enrichment (Fig 4.4.e). This analysis revealed a strong, significant correlation between these two mechanisms of repression for single-nucleotide variants in the library (Fig. 4.4.e) (R^2= 0.85, P < 0.0001, linear regression). This finding suggests that miRNA-target base pairing is a critical determinant of the magnitude of both transcript degradation and translational repression. These data also indicate that the mRNA/pDNA sequencing approach is a good predictor of overall MRE strength.

4.5 Validating high-throughput MRE analysis using conventional expression quantification methods

To evaluate the accuracy of our HTS-based MRE functional analysis platform, we sought to benchmark this approach against routinely using methods of measuring gene expression. To this end, we compared the sequencing assay to RT-qPCR and flow cytometry, methods that can provide accurate measurements of mRNA and protein levels respectively. To compare these assays, we chose a panel of 15 single nucleotide and di-nucleotide miR-17 MRE variants. In order to avoid bias in our choice of MREs, we transformed the miR-17 MRE variant library into *E. coli* and randomly selected and sequenced colonies until we recovered 15 unique MREs. We also included a perfectly complementary miR-17 MRE and a Cel-miR-67 MRE as controls in this validation panel.

We searched for each of the variants from the validation panel in the pDNA and cDNA HTS data that we previously obtained. We then calculated the relative

expression of each variant by dividing the number of corresponding reads in the cDNA library by the number of reads in the pDNA library (Fig. 4.5.a). The randomly selected MREs in this validation set covered a broad range of expression levels representative of the variant library as a whole (Fig. 4.5.a). Next, we independently transfected the reporter constructs comprising each validation MRE into HEK-293T cells in triplicate (Fig 4.5.b). Because each reporter construct encodes ECFP under the control of the MRE variant and iBlue as an un-silenced control, we could readily assess MRE-mediated repression by flow cytometry (Fig 4.5.b). We calculated the expression of each variant as MFI_{ECFP}/MFI_{iBlue} and compared these values to the HTS readout. Linear regression analysis revealed a strong, significant correlation between high-throughput sequencing and flow cytometry measurements of MRE activity (R^2 = 0.95, $P < 0.0001$, Fig. 4.5.c).

Next, we extracted mRNA from HEK 293T cells transfected with each variant. We performed reverse transcription (RT) followed by quantitative PCR (qPCR) with primers targeting iBlue and ECFP for each sample. We used the $\Delta\Delta$ cycle-of-threshold (Ct) method to normalize expression of all MRE variants to a Cel-67-MRE control reporter by comparing the Ct of ECFP to that of iBlue for each variant. RT-qPCR strongly validated high-throughput sequencing as a measure of MRE strength (R^2 = 0.92, $P < 0.0001$, Fig. 4.5.d). Finally, we compared the conventional measures of MRE strength, flow cytometry and RT-qPCR, to each other (Fig. 4.5.e). The data obtained from these two assays correlated strongly (R^2 = 0.97, $P < 0.0001$). Together these validation experiments show that the

sequencing approach we established provides precise, reliable quantification of MRE strength.

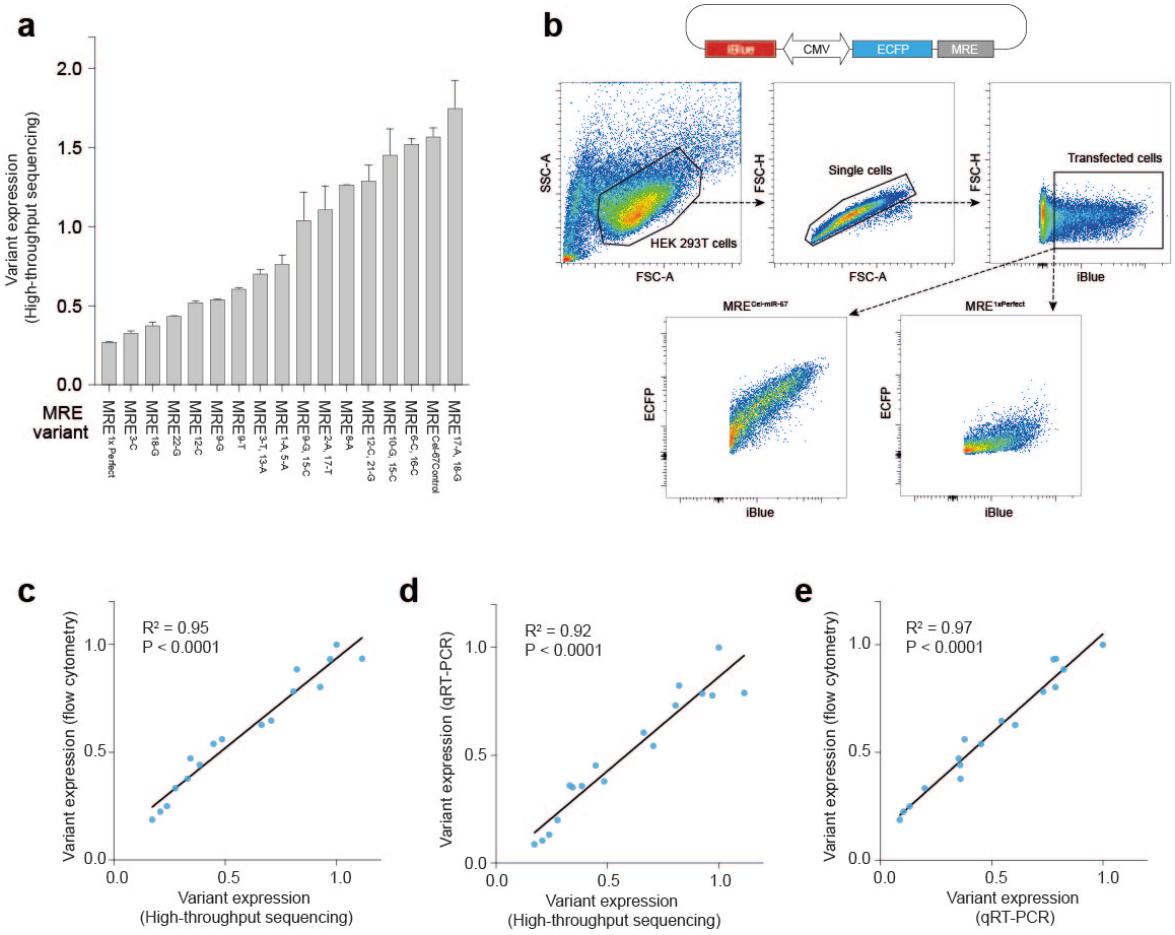

Figure 4.5. RT-qPCR and flow cytometry validate high-throughput MRE screen data.
(a) Ranked impact of 15 randomly selected candidate MRE variants as well as a perfectly complementary miR-17 MRE ($MRE^{1xPerfect}$) and a non-targeted control MRE ($MRE^{Cel\text{-}67\text{-}Control}$) on expression as measured by high-throughput sequencing. Values were calculated by dividing the frequency of reads corresponding to each variant in cDNA pool by the frequency of that variant in the pDNA pool (n=3 biological replicates, mean +/- standard deviation). **(b)** Flow cytometry validation gating strategy. Each variant from the random validation set was independently transfected into HEK-293T cells. **(c)** Linear regression comparing variant expression measured by high-throughput sequencing with expression measured by RT-qPCR in HEK-293T cells transfected with each variant from the validation set. ($P < 0.0001$, slope differs from 0). Expression was calculated with the $\Delta\Delta$Ct method using iBlue as a reference gene. **(d)** Linear regression comparing variant expression measured by high-throughput sequencing and flow cytometry (flow cytometry expression was calculated by normalising ECFP expression to iBlue expression on a single-cell basis, and taking the mean of that normalised value for each MRE variant), ($P < 0.0001$, slope differs from 0). **(e)** Linear regression comparing expression measured by RT-qPCR to flow cytometry ($P < 0.0001$, slope differs from 0).

4.5 Discussion

In this chapter, I described a high-throughput sequencing (HTS) approach that we developed to study miRNA targeting preferences at single-nucleotide resolution. We confirmed that increasing complementarity between a miRNA and its target leads to a corresponding increase in the magnitude of repression (Fig. 4.1). Focusing on single-nucleotide MRE variants, we evaluated how each position within an MRE contributes to the strength of target silencing (Fig. 4.2). These single-nucleotide data demonstrate that miR-17 and miR-21 rely on complementarity at distinct position to exert their function, but for both miRNAs, complementarity with non-seed nucleotides is important for strong repression (Fig. 4.3). We showed that translational repression and transcript degradation are correlated for miR-17 MRE variants (Fig 4.4). Finally, we used RT-qPCR and flow cytometry to validate the accuracy of this HTS approach (Fig 4.5). Together, these findings improve our understanding of miRNA targeting preferences and provide key insights that subsequently enabled us to tune transgene expression (See chapter 5). However, our findings also raise new questions about miRNA biology. In this discussion sub-chapter, I will describe the advantages and limitations of our HTS MRE analysis method and consider the key questions raised by the data obtained using this approach.

By using a degenerate oligonucleotide library, we were able to rapidly create a set of thousands of MRE variants. There are other methods of generating sequence diversity in DNA of interest such as error-prone PCR[134,135] but degenerate oligo libraries offer two key advantages [129]. First, by ordering hand-mixed, degenerate oligonucleotides it is possible to specify which positions will be variable and which

positions will stay constant. This allowed us to diversify the MRE itself but keep the flanking sequence fixed without requiring additional cloning steps. The second advantage of using a degenerate oligo is that by adjusting the percentage of degeneracy at each position in the oligonucleotide pool we were able to skew the library towards single and double-nucleotide mismatches (Fig. 4.1.c). To achieve this distribution, we calculated the binomial probability of recovering single or double nucleotide variants of a 23 nucleotide MRE. 9% degeneracy maximises the predicted representation of single and double mismatches. Degeneracy below 9% would have skewed the library towards perfectly complementary sites (no mismatches) while greater degeneracy would have favoured more than two mismatches. The actual MRE pool bears striking resemblance to our intended mismatch distribution, confirming the effectiveness of using a degenerate oligo in this setting (Compare Fig. 4.1.c with d). Degenerate oligonucleotides allowed us to specify both the position and magnitude of diversity in our MRE library.

Degenerate oligonucleotides made it possible to create thousands of MRE variants and HTS technology allowed us to assess all of those variants in parallel. Although HTS enabled us to screen more MREs than would have been feasible using low-throughput methods like RT-qPCR and flow cytometry, HTS is prone to bias[136]. Sequencing errors can arise during PCR amplification of the target locus or during the sequencing process itself. For example, it is known that the Illumina sequencing platform that we used in this study is prone to systematic errors at GGC sequences, inverted repeats and the nucleotides near the ends of the read[136]. To mitigate errors from amplification, we used Phusion, a hi-fidelity polymerase and performed the minimum number of PCR cycles required to

observe an amplicon by gel electrophoresis (See materials and methods for more details). To mitigate the impact of sequencing errors at the ends of reads, we designed PCR primers that place the MRE of interest in the middle of the sequencing amplicon, allowing us to computationally trim the ends off of each read. We also ensured that our individual sequencing libraries were of high-quality using FastQC[137]. Finally, we validated the HTS data against RT-qPCR (Fig. 4.5.d). Although our results are subject to sequencing error, validation experiments suggest that any bias or systematic error that remains following end-trimming is not sufficient to confound the utility of this HTS approach.

We chose to focus the majority of our analyses on MREs with a high degree of complementarity to a cognate miRNA. By focusing on MREs with a single mismatch to miR-17 or miR-21, we were able to determine how each base in the MRE contributes to the strength of repression (Fig. 4.2.a, 4.3.b). These single-nucleotide MRE variants displayed a good dynamic range and broad coverage in expression output, indicating that they will be useful for precisely tuning gene expression levels. Although these data provide some insight into physiological miRNA targeting preferences, they also raise new questions. For example, our results suggest that certain base changes within a MRE are more deleterious than others (Fig. 4.2.b, 4.3.c), but what biochemical properties influence the positional and nucleotide preferences of miRNA targeting? We noted that G:U wobble pairing contributes to stronger repression than non-pairing interactions (Fig. 4.2.c) but beyond this insight we have yet to elucidate the underlying factors that dictate repressive strength. Wee *et al*. demonstrated that binding affinity between a siRNA loaded into RISC and its cognate target contributes to mRNA slicing

kinetics in cell lysates[118]. I speculate that binding affinity also drives repressive strength in live cells. This hypothesis could be tested by reconstituting our MRE library *ex vivo* and quantifying the affinity of each variant for miRNA-loaded-RISC. Although our current data set does not provide an obvious explanation for why certain targets are preferred over others, it does provide a comprehensive empirical summary of how strongly MRE variants will be repressed by an endogenous miRNA.

Another question raised by the dataset presented here is whether or not our findings can provide insight into the repressive potential of endogenous MREs. In mammals, it is extremely rare for a naturally occurring MRE to exhibit perfect, or near perfect complementarity to an endogenous miRNA[138,139]. This is an important distinction as researchers have shown that high target complementarity promotes rapid miRNA mediated mRNA slicing[118]. In contrast, MREs with lower complementarity are thought to promote mRNA de-adenylation, de-capping and translational repression[138,139]. If the MREs in our variant library are repressed through a distinct mechanism from naturally occurring MREs, then the relative importance of each nucleotide that we observed here may not match their relative importance for endogenous targets. Two lines of future experiments could help resolve this apparent inconsistency. First, it is possible to knockout AGO2, the protein responsible for RISC's slicing activity[140], and repeat the high-throughput MRE variant screen. In the absence of AGO2, miRNAs could only be loaded into other members of the AGO family which do not possess slicing capacity. This would reveal MRE targeting preferences in the absence of slicing. Second, it would be interesting to design a new MRE variant library skewed towards lower

complimentarily to miR-17. This would produce MREs that are more similar to the majority of endogenous miRNA targets[138,139]. Both of these experiments could improve our understanding of endogenous miRNA targeting preferences. These limitations considered, the MRE variant library used in this study provided precise control of gene expression levels over a broad dynamic range, contributing to my primary objective of developing a tuning system.

Another interesting observation from the data presented in this chapter is the apparent difference in targeting preferences between miR-17 and miR-21 (Fig. 4.3.d). The poor correlation between single-nucleotide MRE variants at the same positions in miR-17 and miR-21 is somewhat surprising given our previous experiments with a small set of di-nucleotide mismatched MREs. In chapter 3, we saw good agreement between kinetic predictions based on let-7 and actual repression for MREs with adjacent di-nucleotide mismatches to both miR-17 and miR-21 suggesting these three miRNAs have similar targeting preferences (Fig. 3.4.d). I speculate that the adjacent di-nucleotide mismatches studied in chapter 3 could be more deleterious to structural interactions between RISC and the MRE. Perhaps interactions with broad domains within RISC are important irrespective of miRNA identity. In contrast, individual base pairs that do not impact RISC-MRE interactions on a structural level could play a more sequence-specific role. To test this prediction and to improve our understanding of miRNA targeting, it would be interesting to screen MRE variant libraries against a larger set of perhaps 10 miRNAs. This could help distinguish between general and specific properties of miRNA target selection. Although our single-nucleotide analysis demonstrates that there are indeed positional differences between miRNAs, the libraries of miR-17

and miR-21 MREs that we studied will be useful for tuning gene expression. According to the miRNA expression atlas, miR-17 and miR-21 have been detected at significant levels in 27% and 73% of surveyed cell types respectively[95]. This suggests that our current data sets will be useful for tuning genes in a broad range of contexts and applications.

By subjecting cells transfected with the MRE variant library to polysome profiling, we were able to determine that miR-17 repressed translation of MRE variants in a manner that correlates with transcript degradation (Fig. 4.4.e). At first glance, the possibility that MREs in the variant library were translationally repressed appears to be challenged by two pieces of evidence. First, researchers have shown that transcripts harbouring MREs with near-perfect complementarily to a miRNA are preferentially repressed by Ago2-mediated mRNA slicing[118]. Second, when we compared steady-state mRNA levels to protein levels for a validation set of MREs from the library, we observed a correlation with a slope nearly equal to 1 (Fig. 4.5.e). This is consistent with a scenario whereby nearly all changes in protein levels can be explained by changes in mRNA levels, a finding that diminishes the contribution of translational repression. A recent study by Tat *et a*l. helps to reconcile these seemingly contradictory findings[111]. The authors used polysome fractionation to show that miRNA-mediated transcript degradation can actually occur co-translationally[111]. If, in fact, the transcripts in our MRE variant library were being sliced or degraded during translation, fewer ribosomes could amass on those transcripts and they would appear to be translationally repressed. The strong correlation that we observed between transcript abundance and translational efficiency for MRE variants in the library is consistent with this

possibility (Fig. 4.4.e). Nonetheless, it is difficult to disentangle these two mechanisms of repression, namely inhibition of translational elongation or co-translation transcript degradation, using polysome profiling alone. Ultimately, for developing a gene tuning system we are primarily concerned with protein output levels. Despite the intricacies underlying miRNA's mechanism of action, our validation experiments show that HTS analysis of MRE strength is a very good predictor of gene expression output (Fig. 4.5.c) thus supporting the utility of this method for gene tuning purposes.

Chapter 5: Fine tuning gene expression levels with miRNA silencing-mediated fine-tuners (miSFITs)

5.1 A miSFIT dictionary relates MRE sequence and gene expression output

In the previous chapter, I described a high-throughput method for assaying MRE function. This approach allowed us to dissect microRNA targeting preferences at single-nucleotide resolution. In this chapter I will describe how this functional insight can be translated for predictably and precisely engineering expression levels of user-specified genes. We show that it is possible to fine tune a fluorescent reporter (ECFP) and a T-cell inhibitory receptor (PD-1) with important implications for anti-cancer immunotherapy[27,36,38,40], using MREs with varying complementarity to miR-17.

In chapter 4, we demonstrated that introducing a mismatch into an MRE can drastically alter the strength of miRNA-mediated repression in a manner that depends on the position and base identity of the mismatch (Fig. 4.2.a,b). By ranking all 69 single-nucleotide miR-17 MRE variants from strongest to weakest, we created a dictionary of microRNA silencing-mediated fine-tuners (miSFITs) (Fig 5.1.a). This miSFIT dictionary relates MRE genotype (sequence) to phenotype (repressive strength). Sorting all miSFITs according to their predicted strength revealed that the system has the capacity to achieve precise, stepwise control of gene expression levels. To determine just how precise this method actually is, we plotted the cumulative distribution of the expression difference between adjacent miSFITs in the dictionary (Fig. 5.1.b). This analysis revealed

that the difference in expression between adjacent single-nucleotide miSFIT variants is 1.3% of maximal expression on average (0.87% to 1.77%, 95% CI, Fig. 5.1.b). Based on these data, we conclude that miSFITs are an effective tool for precisely modulating ECFP levels in HEK-293T cells.

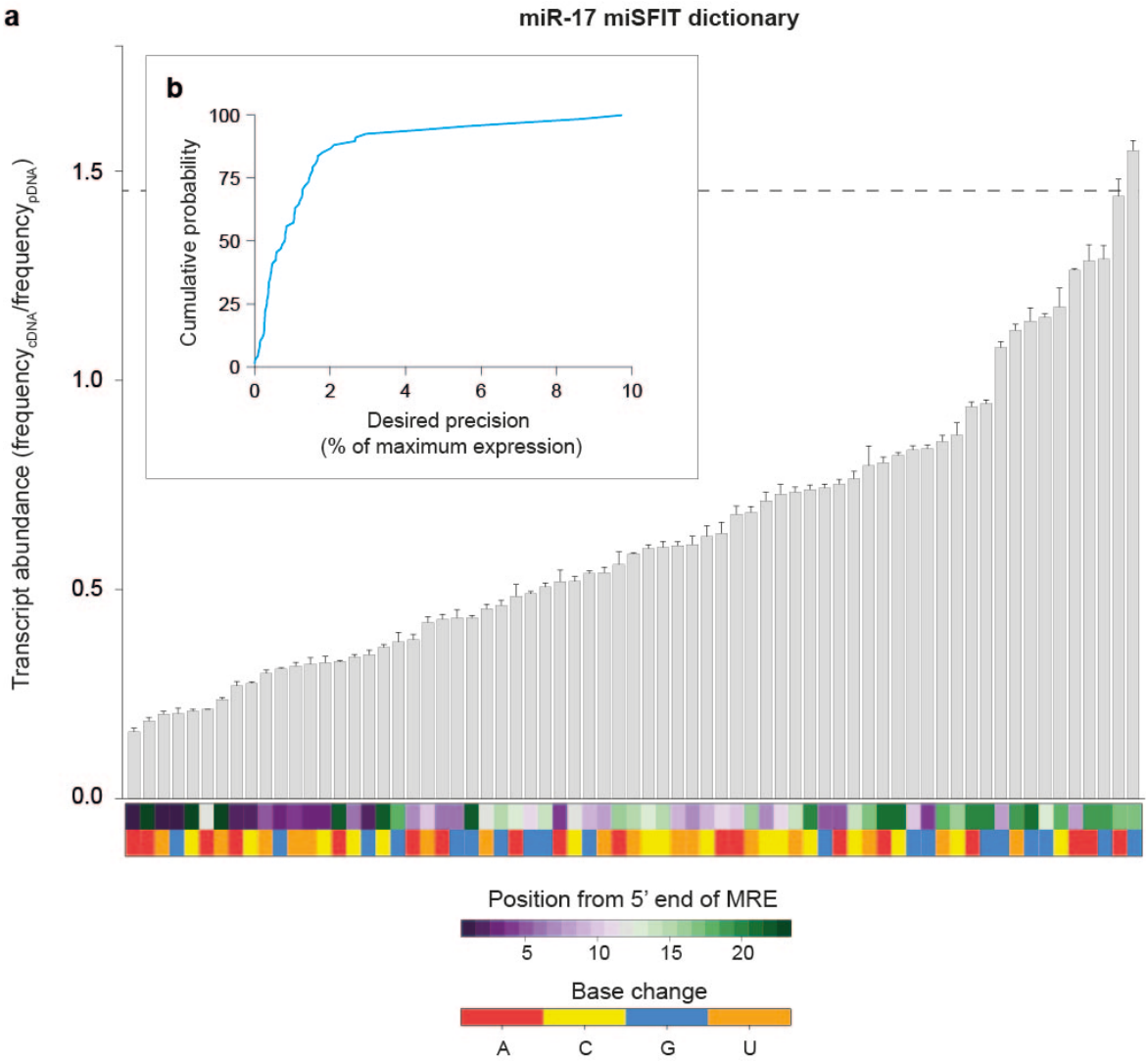

Figure 5.1. Synthetic miSFIT variants enable precise fine-tuning of reporter expression in mammalian cells.
(a) Impact on transcript abundance of all single-nucleotide miR-17 miSFIT variants ranked by expression output. Dashed line indicates expression level of a Cel-miR-67-MRE control. Coloured rectangles beneath each bar indicate the position (top) and base change (bottom) of the synthetic MRE variant (n = 3 biological replicates, mean +/- standard deviation). **(b)** Cumulative probability distribution of the precision that can be achieved using miSFIT technology. The distribution reflects the difference in ECFP expression between nearest single-nucleotide miR-17 MRE in variants in HEK-293T cells.

5.2 MRE-mediated fine-tuning of PD-1 expression in Jurkat T cells

We next asked if a selection of miSFIT variants from this dictionary, could be used in a different human cell type to tune expression of a protein with an important biological function. Programmed cell death 1 (PD-1) is a co-inhibitory receptor expressed on effector T cells and an important target for anti-cancer immunotherapy[27,36,41,44]. PD-1 is upregulated following T-cell activation and serves as a negative regulator of T-cell effector function and proliferation upon recognition of its cognate ligand PD-L1 on tumour cells or antigen presenting cells (APCs) [40]. In states of chronic stimulation, as is the case in cancer and prolonged viral infection, PD-1 contributes to T-cell differentiation into an exhausted and dysfunctional state[27,36,40]. Antibodies that block PD-1 from interacting with PD-L1 confer durable clinical benefit in patients with melanoma, lung cancer and other malignancies[27,44]. However, these blocking antibodies also lead to systemic immune-upregulation and auto-immunity[45], a shortcoming that currently limits wider usage of this class of therapies. A previous study made use of mRNA electroporation to transiently titrate PD-1 expression on T-cells from healthy donors[38]. The authors concluded that T-cell cytotoxicity is a function of PD-1 expression in a viral infection model[38]. We reasoned that tuning PD-1 expression on tumour-specific T-cells could help strike a balance between tumour killing and auto-immune reactivity.

We aimed to precisely modulate PD-1 expression using a panel of engineered MREs. we selected four miR-17 miSFITs from the ECFP dictionary, a perfectly complementary MRE (1x perfect site), tandem perfectly complementary MREs (2x perfect sites), and the control Cel-miR-67 MRE. We appended each variant

downstream of PD-1 in a bi-cistronic lentiviral vector[113] that also encodes a control reporter gene (truncated nerve growth factor receptor, NGFR) that is not under MRE control (Fig. 5.2.a). We then transduced Jurkat T-cells, an immortalized T-cell leukaemia line that expresses very low levels of PD-1 at baseline, with each of these constructs at low multiplicity of infection (MOI). We sorted pools of NGFR+ cells and expanded the sorted poly-clonal lines for two weeks. Following this expansion step, we assayed PD-1 expression by flow cytometry (Fig. 5.2.b,c). The selected miSFITs elicited discrete, stepwise control over PD-1 levels. Furthermore, PD-1 tuning was accurately predicted by the ECFP MRE dictionary (Fig. 5.2.d, R^2 = 0.94, P = 0.0015, linear regression). These results indicate that our strategy can be used to precisely tune PD-1 expression in a manner that is stable over time. They also suggest that the same panel of miSFITs can predictably tune multiple target genes in distinct human cell types, highlighting the versatility of this approach.

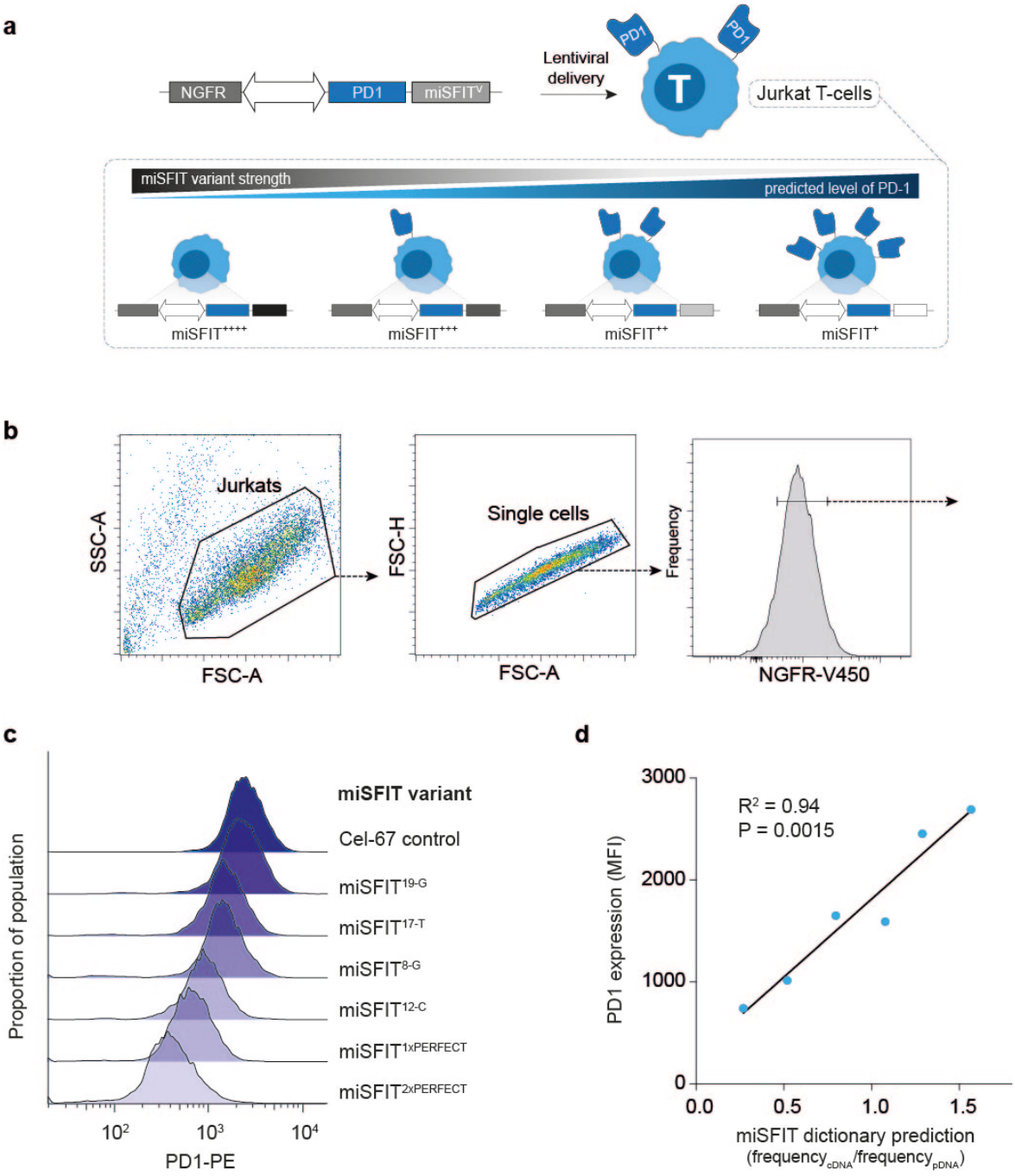

Figure 5.2. Tuning PD-1 expression in Jurkat T-cells with miSFITs.
(a) Schematic of miSFIT tuning strategy. PD-1 expression is controlled by miSFIT variants while NGFR serves as an un-silenced internal control. **(b)** Flow cytometry gating strategy used to analyse PD-1 expression on Jurkat miSFIT cell lines. **(c)** Flow-cytometry histograms of PD-1 expression on Jurkat T-cell lines transduced with one of six different miSFIT variants (x-axis = log10 transformed PD1-PE fluorescence). **(d)** Correlation between predicted expression in the miSFIT dictionary and observed PD1 expression on Jurkat T-cell lines (linear regression).

5.3 Genomic integration of miSFITs in embryonic stem cells with CRISPR/Cas9

We successfully tuned transgenic PD-1 expression in Jurkat T-cells using a panel of lentiviral vectors (Fig. 5.2.c). However, in a clinical setting, it would be necessary to tune endogenous PD-1 to improve primary T-cell effector function. To address this requirement, we set out to demonstrate that miSFITs can be introduced into the endogenous PD-1 locus by CRISPR/Cas9 mediated genome engineering. As a proof-of-principle demonstration, we attempted to introduce a panel of miR-21 miSFITs into the PD-1 3'UTR in mouse embryonic stem cells (mESCs). We chose miR-21 because it is highly expressed on activated T-cells[141]. Furthermore, we previously designed a panel of MREs with various mismatches to miR-21 that conferred a range of gene-expression output levels (Fig 3.2.b). We chose mESCs because they are amenable to single-cell cloning and expansion *in vitro*. Clonal mESC lines can be used to create transgenic mice, and mESCs can be differentiated into CD8+ T-cells *in vitro*[142,143]. mESCs are also reportedly amenable to highly efficient genome editing[68].

The WIMM transgenics facility derived a mESC line from an OT-I mouse. The OT-I strain carries a transgenic T-cell receptor that is specific for SIINFEKL, a model antigen from the protein Ovalbumin[144]. Next, we devised a genome engineering strategy for inserting a panel of miR-21 miSFITs into the endogenous PD-1 locus. We decided to co-deliver an expression vector[121] that drives a selection marker (Puro), Cas9 and one of 5 sgRNAs designed to target the PD-1 3'UTR along with a single stranded oligonucleotide homology donor (ssODN) (Fig 5.3.a). We designed two ssODNs at positions in the PD-1 3'UTR adjacent to the predicted cut

sites of the sgRNAs. Each ssODN comprises the MRE insert and a BamHI restriction site to enable rapid genotyping. The MRE and restriction site are flanked by 45nt homology arms on either side. After nucleofecting mESCs with each sgRNA (which we labelled sgA,B,C,D,E) along with their corresponding ssODN, we applied a 48 hour puromycin selection (Fig 5.3.b). To genotype the resulting bulk cell populations, we PCR amplified the targeted locus and digested with BamHI to assess the efficiency of HDR mediated MRE insertion (Fig 5.3.c). This analysis revealed that one sgRNA, sgD, catalysed efficient HDR when co-delivered with an ssODN (Fig 5.3.c).

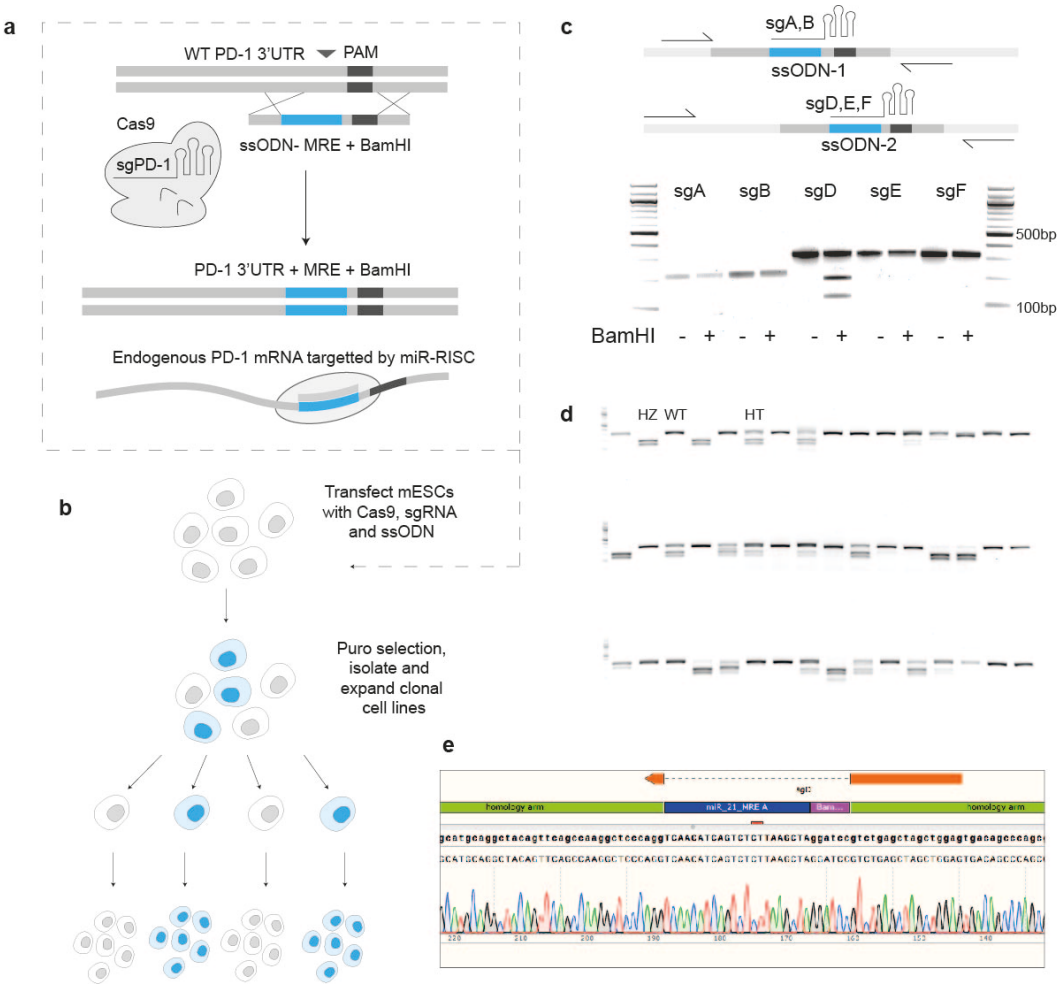

Figure 5.3. Introducing miSFITs into the genome with CRISPR/Cas9.
(a) Schematic representation of the genome editing strategy used for targeted miSFIT integration into the PD-1 locus. MRE is shaded in blue. **(b)** Schematic of mESC cloning strategy. Successfully edited cells are shaded in blue. **(c)** Restriction digest-based genotyping analysis of bulk mESC populations following transfection with the indicated sgRNA and puromycin selection. sgA and B were co-transfected with ssODN-1 and sgD,E and F were co-transfected with ssODN-2. **(d)** Restriction digest-analysis of clonal cell lines selected and expanded from mESCs electroporated with sgD and ssODN-2. Examples of the digestion pattern associated with wildtype (WT), bi-allelic (HZ) and mono-allelic (HT) integration are indicated. **(e)** Sanger sequencing confirms correct miSFIT insertion in a homozygous clone identified in (d).

Next, we used sgD in combination with distinct ssODNs to introduce each of 6 different MREs (4 miR-21 miSFITs, a MRE with perfect complementarity to miR-21 and a Cel-miR-67 control MRE) into the PD-1 3'UTR. Following transfection and puromycin selection, we selected 48 individual clones for each MRE in our

panel. We expanded each clone and isolated cells for genotyping. We checked each clone for the MRE insert using the same BamHI genotyping method that we applied to the bulk population (Fig 5.3.c). This editing strategy resulted in very efficient HDR. For the representative ssODN shown in figure 5.3.d, 44% (21/48) of genotyped clones contained an MRE knockin. The BamHI digestion strategy also allowed us to distinguish between mono-allelic and bi-allelic(homozygous) knockins. At least 6 clones appear to have undergone homozygous MRE insertion (Fig 5.3.d). Finally, to validate the aforementioned BamHI genotyping strategy, we further expanded clones that appeared to contain bi-allelic insertion and performed Sanger sequencing on the targeted locus (Fig 5.3.e). This approach confirmed correct homozygous miSFIT integration in the PD-1 3'UTR for each MRE (Fig 5.3.e). These data suggest that MREs can readily be inserted into genomic loci in mESCs using the CRISPR/Cas9 system. In the future, the miSFIT knockin lines that we have generated can be injected into blastocysts to create mosaic miSFIT knockin mice whose T-cells will express tuned levels of PD-1. The mESCs can also be differentiated into CD8+ T-cells *in vitro*[142]. Thus, we expect miSFIT mESC lines will serve as a renewable resource for generating T-cells with tuned PD-1 levels.

5.4 Implementing miSFITs in primary mouse T-cells with CRISPR/Cas9 is inefficient

Fine-tuning PD-1 expression levels holds the potential to improve anti-cancer immunotherapy. CRISPR/Cas9 mediated genome engineered enabled efficient insertion of a panel of miSFITs in the endogenous PD-1 3'UTR in mESCs (Fig 5.3.c,d,e). Although stem cells can provide a renewable supply of T-cells[142,143], fine tuning PD-1 expression directly in primary T-cells directly could provide a

more rapid therapeutic alternative. Minimizing the time between diagnosis and treatment is paramount for effective cancer therapy, especially for melanoma patients[145,146]. Furthermore, engineering PD-1 expression in tumour infiltrating T-cells or tumour-reactive peripheral blood T-cells could provide an additional layer of cancer specificity. With these benefits in mind, CRISPR/Cas9 has previously been employed to knockout PD-1 in primary human T-cells[47,48]. As discussed in section 5.2, tuning PD-1 has the potential to promote more effective anti-tumour therapy than complete PD-1 abrogation.

In an effort to strike a balance between anti-tumour reactivity, autoimmunity and T-cell dysfunction, we set out to introduce miSFITs into the endogenous PD-1 locus in primary mouse T-cells. We decided to use sgD and two corresponding ssODNs since this combination proved highly efficient in mESCs (Fig. 5.3.c). As with previous experiments, we included a BamHI recognition site in the ssODN insert to simplify downstream genotyping. Primary murine T-cells are difficult to transfect with plasmid DNA and RNPs have been shown to catalyse efficient and facile HDR in human T-cells[47]. Therefore, rather than delivering Cas9 and the sgRNA on an expression plasmid, we opted to use a ribonucleoprotein (RNP) composed of recombinant Cas9 protein and *in vitro* transcribed sgRNA (Fig. 5.4.a).

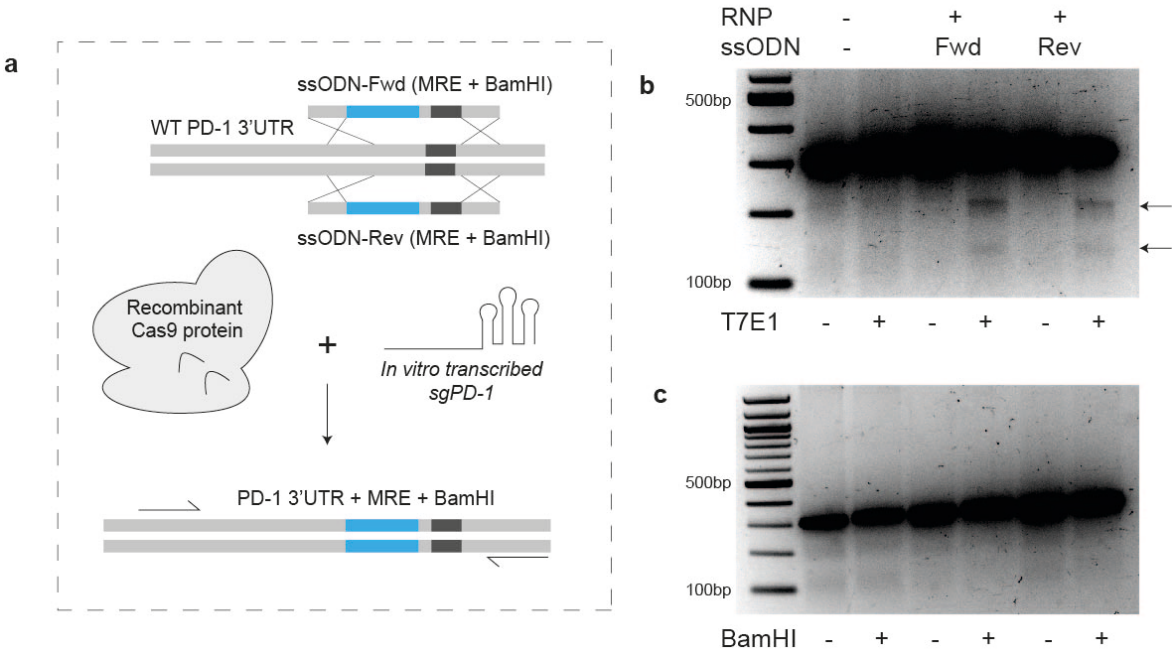

Figure 5.4. CRIPSR/Cas9 mediated miSFIT integration is inefficient in primary mouse T-cells.
(a) Schematic representation of the Cas9-RNP strategy used to introduce miSFITs into the PD-1 3'UTR in primary mouse T-cells. **(b)** T7E1 genotyping of total edits in bulk T-cell cell populations electroporated with recombinant Cas9, sgRNA targeting the PD-1 3'UTR and the indicated ssODN. Arrows indicate the expected digestion products. **(c)** Restriction digest-based genotyping reveals a lack of HDR in the same populations shown in (b).

We isolated splenocytes from OT-I mice and stimulated them with SIINFEKL peptide to promote expansion. Next, we purified CD8+ T-cells by negative selection and nucleofected them with RNPs targeting the PD-1 3'UTR. After two days, we harvested genomic DNA from the nucleofected cells and PCR amplified the locus of interest (Fig. 5.4). To assess total CRISPR/Cas9 mediated mutagenesis, we re-annealed and digested PCR amplicons with T7E1, an endonuclease which cleaves mismatched double stranded DNA[147] (Fig. 5.4.b). We observed editing events in cells nucleofected with Cas9, sgD and two ssODNs, albeit at very low frequencies (Fig. 5.4.b). Next, we digested the same PCR amplicons with BamHI to check for HDR mediated MRE insertion (Fig. 5.4.c).

Unfortunately, we could not detect any HDR events by gel electrophoresis. Although further optimization may improve the efficiency of HDR in primary mouse T-cells, this data suggests that mESCs are much more amenable to knockins.

5.5 Discussion

In summary, in this chapter I showed that single-nucleotide miR-17 MRE variants, which we call miSFITs, enable precise, stepwise control of fluorescent reporter gene expression (Fig. 5.1). Furthermore, we demonstrated that miSFITs are portable to another gene and to a different cell type by tuning PD-1 transgene expression in Jurkat T-cells (Fig. 5.2). Next, we successfully integrated miR-21 miSFITs into the endogenous PD-1 3'UTR in mESCs with high efficiency using CRISPR/Cas9 genome engineering (Fig. 5.3). Finally, we attempted to introduce the same miSFITs into the genome of primary mouse T-cells. Unfortunately, we were unable to achieve detectable miSFIT insertion in this difficult-to-edit cell type (Fig. 5.4).

This discussion section addresses some noteworthy aspects of these results. I comment on how the precision that can be achieved with miSFITs compares to that of other gene expression control methods. This section next explores the value of fine-tuning PD-1 and explains why implementing miSFITs into the endogenous PD-1 locus in mESCs will provide a useful research tool. Finally, I address our attempt to edit primary mouse T-cells and discuss how to improve on these results in the future.

Our data demonstrate that miSFIT variants provide fine control over gene expression levels. In fact, the average difference in gene expression level between adjacent miR-17 miSFIT variants is 1.3% (Fig. 5.1.b). This represents a marked improvement in tuning resolution when compared to other methods. For example, the resolution of expression control that can be achieved using heterozygous/homozygous knockout strategies is on the order of 50% in the best case. siRNA and shRNA yield unpredictable knockdown and are often completely ineffective[148,149]. Tetracycline inducible promotor systems have been optimised to provide greater than 1000-fold dynamic range in expression control and may rival the precision achieved with miSFITs[150]. However, tetracycline is an immunomodulatory small molecule known to impact cellular metabolism[81]. Titrating the concentration of this drug across experimental conditions creates a confounding variable, especially for assays of immune-function. miSFITs, on the other hand, provide excellent precision and I speculate that they impart minimal impact on cellular function.

In this chapter, we sought to illustrate the utility of miSFIT technology by tuning PD-1 expression. We achieved precise, predictable control over PD-1 levels in Jurkat T-cells (Fig. 5.2.c). There are several reasons to believe that tuning PD-1 levels is a superior therapeutic strategy to knocking out PD-1. PD-1 deletion causes severe lupus-like autoimmunity in mice[41] and PD-1 sequence variation is associated with multiple autoimmune disorders in humans[40,42]. In the absence of this negative regulator, T-cells undergo strong and rapid activation in response to TCR ligation. Interestingly, other co-inhibitory receptors are activated to compensate for PD-1 loss[49]. After their initial burst in activity, PD-1 null T-cells

succumb to terminal dysfunction as a result of this compensatory inhibition[49]. By transiently transfecting primary T-cells with varying doses of PD-1 mRNA, Wei and colleagues showed that PD-1 levels inversely correlate with T-cell effector function[38]. Stably and precisely suppressing PD-1 expression on patient T-cells could provide a means of balancing anti-tumour reactivity with autoimmune harm[45].

We decided to introduce miSFITs into the endogenous PD-1 3'UTR in mESCs as a proof-of-concept demonstration that engineered MREs can be integrated at specific genomic loci. There are two reasons why we chose mESCs for these experiments. First, this cell type is highly amenable to genome editing[68]. We were able to achieve miSFIT knockin efficiencies exceeding 40% (Fig. 5.3.d). Second, mESCs constitute a potential source of T-cells. In the future, we can either utilize our miSFIT mESC lines to generate transgenic mice or we can subject these mESCs to *in vitro* differentiation. The Zuniga-Pflucker lab has shown that mESCs differentiate into CD8+ T-cells upon co-culture with stromal cells expressing Delta-ligands[142]. This protocol could offer a renewable supply of PD-1 miSFIT T-cells for functional characterisation experiments.

After successfully introducing miSFITs into the genome in mESCs by CRISPR/Cas9 mediated HDR, we attempted to make the same insertions in primary mouse T-cells. The most practical way to tune PD-1 expression levels for immunotherapy would be to insert miSFITs into the genome of patient-derived peripheral blood T-cells or tumour infiltrating T-cells[47]. Unfortunately, in our experiments we were unable to detect HDR mediated miSFIT integration in mouse

T-cells. Although this result was disappointing, it was not wholly surprising. Based on our lab's first-hand experience, as well as reports in the literature, it is difficult to efficiently deliver proteins and nucleic acid to murine T-cells[151]. In mESCs, we were able to overcome any limitations in delivery efficiency by selecting transfected cells with puromycin. In contrast, it is challenging to maintain actively dividing mouse T-cells in culture long enough to apply a selection. Although we could not detect HDR in primary mouse T-cells (Fig. 5.4.c), we did observe some NHEJ events (Fig. 5.4.c). This confirms that we were able to deliver functional Cas9 to these cells. NHEJ is an efficient repair mechanism that competes with HDR in most cell types[67,75]. In addition to improving delivery efficiency, precise miSFIT knockin will also require methods that direct cellular DNA repair towards HDR. Encouragingly, several groups are working to improve HDR efficiency by better design of editing reagents[152,153] or inhibition of competing NHEJ[154]. Furthermore, while reports of efficient HDR in murine T-cells remain scant, multiple groups have reported reasonably high rates of HDR in primary human T-cells[155,156]. Despite these discouraging preliminary data, recent methodological advances in genome engineering and T-cell transfection suggest that it will be possible to integrate miSFITs into the genome in primary T-cells in the near future.

Chapter 6: Fine tuning the adaptive immune response with miSFITs

In the preceding chapters, I described the development and implementation of a gene expression tuning system based on synthetic MREs with varying complementarity to an endogenous miRNA. These engineered MREs, which we call miSFITs elicited predictable control over ECFP and PD-1. In this chapter, I demonstrate that fine-tuning gene expression levels with miSFITs can provide new insight into important biological processes. Specifically, miSFITs allowed us to determine that the strength of the anti-tumour immune response is a function of tumour-associated antigen expression level.

6.1 Fine-tuning tumour antigen expression and T-cell activation in mouse melanoma

To illustrate the utility of miSFITs as an effective tool for modulating gene-expression, we sought to apply this technology towards a biological question that has previously been confounded by technical limitations. More specifically, we set out to explore how peptide-antigen expression levels influence the strength of the anti-tumour immune response in a murine melanoma model. Cancer immunotherapy is a promising class of treatments that aim to enhance anti-tumour cytotoxicity by the adaptive immune system[27,32,157]. Several sub-types of immunotherapy, including checkpoint blockade[36] and adoptive cell transplant[50], rely on T-cell receptor (TCR) mediated recognition of peptide antigens presented by MHC-I molecules on the surface of tumour cells[51]. Although *in silico* algorithms can accurately predict which peptide antigens are likely to elicit an immune response [158-162], understanding how peptide-antigen expression levels influence

the strength of the anti-tumour immune response *in vivo* remains elusive. A quantitative analysis of this relationship could provide an important benchmark for predicting which tumours might respond to anti-cancer immunotherapy.

Previous efforts to titrate peptide-MHC concentrations have relied on coating culture vessels with recombinant peptide-MHC multimers[163] or by briefly adding varying concentrations of peptide to cellular growth media (a process known as peptide pulsing)[164]. Although valuable, these methods cannot accurately recapitulate the endogenous pathway of antigen expression, proteolytic processing and subsequent surface presentation. Furthermore, because peptide pulsing is inherently transient, this method precludes tracking the survival of antigen-expressing cells in vivo. To understand how antigen-expression influences the anti-tumour immune response and the relative fitness of cancer cells *in vitro* and *in vivo*, we used miSFITs to finely tune expression of ovalbumin (OVA), a model immunogenic protein, in a stable and physiologically accurate fashion.

To this end, we created a panel of seven bi-cistronic OVA expression vectors, each encoding a distinct miSFIT variant in the 3'UTR of ovalbumin (Fig. 6.1.a). We also coupled EGFP downstream of ovalbumin via a self-cleaving T2A peptide, enabling us to monitor expression levels by flow-cytometry (Fig. 6.1.a). In each vector, NGFR was included as an unsilenced internal control reporter[113]. We transiently expressed these constructs in B16-F10 melanoma cells to evaluate gene expression output (Fig. 6.1.a). This analysis revealed discrete, stepwise tuning of target levels, although the exact ranking of miSFIT variant strength differed from what we observed when tuning PD-1 in human Jurkat T-cells (Fig

6.1.b). To generate stable cell lines expressing varying levels of ovalbumin, we then transduced B16-F10 cells with a subset of five OVA-miSFIT constructs at low MOI (<2% transduction efficiency) (Fig 6.1.c). The semi-random nature of lentiviral integration results in heterogeneity of gene expression between individual cells[165]. To mitigate this effect, we sorted and expanded pools of 150,000 cells on the basis of NGFR expression. We analysed the resulting five cell lines by flow cytometry (Fig. 6.1.d) and confirmed that we successfully tuned ovalbumin expression (Fig. 6.1.e).

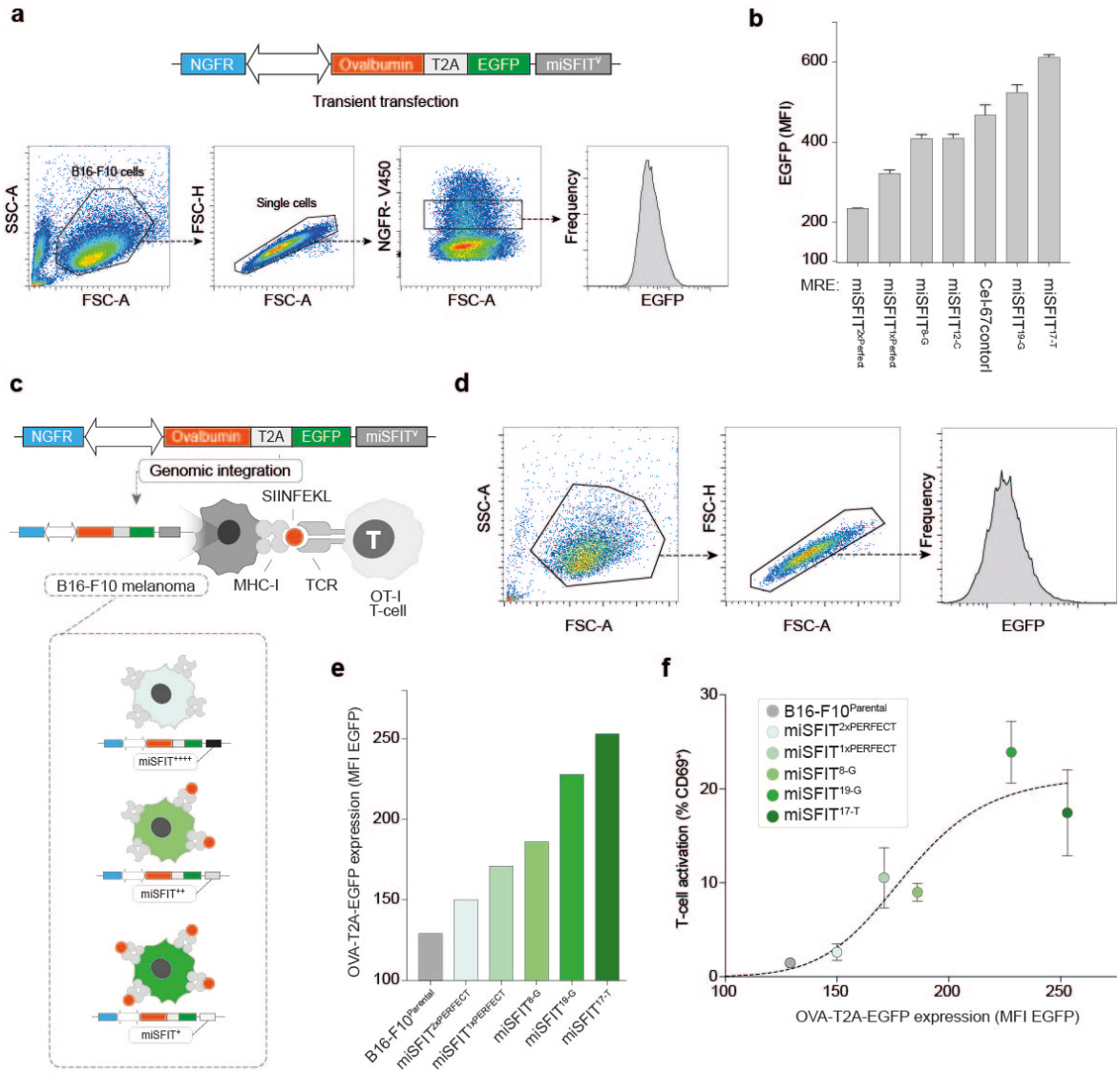

Figure 6.1. Tuning Ovalbumin expression in B16-F10 melanoma.
(a) Flow-cytometry gating strategy used to assess OVA-T2A-EGFP expression following transient transfection of seven different miSFIT variant constructs in B16-F10 cells. **(b)** EGFP expression in B16-F10 cells following transient transfection of seven OVA-miSFIT constructs (n = 3 biological replicates, mean +/- standard deviation). **(c)** Pipeline for stably tuning ovalbumin expression with lentiviral miSFIT integration. (d) Flow-cytometry gating strategy for analysing OVA-T2A-EGFP expression on stable miSFIT cell lines. **(e)** OVA-T2A-EGFP expression on five stable miSFIT cell lines. **(f)** CD8+ OT-I T-cell activation by OVA-miSFIT B16-F10 cell lines. CD69 expression was quantified by flow cytometry (n = 5 biological replicates, mean +/- standard deviation).

Next, we asked how antigen expression levels influence CD8+ T-cell activation. The OT-I T-cell receptor (OT-I) is specific for SIINFEKL, a short peptide antigen derived from ovalbumin, presented by MHC-I[166]. We co-cultured each of the five

B16-F10 lines expressing differential ovalbumin levels and the OVA-negative parent line with CD8+ OT-I T-cells and assayed activation by measuring CD69 expression (Fig. 6.1.f). Indeed, increasing OVA expression resulted in a concomitant increase in the proportion of activated T-cells, presumably due to the greater probability of each T-cell encountering and responding to a SIINFEKL-MHC-I complex (Fig. 6.1.f).

6.2 Immuno-editing and tumour fitness are functions of antigen expression level in vitro

Under selective pressure by the adaptive immune system, tumours have been shown to acquire mutations that prevent effective T-cell surveillance in a process known as immunoediting[33]. In mouse models, this is generally achieved through loss of function mutations in MHC genes, up-regulation of immunosuppressive molecules or by elimination of clones expressing neo-antigens[33,167]. Evidence for immune-mediated clonal selection is now also beginning to emerge in patient studies[168]. Colorectal tumours exhibit lower-than-predicted neoantigen burdens as well as prevalent loss of function mutations in HLA genes that are indicative of immune-selection[168]. It is clear that T-cells can specifically lyse antigen-expressing tumour cells but the variables that determine the selective pressure against antigen-expressing cells have yet to be elucidated.

We hypothesized that tumour cells might also be selected on the basis of antigen expression levels. To address this possibility, we devised an *in vitro* fitness assay whereby two cell populations are grown together in the presence or absence of a T-cell challenge. By quantifying the frequency of each population in the T-cell

condition and comparing that to the corresponding frequency in the no-T-cell condition after one cell-doubling time, it is possible to estimate the relative fitness of each population. To this end, we first mixed the five OVA-miSFIT B16-F10 cell lines at a 1:1 ratio with OVA-negative B16-F10 cells (Fig. 6.2.a). We then allowed these mixed cultures to grow overnight in the presence or absence of OT-I T-cells at two different ratios (1:1 and 3:1 T-cells:B16-F10 cells).

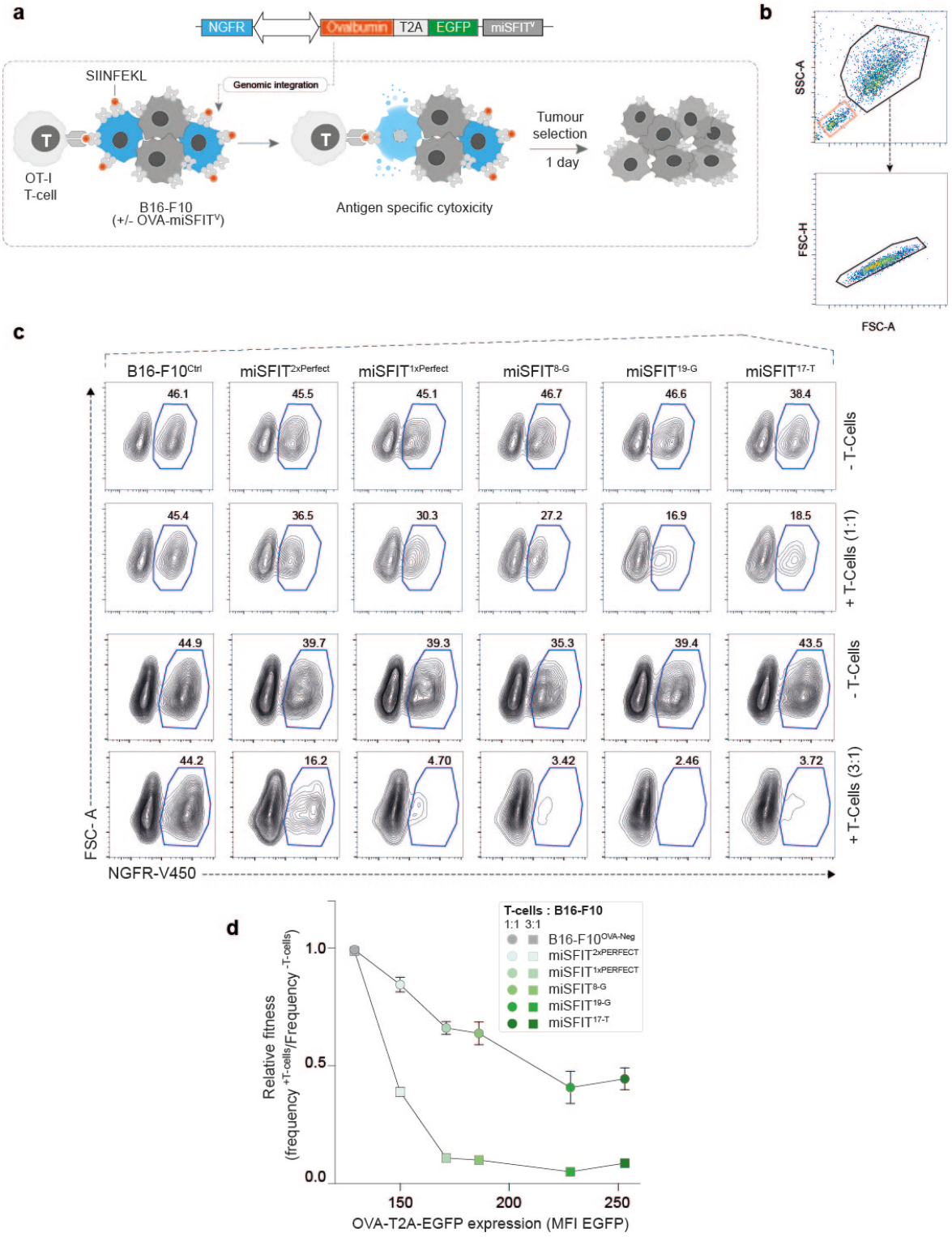

Figure 6.2. Melanoma cellular fitness is a function of tumour-associated antigen expression *in vitro*.
(a) Schematic representation of mixed-culture experimental design. OVA-negative (NGFR-) are mixed with OVA-miSFIT (NGFR+) B16-F10 cells and are challenged overnight with OT-I T-cells. **(b)** Flow-cytometry gating strategy for evaluating B16-F10 fitness. **(c)** Representative flow cytometry plots of mixed culture experiments. The percentage of NGFR+ (OVA-miSFIT) cells (blue polygon gate) surviving after overnight selection in the presence or absence of CD8+ OT-I T-cells (at the

specified ratio of T-cells:tumour cells) is indicated for each condition. **(d)** Relative fitness of B16-F10 cell lines as a function of OVA expression. Relative fitness was calculated by dividing the frequency of NGFR+ cells with T-cells by the frequency of NGFR+ cells without T-cells (n = 3 biological replicates, mean +/- standard deviation).

After approximately 20-hours in co-culture, we analysed the remaining B16-F10 cells by flow-cytometry (Fig. 6.2.b,c). Because all OVA-miSFIT lines express NGFR whilst the OVA-negative parent line does not, we quantified the relative abundance of OVA+ (NGFR+) and OVA-negative (NGFR-) cells following the T-cell challenge (Fig. 6.2.b,c). Tuning antigen expression using miSFITs modulated the strength of T-cell mediated selection in a dose-responsive manner at both T-cell: tumour cell ratios (Fig. 6.2.c,d). Notably, even low antigen expression was sufficient to elicit a strong reduction in relative fitness at a high T-cell: tumour cell ratio (Fig. 6.2.c,d). Our findings also indicate that it is possible to achieve efficient killing of tumours expressing low and high antigen levels. However, a greater number of T-cells are required to achieve equivalent control of low-antigen expressing melanoma cells. In summary, we demonstrate that T-cell mediated tumour selection is a function of antigen expression level and T-cell:tumour cell ratio *in vitro*.

6.3 Antigen expression levels determine tumour growth and survival in vivo in an adoptive cell transfer model

Next, we asked if the effect of antigen expression on melanoma survival that we observed *in vitro* correlates with tumour growth rates *in vivo*. B16-F10 is a melanoma line derived from C57/BL6 mice line that readily engrafts into recipient mice of the same strain upon intradermal injection[169]. Once tumours have established, it is possible to model the adaptive anti-tumour immune response by

adoptively transferring T-cells into the recipient mouse. Adoptively transferred T-cells can be delineated from host cells by injecting T-cells of a different allotype from the host. For example, CD45.1+ T-cells can be distinguished from CD45.2+ host T-cells by flow cytometry. We decided to use this model system to study the impact of tumour associated antigen expression level on the magnitude of the adaptive immune response *in vivo*.

We injected 6 mice each with a subset of three engineered OVA-miSFIT-B16-F10 cell lines and an OVA-negative control line into syngeneic recipient mice (Fig. 6.3.a). After allowing intradermal tumours to establish for seven days, we adoptively transferred CD-8+ OT-I T-cells into each recipient mouse and monitored tumour growth for an additional 22 days (Fig. 6.3.a). Antigen expression levels significantly impacted tumour growth *in vivo* in a manner that faithfully mirrored *in vitro* T-cell activation and killing (Fig. 6.3.b) ($P = 0.02$, Kruskal-Wallis test, comparison of tumour volumes at day 19). We continued to monitor mice for 46 days and observed that antigen expression markedly influenced survival ($P = 0.0038$, Logrank test for trend, Fig. 6.3.c). Mice bearing tumours with no, or low antigen expression all met our endpoint criteria by day 27. In contrast, medium or high OVA expressing tumours displayed a substantial increase in survival. One third of the mice bearing high-antigen B16-F10-OVA cells (2/6) survived for 46 days with tumours that were nearly undetectable by the experiment's endpoint (Fig. 6.3.c). Together, these findings demonstrate that miSFITs function *in vivo* and illustrate the importance of tumour-associated antigen expression levels in determining the strength of the immune-response.

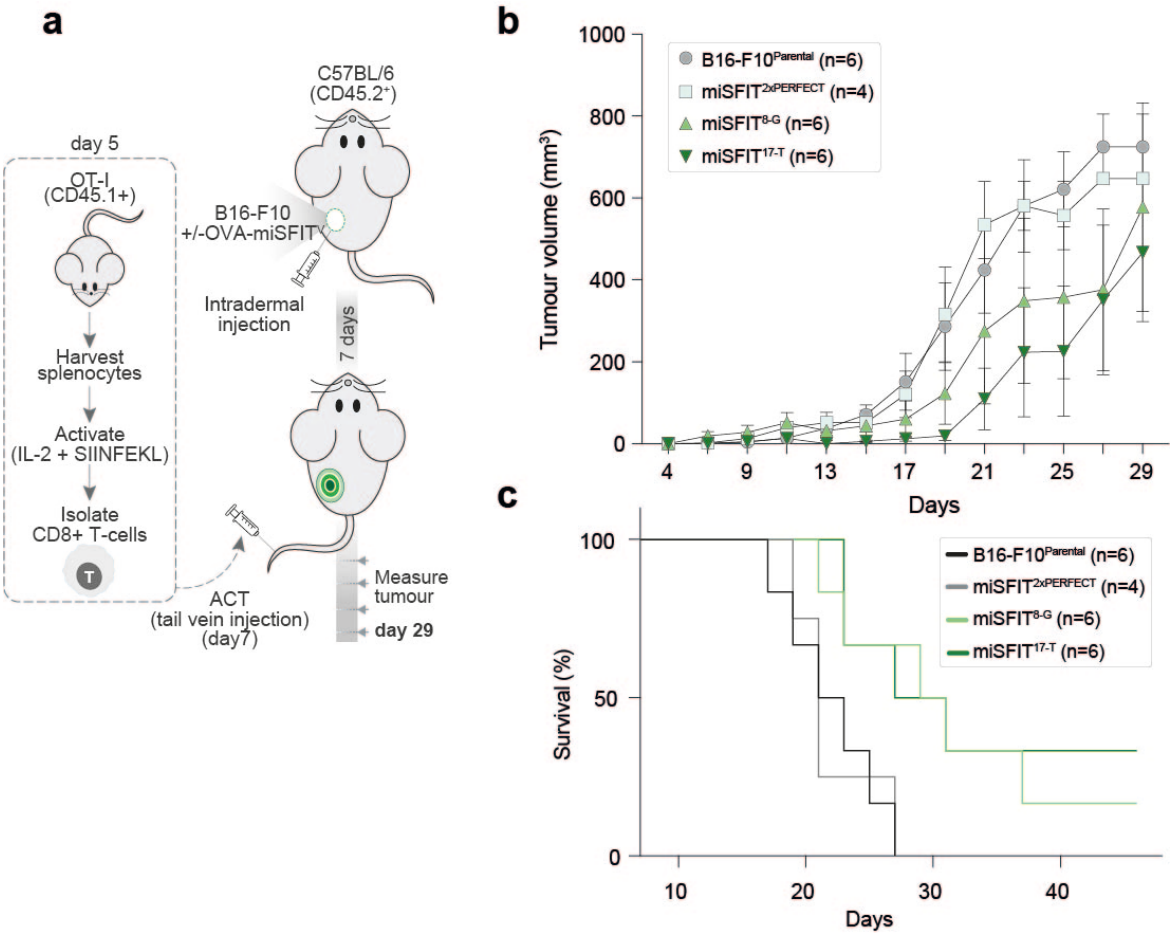

Figure 6.3. Tumour growth and mouse survival are functions of tumour-associated antigen expression.
(a) Experimental design for *in vivo* OVA-miSFIT B16-F10 tumour growth experiments. **(b)** Analysis of tumour volume over time for four B16-F10 cell lines (3 OVA-miSFIT variant lines and one B16-F10 parental control line) challenged with OT-I CD8+ T-cells (x-axis = number of days from tumour cells injection; mean +/- s.e.m.) **(c)** Survival curves for mice injected with B16-F10 lines following the same experimental setup as in (a).

6.4 The effect of antigen expression on tumour growth is mediated by differential T-cell infiltration

To understand why higher antigen expressing tumours were more effectively controlled, we harvested and analysed tumour infiltrating lymphocytes (TILs) at eight days after adoptive T-cell injections. We analysed a panel of T-cell surface markers by flow cytometry to compare the phenotype of TILs between tumours (Fig 6.4.a). Since the OT-I T-cells that we adoptively transferred have a CD45.1+

allotype, while the recipient mice are CD45.2+, it is possible to distinguish each T-cell's origin by flow cytometry.

To control for differences in T-cell phenotype that are not associated with tumour recognition, we first compared surface marker expression on T-cells harvested from recipient mouse spleens (Fig. 6.4.b). Because we stimulated OT-I T-cells prior to intravenously injecting them into recipient mice we expected these cells to exhibit a memory phenotype. Consistent with this hypothesis, nearly all transferred T-cells expressed the memory marker CD44 (Fig. 6.4.b). Expectedly, splenic T-cells did not express the early activation marker CD69, nor the intermediate activation marker CD25 (Fig. 6.4.b). A small proportion of spleen-derived T-cells from each experimental condition expressed PD-1 which marks activated and exhausted T-cells (Fig. 6.4.b).

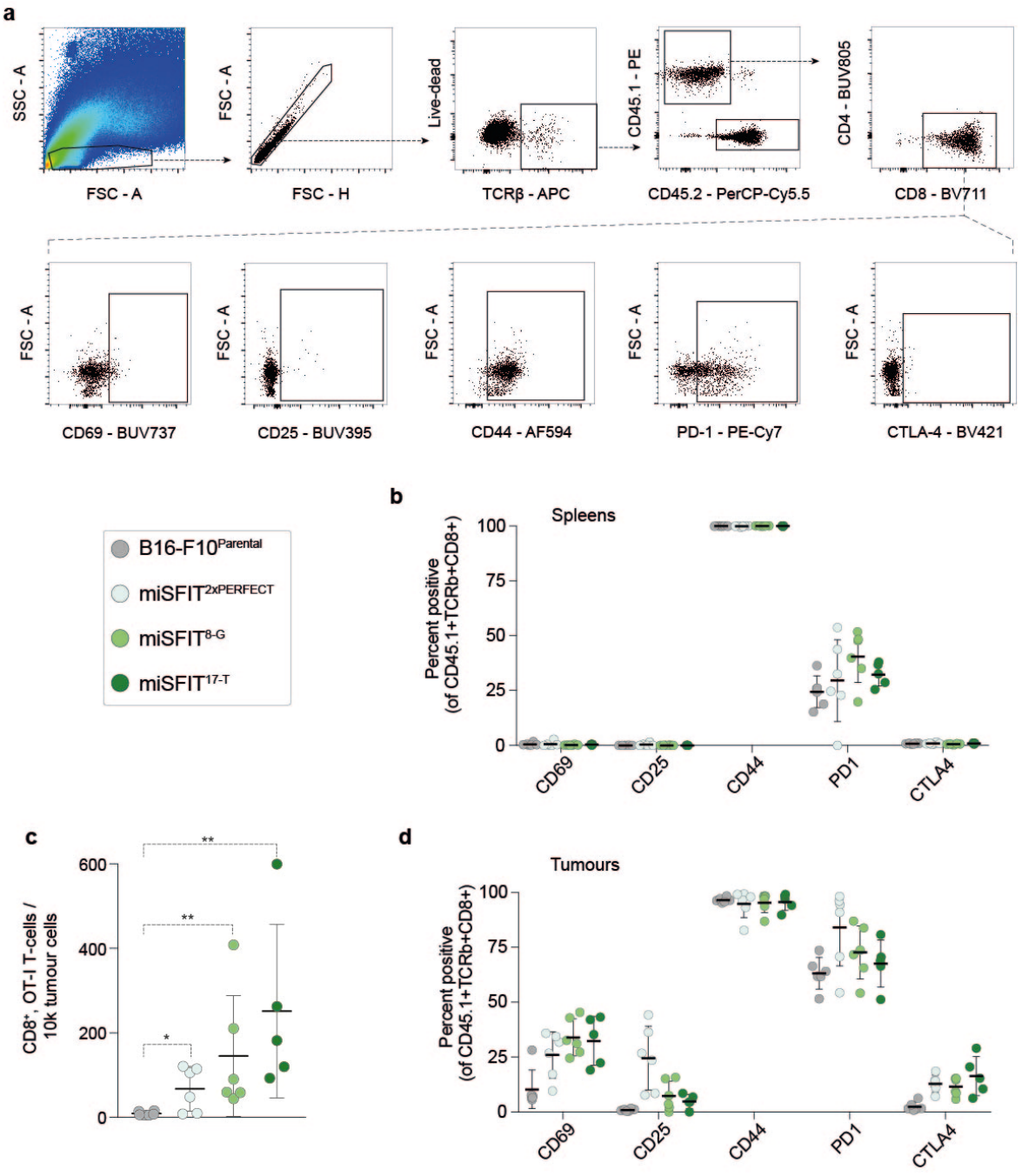

Figure 6.4. T-cell infiltration is correlated with tumour antigen expression level and tumour growth.
(a) Flow cytometry gating strategy used to quantify and analyse TILs from mice bearing each of three OVA-miSFIT variant tumour lines or a B16-F10 control line. **(b)** Immunophenotype of CD8+, CD45.1+, TCRß+ splenocytes from mice bearing each tumour type (mean +/- standard deviation). **(c)** Frequency of CD8+ OT-I T-cells per 10,000 tumour cells for each tumour type (mean +/- standard deviation, Mann-Whitney U test, *P < 0.05, **P < 0.01). **(d)** Immunophenotype of CD8+, CD45.1+, TCRß+ TILs from mice bearing each tumour type (mean +/- standard deviation).

Next, we compared CD45.1+ TILs between recipient mice bearing each OVA-miSFIT-B16-F10 tumour variant. Strikingly, increasing levels of OVA lead to a dose-responsive increase in the frequency of TILs *in vivo* (P = 0.003, Kruskal-Wallis test, (Fig. 6.4.c). This could be a result of improved T-cell infiltration, increased T-cell proliferation or a combination of both factors. We proceeded to examine changes in surface marker expression of TILs from each tumour type (Fig. 6.4.d). The early activation marker CD69 and the exhaustion marker CTLA-4 appeared to increase with increasing OVA expression (Fig. 6.4.d). On the other hand, CD25 and PD-1 expression did not respond monotonically to increasing antigen levels (Fig. 6.4.d). Although antigen expression may differentially influence T-cell phenotype, the most definitive consequence of higher antigen expression levels is an increase in the number of TILs.

6.5 Discussion

Taken together, the data presented in this chapter demonstrate that miSFIT technology is an effective method of tuning the expression of a physiologically relevant protein. We stably tuned Ovalbumin expression in a mouse melanoma cell line and showed that antigen expression levels influence T-cell activation and tumour cell fitness *in vitro*. We also identified tumour associated antigen level as a determinant of T-cell infiltration and the anti-tumour immune response *in vivo*. In this sub-chapter, I discuss limitations of our results as well as new lines of investigation raised by our findings. I also describe how our efforts to better understand the relationship between tumour-associated antigen expression levels and the anti-tumour immune response can be applied clinically.

In this study, we designed a lentiviral expression vector that produces Ovalbumin coupled to EGFP by a self-cleaving peptide. The expression level of Ovalbumin and EGFP are both regulated by a downstream miSFIT variant (Fig. 6.1.e). This proved to be an effective method of studying the relationship between antigen expression and T cell response. However, there are two limitations of this approach that could be improved on in future iterations. First, we used EGFP as a proxy for Ovalbumin, rather than directly quantifying Ovalbumin itself. Researchers have previously raised an antibody against SIINFEKL (the immunogenic antigen from Ovalbumin that is recognized by the OT-I TCR) bound to the MHC[170]. Unfortunately, this antibody displays extremely poor sensitivity and is incapable of reliably detecting SIINFEKL on our highest EGFP-expressing B16-F10 cell line. Despite our inability to directly detect the antigen, EGFP serves as a reliable proxy since both proteins are produced from the same polypeptide before they are separated by the self-cleaving T2A peptide linker.

It should also be noted that the ovalbumin-derived model antigen SIINFEKL is recognized by the OT-I TCR with very high affinity[171]. However, patient-derived tumour-associated antigens have varying affinity and avidity for their cognate TCRs[172]. Applying the miSFIT technology to *bona fide* tumour antigens, such as NY-ESO-1, will enable scientists to understand how antigen expression levels and immunogenicity interact to influence the immune response.

In spite of the limitations associated with Ovalbumin, this model antigen allowed us to address fundamental questions about immunotherapy. In our experiments,

we observed an inverse correlation between antigen expression level and tumour fitness in response to T-cell challenge (Fig. 6.2.d). It is tempting to speculate that these *in vitro* results are indicative of how a tumour that displays heterogeneity in antigen expression will evolve under immune selection *in vivo*. Based on our data, we would expect a loss in high-antigen expressing tumour sub-clones over time. However, another possibility is that a small number of high-antigen expressing cells in a tumour are able to recruit and activate sufficient quantities of T-cells to kill low-antigen expressing cells as well. If high-antigen expressing cells indeed drive collateral killing in a heterogeneous tumour, this will alter the relative fitness cost associated with high-antigen expression. We could use our *in vitro* mixed culture system to assess this possibility by mixing each of our B16-F10 OVA-miSFIT lines with low-Ova cells instead of Ova-negative cells. By quantifying the absolute survival of the low-Ova cells in the mixed culture during a T-cell challenge, we could determine whether high-antigen cells contribute to collateral killing.

One matter of contention in the field is which classes of tumour associated antigens can be targeted effectively in a therapeutic setting. Tumour associated antigens fall into three categories- *i.)* new antigens that are produced by tumour mutations that do not exist elsewhere in the transcriptome (these are known as neoantigens), *ii.)* antigens which are otherwise only expressed in the testes, an immune-privileged tissue, but are upregulated in tumours (known as cancer-germline antigens) and *iii.)* antigens which are expressed in healthy tissues but are significantly upregulated in tumours (known as over-expressed antigens) [51]. It is clear that neoantigens represent the safest class of targets for immunotherapy

since they are not expressed on healthy tissues[173]. Due to a lack of data, it has remained unclear whether T-cells specific for overexpressed antigens can target a tumour while sparing healthy tissues expressing low levels of the same antigen. Here, we provide compelling evidence that it is difficult for T-cells to selectively target cancer cells expressing high antigen levels without killing low antigen expressing cells (Fig. 6.2.d). At the dose of antigen-specific T-cells required to elicit >90% killing of Ovalbumin-high cells, we observed >50% killing of Ovalbumin-low cells (Fig. 6.2.d). This finding suggests that it will be challenging to specifically target overexpressed antigens with current adoptive T-cell strategies as considerable damage to non-tumour tissue may occur.

In addition to suggesting that overexpressed antigens are difficult to target safely, our mixed co-culture and *in vivo* experiments may also help us predict how patients with cancer-germline antigens or neoantigens are likely to respond to checkpoint blockade. Better methods of predicting patient responses are critical for sparing those who are unlikely to respond to a given treatment from incurring unnecessary side effects. Our data reveals that T-cell mediated tumour killing is a function of antigen expression level *in vitro* (Fig. 6.2) and *in vivo* (Fig. 6.3). It would be interesting to determine how checkpoint blockade would influence this response function. Clinical results suggest that higher neoantigen burden is associated with better response to PD-1 blockade[174] but it is unclear how the expression level of those neoantigens alters the probability of response. In the future, we could address this question by repeating the mixed co-culture and *in vivo* engraftment experiments with and without a PD-1 blocking antibody. I speculate that PD-1 blockade might lower the activation threshold, allowing T-cells

to respond at lower antigens levels. Ultimately, defining the range of antigen expression where PD-1 inhibition confers a benefit could help clinicians determine which patients are likely to respond to this treatment modality. Some tumour antigens, such as NY-ESO-1, are shared amongst a substantial portion of melanoma patients. In the future, we could use miSFITs to study how NY-ESO-1 levels influence the efficacy of PD-1 blockade. This would allow us to identify the window of NY-ESO-1 expression where PD-1 blockade is actually beneficial, allowing patients who are unlikely to respond to immunotherapy to avoid unnecessary side effects.

Chapter 7: General discussion and future directions

The over-arching goal of the research project described herein was to develop a technology for precisely tuning gene expression levels in mammalian cells. In this thesis, I document my efforts to understand how base pairing between a miRNA and MRE influences the strength of miRNA-mediated gene silencing. We used this information to design and implement miSFITs that confer predictable control over transgene expression in multiple mammalian cell types, thereby achieving the main aim of the project. Finally, to demonstrate the value of this new technology, we used miSFITs to dissect the relationship between tumour associated antigen expression and T-cell response. In this chapter I will discuss some limitations of this new technology that still need to be addressed. This chapter will also highlight key advantages of miSFITs compared to existing gene-modulation strategies. Finally, I will close by proposing how miSFIT technology will help address outstanding research challenges and contribute to improving cell-based therapy in the future.

7.1 Current limitations of miSFIT technology

In this thesis, I attempted to systematically assess the generalisability of miSFIT technology but some important considerations still need to be addressed for miSFITs to be widely adopted by the scientific community. Although we have tested MRE libraries for two different miRNAs and applied one of these libraries to tune three different genes in three different cell types, it would be useful to further characterise the portability of this new tuning method.

One of my design criteria was to create a tuning platform that is portable to different cell types and tissues. In mammals, heterogeneity in transcription factor expression and epigenetic landscapes across cell types means that transcriptional regulation is a tissue specific phenomenon[7,93]. Consequently, I reasoned that gene tuners based on post-transcriptional miRNA targeting would perform more consistently than engineered promotor/enhancer elements. A comparison of our miSFIT data to an effort by Schlabach *et al*. to engineer synthetic mammalian promotors supports the hypothesis that miSFIT-based tuning is the more portable approach[175]. Schlabach *et al.* set out to identify novel strong mammalian enhancers from random sequence libraries. Although they successfully recovered strong enhancer elements that functioned well in HEK 293T cells, when they transferred these elements to HeLa, a second human cell line, they failed to observe a significant correlation in activity (R^2 = 0.56, P = 0.15)[175]. In contrast, when we transferred a panel of miSFITs from HEK 293T cells to Jurkat T-cells, we observed a strong, significant correlation in expression output between these two cell types (R^2 = 0.94, P = 0.0015, Fig. 5.2.d). Although this finding indicates that miSFITs are more portable than enhancers or promotors, more work needs to be done to draw a rigorous conclusion. Testing the same panel of miR-17 miSFIT variants in an additional 5 to 10 commonly used cell types would help address this current limitation.

In addition to comprehensively evaluating how predictably miSFITs behave across cell types, understanding how this technology will perform in other organisms will also require further experimentation. When we tested a panel of miR-17 miSFIT variants in B16-F10 mouse melanoma cells, the order of repression did not match

the order that we observed in human cells (compare Fig. 6.1.b to Fig. 5.2.d). This difference may be attributable to several factors. First, miR-17 is a member of a conserved family of miRNAs that share the same seed sequence[133]. Perhaps other miR-17 family members are expressed in B16-F10 cells that display distinct targeting preferences from miR-17. However, this explanation is unlikely as miR-17 family members differ from miR-17 in positions that were held constant in the miSFIT panel that we tested[133]. To determine whether the observed difference in miSFIT strength between HEK 293T and B16-F10 cells is anomalous, or indicative of a species-portability barrier, it would be useful to test a larger panel of miSFITs in multiple murine cell lines. Fortunately, in spite of the unexpected order of miSFIT strength in B16-F10 cells, the panel we selected still afforded precise, stepwise control of Ovalbumin expression levels in this cell line (Fig. 6.1.e).

Another limitation of the data presented in this thesis is that we have, thus far, only dissected the targeting preferences of two miRNAs, miR-17 and miR-21. Interestingly, the MRE positions that are important for miR-17 regulation do not correlate well with the positions that are important for miR-21 (Fig. 4.3.d). This lack of correlation in miRNA targeting preferences suggests that if users would like to construct miSFIT libraries that are triggered by miRNAs other than miR-17 and miR-21 they will need to dissect the targeting landscapes of each miRNA on an individual basis. There are three ways that this limitation can be mitigated or overcome. Firstly, the high-throughput assay that we have developed and optimised for dissecting miRNA targeting preferences can be readily applied to other miRNAs. Following the protocols described in chapter 2.7 of this thesis, it is possible to build and test a new MRE variant library in under two weeks.

Secondly, if a user is unable to generate a new miSFIT library, it is probable that they will be able to apply either the existing miR-17 or miR-21 library for their desired application. In particular, miR-21 is expressed at detectable levels in over 70% of human and mouse cell types[95]. Finally, there may be cases where a user seeks to tune a gene in a cell type that does not express miR-17 or miR-21 and that user is unable to profile the targeting preferences of another miRNA. In this situation, it may be effective to generate a small, random set of MRE variants for a miRNA that is expressed in the relevant cell type. We actually tested this solution inadvertently when we selected a set of 15 miR-17 MREs at random to validate the high-throughput sequencing data (Fig. 4.5.). Intriguingly, this random set of MRE variants conferred excellent stepwise control over ECFP expression in HEK 293T cells (Fig. 4.5.a). I suggest that intentionally synthesising a small panel of random or pseudorandom MREs is a quick and easy solution for tuning gene expression using miRNAs with unknown targeting preferences.

In summary, the data presented in this thesis indicate that miSFITs are a generalizable tool for tuning gene expression but further work is required to determine how broadly miSFITs can be applied to different cell types, organisms and miRNAs.

7.2 Advantages of miRNA-based gene tuning

miSFITs exhibit key advantages over other gene tuning strategies. In this subchapter, I compare miSFITs to doxycycline-inducible promotors[79,150], uORFs and TIS libraries[89]. I discuss the fact that miSFITs do not require a small molecule trigger. Next, I highlight the benefits of a genetically encoded tuning system.

Finally, I explain how miSFITs can be co-opted for conditional and tissue-specific gene tuning, a feat that is not possible with other methods.

The tetracycline/doxycycline responsive transcriptional regulator (rtTA) enables fine control of transgene expression over a broad dynamic range[79,150]. Unfortunately, tetracycline, and its derivatives, significantly impact mammalian cellular function. Ahler *et al.* demonstrated that doxycycline alters cellular metabolism and impedes proliferation in commonly used human cell lines[81], highlighting the confounding effect of this antibiotic *in vitro*. *In vivo*, doxycycline delays the onset of transplant rejection[176], validating this drug's immunosuppressive activity. Doxycycline also impedes tumour growth *in vitro* and *in vivo*[177]. Collectively, these results demonstrate that tetracycline/doxycycline responsive transcriptional regulators require carefully designed controls and are especially incompatible with studying the anti-tumour immune response. In contrast, miSFITs do not require small molecule/antibiotic triggers which is a significant advantage for studying mammalian biology.

In addition to confounding cellular function, tuning gene expression with the rtTA system suffers from additional limitations. When doxycycline is delivered *in vivo*, not all cells respond, resulting in a bimodal distribution in transgene expression[178]. Increasing the dose of this drug might increase the proportion of cells expressing the transgene, rather than the level of transgene expressed by the responding cells[178]. In contrast, miSFITs cause a unimodal shift in expression (Fig. 5.2.c). Finally, because miSFITs are short sequences that can be placed at flexible positions within a 3'UTR they are suitable for tuning expression of endogenous

genes. I speculate that replacing an endogenous gene promoter with the Tet-operator would require substantially more optimisation that introducing miSFITs into flexible locations in the genome.

Another advantageous characteristic of MRE-based gene tuning is the fact that expression levels are genetically encoded by each miSFIT. This obviates the need to constantly supply an external trigger, resulting in tuning that is stable over time in the absence of user intervention. This allowed us to tune antigen expression *in vivo*, producing a phenotype that was stable for several weeks (Fig. 6.3.a). Furthermore, because miSFITs are genetically encoded it is theoretically possible to tune multiple genes to independent levels. Titrating doxycycline would have a similar impact on two genes under rtTA control. Conversely, two genes under control of different miSFIT variants could each be tuned to distinct levels. This property will enable complex studies of multi-gene interactions *in vitro* and *in vivo*. Despite the aforementioned limitations, the rtTA system is clearly an invaluable tool, especially when inducible expression is advantageous. I believe that miSFITs will serve to complement this technology.

Other methods for encoded gene tuning in mammalian cells include altering the sequence of the translation initiation site (TIS), and appending artificial upstream open reading frames (uORFs) to the gene of interest[89]. Although TIS or uORF libraries alone are insufficient to tune genes over a broad dynamic range, Ferreira *et al*. elegantly combined these approaches to robustly tune gene expression in mammalian cell lines[89]. This combined TIS/uORF approach achieves comparable precision to miSFITs[89], but the latter method has one key advantage. Because

miRNA expression patterns vary with cell type, miSFITs can enable tissue-specific gene tuning. Gene circuits that are triggered by differentially expressed miRNA display remarkable tissue-specificity[78,112]. I envision co-opting this strategy for gene tuning by designing miSFITs with complementarity to tissue- or state-specific miRNAs[95]. Because TIS/uORF tuning relies on modulating ribosome binding and elongation, properties that are ubiquitous across cell types[89], they are not suitable for context-specific gene tuning. In the next section, I discuss exciting future applications of contextual gene tuning that are enabled by miSFITs.

7.3 Applications and future directions

miSFIT-based gene tuning has several advantages that collectively enable a variety of new applications in research and therapy. Here I will propose three such applications. I explain how miSFITs can improve CAR-T therapy, I describe how misFITs can augment immunotherapy for solid tumours, and finally I explain how tuning Cas9 expression could contribute to our understanding of mammalian cell biology.

7.3.1. Tuning CAR expression levels

Fine tuning Chimeric Antigen Receptor (CAR) expression with miSFITs could substantially improve adoptive T-cell therapy for B-cell leukaemia and lymphoma. CARs are fusion proteins comprising a synthetic sensing motif (often derived from an antibody), that is coupled to one or more intracellular activating domains derived from the human TCR[157,179,180] (Fig. 7.1.a). When the sensing motif engages its cognate ligand on the surface of a target cell, an activating signal is transduced via the CAR's intracellular domain that ultimately leads to T-cell

stimulation and target killing[157,179,180]. Transducing patient-derived T-cells with a CAR that senses a tumour associated antigen, such as CD19[181] or HER-2[182], can redirect those T-cells to seek and destroy a tumour (Fig. 7.1.a). CD19 CAR T-cell therapy has shown substantial clinical promise and recently gained FDA approval for some types of B-cell leukaemia and lymphoma[157]. Despite producing encouraging clinical results, CAR T-cells are limited by their propensity to initiate cytokine release syndrome[183] and by their susceptibility to co-inhibitory signalling[156]. Rapid lysis of CD19+ B-cells following CAR T-cell infusion results in potentially life-threatening cytokine overproduction in a large proportion of treated patients[183].

Because CARs are expressed from strong viral promotors, they reach concentrations on the cell surface that far-exceed endogenous TCR expression levels[156]. Eyquem *et al.* recently demonstrated that high CAR expression leads to excessive T-cell stimulation that ultimately results in co-inhibitory receptor mediated T-cell dysfunction[156]. They overcame this limitation by inserting the CAR directly into the endogenous TCR locus with CRISPR/Cas9 rather than using a randomly integrating viral vector[156]. This resulted in significantly lower CAR expression, reduced T-cell dysfunction/exhaustion and better tumour killing[156]. In addition to driving T-cell exhaustion, I speculate that high CAR expression also contributes to acute T-cell hyperactivation and excessive cytokine production in patients. Careful optimisation of CAR expression levels has the potential to mitigate these limitations by identifying the minimum expression level necessary to kill tumour cells without inducing hyperactivation (Fig. 7.1.b).

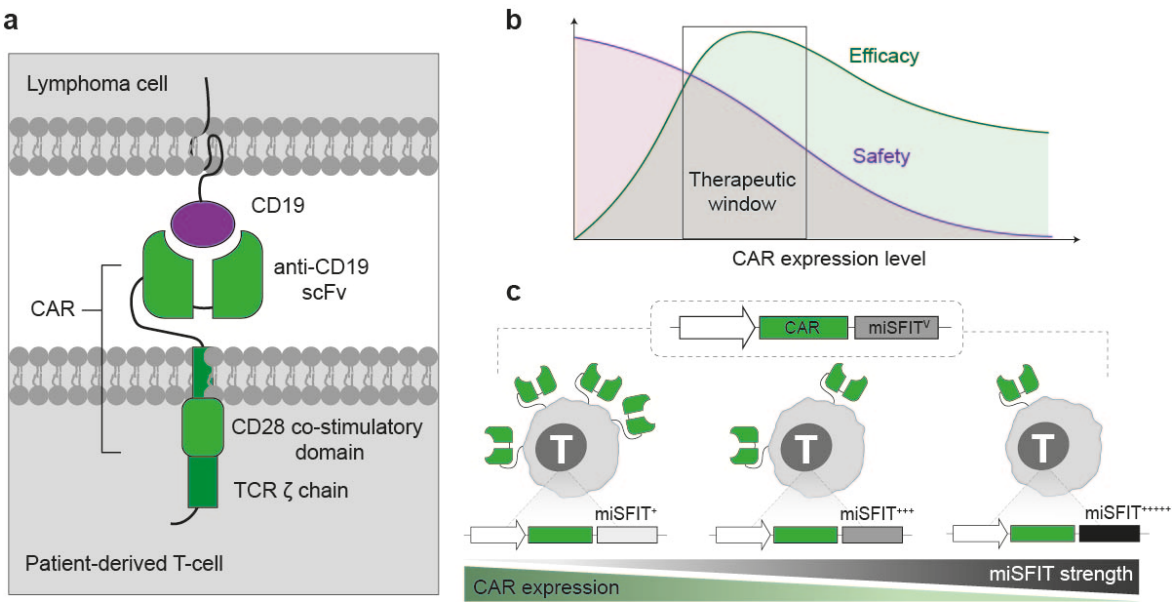

Figure 7.1. Fine-tuning CAR expression and T-cell function.
(a) Schematic representation of a CAR T-cell recognizing CD19 expressed on a lymphoma cell. **(b)** Predicted relationship between CAR expression level and CAR T efficacy and safety. **(c)** Proposed approach for tuning CAR expression with miSFITs.

Improving expression levels by targeting CARs to the endogenous TCR locus with CRISPR/Cas9 has long-term therapeutic potential[156], but the use of genome engineering in primary T-cells must first traverse a lengthy regulatory review process. Furthermore, although expression from the endogenous TCR locus improved CAR T-cell efficacy in a mouse model, the authors did not investigate how this approach might impact autoimmune side effects such as cytokine release syndrome[183]. In this thesis, I have demonstrated that miSFITs impose fine control over transgene expression levels in human cells. Integrating miSFITs into the 3'UTR of existing CAR vectors will allow precise dampening of their expression levels, helping clinicians strike a balance between tumour recognition, T-cell exhaustion and auto-immunity. Fine-tuning CAR-T therapy with miSFITs is a promising application of this new tuning technology with direct clinical relevance (Fig. 7.1.c).

7.3.2. Tuning endogenous co-inhibitory receptor expression

In chapter 5 of this thesis, I explained why it would be useful to tune PD-1 expression levels with miSFITs and demonstrated that it is feasible to introduce miSFITs into the endogenous PD-1 3'UTR using CRISPR/Cas9 genome engineering (Fig. 5.3). Here, I elaborate on how tuning co-inhibitory receptors will augment anti-tumour immunotherapy and address the practicalities of implementing this technology in the clinic.

The motivation for tuning PD-1 expression levels, rather than completely abrogating PD-1 signalling is introduced in subchapter 1.2 and detailed in sub-chapters 5.2 and 5.5. Briefly, tumour-specific T-cells often upregulate PD-1 which, in turn, suppresses T-cell effector function[36]. Antibodies that block PD-1 signalling, or genetic PD-1 ablation, promote effector activity and tumour killing but also increase off-tumour autoimmunity in animals and patients[27,36,44]. Furthermore, complete loss of PD-1 leads to compensatory T-cell inhibition through other co-inhibitory receptors such TIM-3 and CTLA-4. Fine tuning PD-1 expression with miSFITs could optimise the trade-off between tumour killing, auto-immunity and compensatory exhaustion[49].

Three classes of anti-cancer adoptive T-cell therapy are either approved for patient use or are currently undergoing clinical trial. These three classes, namely ex vivo expanded TILs[50], T-cells engineered to express a defined TCR (Such as the NY-ESO-1 specific 1G4 TCR)[184] and CAR-T cells[157], can all be impeded by T-cell exhaustion. I propose that fine tuning endogenous PD-1, as well as CTLA-4,

TIM-3 and other co-inhibitory receptors, could improve the safety and efficacy of all classes of adoptive cell therapy.

Because miSFITs respond to miRNA triggers that can vary in expression patterns across cell states[78,112], miSFITs could be adapted to confer context-specific tuning of co-inhibitory receptors. Exemplifying state-specific miRNA expression, miR-21 is strongly upregulated on activated T-cells compared to their antigen-naïve counterparts[141]. Introducing miR-21 miSFIT variants downstream of co-inhibitory genes, such as PD-1, would restrict tuning to activated T-cells (Fig. 7.2). Tumour infiltration induces significant alterations to gene expression patterns in effector T-cells[185] and it is likely that miRNA expression profiles also change upon tumour engagement. By identifying miRNAs that are specifically upregulated in TILs, we could foreseeable restrict PD-1 tuning to tumour-reactive T-cells, adding another layer of protection against off-tumour auto-immunity. The same principle can be applied to achieve context-restricted tuning of a range of effector genes for a myriad of therapeutic indications such as immunodeficiency and autoimmunity. Developing the capacity to efficiently introduce miSFITs at defined locations in the genome will enable improvements to the efficacy and safety of cellular therapy.

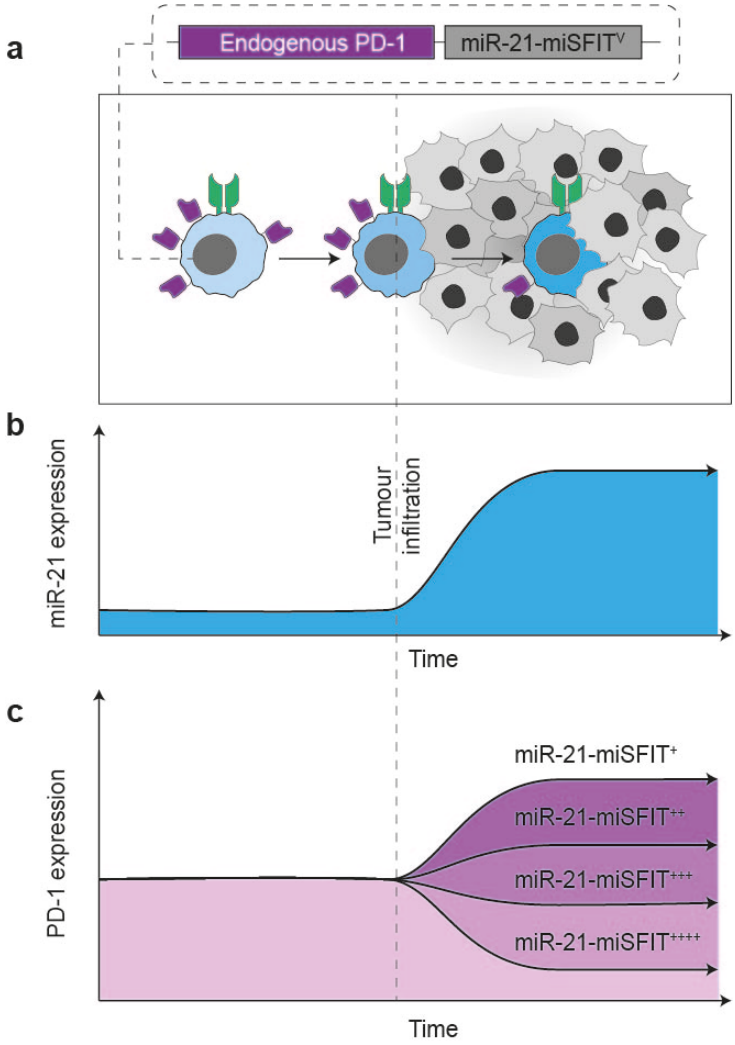

Figure 7.2. Conditional gene tuning with miSFITs.
(a) Proposed method of tuning endogenous PD-1 with miSFITs. A miSFIT variant is introduced into the endogenous PD-1 3'UTR in a patient-derived T-cell. Upon tumour infiltration, **(b)** miR-21 is upregulated in the T-cell and **(c)** miSFIT variants can to tune PD-1.

In the work presented in this thesis, targeting genomic integration of miSFITs in primary mouse T-cells was hampered by poor genome engineering efficiency (Fig. 5.4.c). Fortunately, several recent publications demonstrate that high efficiency HDR is possible in primary human T-cells. In a 2017 study, Bak *et al.* achieved greater than 10% HDR-mediated knockin of a fluorescent reporter gene in human T-cells by combining Cas9 RNP electroporation with an AAV6 HDR donor[155]. That same year, Eyquem *et al.* reported an HDR efficiency of 45% in primary human T-

cells, in this case by delivering an AAV6 donor shortly after Cas9 mRNA and sgRNA electroporation[156]. As genome engineering technologies continue to improve and eventually gain regulatory approval, miSFIT knockin in primary human T-cells could become a reality.

7.3.3. Multiplexed cellular engineering in PSCs

An alternative strategy to overcome low editing efficiency in primary T-cells is to introduce miSFITs in pluripotent stem cells (PSCs). Because of their capacity for self-renewal, it is possible to select for and expand correctly modified PSC clones. I used this approach to create homozygous miSFIT knockin mESC lines (Fig. 5.3.d,e). I envision that it will be possible to perform complex modifications in PSCs that are not feasible in primary T-cells, a cell type with much more limited self-renewal potential (Fig. 7.3). For example, in PSCs it should be possible to multiplex the insertion of CARs or engineered TCRs with the integration of miSFITs downstream of genes like PD-1 (Fig. 7.3). It will perhaps even be possible to further optimise the efficacy and safety of adoptive T-cell therapy by sequentially introducing miSFITs downstream of multiple co-inhibitory genes such PD-1, CTLA-4 and TIM-3. Finally, genes that create graft-host incompatibility such as the endogenous TCR and HLA genes can be knocked out using CRISPR/Cas9, creating an “off-the-shelf” product that bypasses HLA compatibility barriers[186,187] (Fig. 7.3).

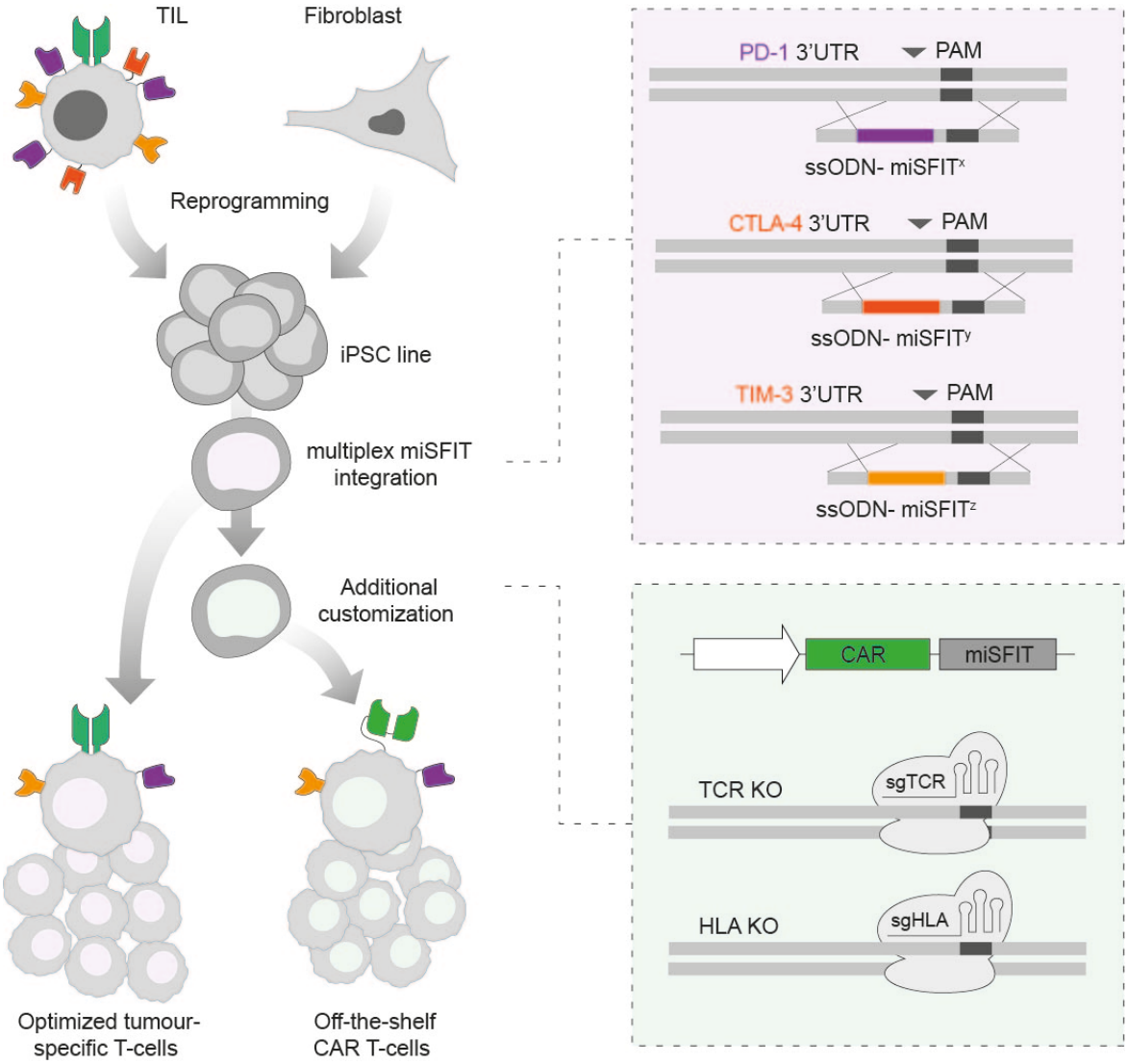

Figure 7.3. Optimising T-cell immunotherapy with multiplex genetic manipulations in PSCs.
A proposed strategy for optimised cell therapy begins with either tumour-biopsy-derived T-cells or patient fibroblasts. These cells are reprogrammed to pluripotency prior to parallel or stepwise genetic modification. The resulting modified clones can be screened prior to *in vitro* differentiation into a therapeutic T-cell product.

For this ambitious vision to come to fruition, it is necessary to improve on current methods of differentiating T-cells, or T-cell progenitors (Pro-T) from PSCs. The Zuniga-Pflucker lab pioneered a method for producing CD8+ T-cells by co-culturing PSCs with a mouse stromal cell line modified to express high levels of Notch ligands[142]. Though useful, this method is limited by high variability and a

reliance on mouse feeder cells that are not compatible with clinical applications. This method was recently improved upon by Shukla *et al.* who successfully replaced the stromal feeder cell line with an *in vitro* niche comprising recombinant DL4 and VCAM-1[188]. This fully defined, feeder-free protocol produced high yields of Pro-T-cells from a starting population of hematopoietic stem and progenitor cells (HSPCs) [188]. If this method can be adapted to work on PSCs rather than HSPCs it would serve as an invaluable resource for producing therapeutic T-cells with complex, multi-loci modifications.

7.3.4. miSFITs meet Cas9 transcriptional regulators

In addition to augmenting anti-tumour immunotherapy, miSFITs can also serve as a research tool, helping scientists discover how quantitative changes in gene expression levels impact mammalian biology. Exemplifying this possibility, I used miSFITs to tune expression of a transgenic tumour-associated antigen (Fig. 6.1). In the future, this same method can also be applied to study the impact of changes in endogenous transcript levels. To make endogenous gene tuning simple to implement for researchers across disciplines, I propose combining miSFITs with dCas9 transcriptional regulators (Cas9-TR).

By supplying dCas9 fused to an activator (CRISPRa)[82] or repressor (CRISPRi)[54,55] domain along with an sgRNA targeting a transcription start site of interest, it is possible to boost or silence transcription in a programmable fashion. Empowered by the simplicity of Cas9-TR targeting, scientists have modulated multiple genes simultaneously[55] and have even carried out genome wide activation and repression screens[83]. Cas9-TRs have generally been deployed for binary

transcription control, comparing cellular phenotypes in the presence vs. absence of CRISPRa/i. By titrating the expression level of the Cas9-TR itself, it may be possible to tune the strength of transcriptional control exerted on the sgRNA-targeted locus (Fig. 7.4.a). I speculate that controlling Cas9-TR expression levels with miSFITs will enable simple and rapid tuning of downstream genes (Fig.7.4.b,c). A panel of Cas9-TR-miSFIT-variant constructs would be a valuable resource for multiplex gene tuning studies and could even facilitate genome-wide tuning experiments. Scientists could use this tool to determine how the dosage of each gene product contributes to a phenotype of interest.

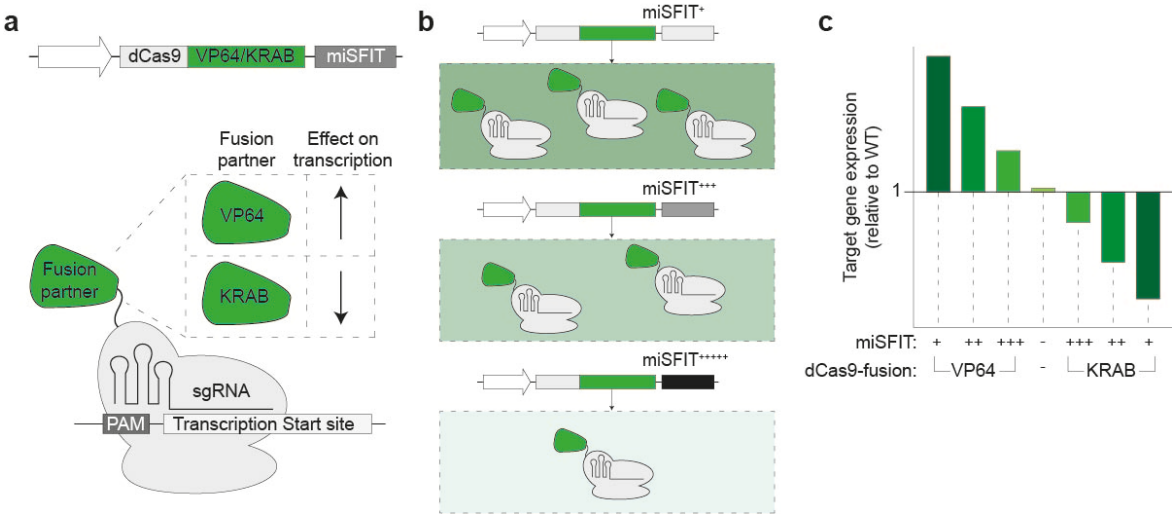

Figure 7.4. Coupling miSFITs with Cas9-transcriptional regulation.
(a) Schematic representation of a vector that enables fine-tuning of dCas9-fusion protein expression levels. **(b)** miSFIT variant strength determines steady-state dCas9-fusion protein abundance. **(c)** Depending of the choice of dCas9 fusion partner, miSFITs can be used to precisely up-, or downregulate endogenous genes.

7.4 Concluding remarks

In this thesis, I have documented my efforts to develop a tuning system suitable for use in mammalian cells. The resulting technology, which I call miSFITs, enables precise tuning of gene output levels in human and murine cells, satisfying the primary objective of this project. In the future, I hope miSFITs will empower scientists to explore how quantitative changes in gene expression contribute to the robust nature of mammalian biology. I envision that miSFITs will help clinicians understand how gene expression levels contribute to disease and I hope that gene tuning will allow biomedical engineers to optimise cellular therapies. Historically, improvements in scientific understanding have often stemmed from technological advances. By offering an alternative to conventional all-or-nothing genetic perturbations, miSFITs provide a new tool in the synthetic biologist's ever-expanding toolbox.

Chapter 8: References

1 Michaels, Y. S. *et al.* Precise tuning of gene expression output levels in mammalian cells. *bioRxiv*, doi:10.1101/352377 (2018).

2 Michaels, Y. S., Wu, Q. & Fulga, T. A. Interrogation of Functional miRNA-Target Interactions by CRISPR/Cas9 Genome Engineering. *Methods Mol Biol* **1580**, 79-97, doi:10.1007/978-1-4939-6866-4_7 (2017).

3 Fulga, T. A., Michaels, Y. S. & Milne, T. A. Method for Modifying Genes. Great Britain patent (2018).

4 Knapp, D. J. *et al.* Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *bioRxiv*, doi:10.1101/342485 (2018).

5 Wu, Q. *et al.* In situ functional dissection of RNA cis-regulatory elements by multiplex CRISPR-Cas9 genome engineering. *Nat Commun* **8**, 2109, doi:10.1038/s41467-017-00686-2 (2017).

6 Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-9367, doi:10.1073/pnas.0903103106 (2009).

7 Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251, doi:10.1016/j.cell.2013.02.014 (2013).

8 O'Connor, L. J. *et al.* Estimating the proportion of disease heritability mediated by gene expression levels. *bioRxiv*, doi:10.1101/118018 (2017).

9 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).

10 Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-212, doi:10.1038/ng.2871 (2014).

11 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

12 Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524, doi:10.1038/nature21411 (2017).

13 Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308 (1981).

14 Banerji, J., Olson, L. & Schaffner, W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**, 729-740 (1983).

15 Amano, T. *et al.* Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* **16**, 47-57, doi:10.1016/j.devcel.2008.11.011 (2009).

16 Razin, A. & Cedar, H. DNA methylation and gene expression. *Microbiol Rev* **55**, 451-458 (1991).

17 Gibney, E. R. & Nolan, C. M. Epigenetics and gene expression. *Heredity (Edinb)* **105**, 4-13, doi:10.1038/hdy.2010.54 (2010).

18 Mitchell, P. J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378 (1989).

19 Xin, B. & Rohs, R. Relationship between histone modifications and transcription factor binding is protein family specific. *Genome Res*, doi:10.1101/gr.220079.116 (2018).

20 Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696-710, doi:10.1016/j.molcel.2016.06.029 (2016).

21 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).

22 Iwakawa, H. O. & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol* **25**, 651-665, doi:10.1016/j.tcb.2015.07.011 (2015).

23 Hershberg, R. & Petrov, D. A. Selection on codon bias. *Annu Rev Genet* **42**, 287-299, doi:10.1146/annurev.genet.42.110807.091442 (2008).

24 Mao, Y., Liu, H., Liu, Y. & Tao, S. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. *Nucleic Acids Res* **42**, 4813-4822, doi:10.1093/nar/gku159 (2014).

25 Floor, S. N. & Doudna, J. A. Tunable protein synthesis by transcript isoforms in human cells. *Elife* **5**, doi:10.7554/eLife.10921 (2016).

26 Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat Rev Mol Cell Biol* **19**, 59-70, doi:10.1038/nrm.2017.83 (2018).

27 Farkona, S., Diamandis, E. P. & Blasutig, I. M. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med* **14**, 73, doi:10.1186/s12916-016-0623-5 (2016).

28 Roybal, K. T. *et al.* Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors. *Cell* **167**, 419-432 e416, doi:10.1016/j.cell.2016.09.011 (2016).

29 Cho, J. H., Collins, J. J. & Wong, W. W. Universal Chimeric Antigen Receptors for Multiplexed and Logical Control of T Cell Responses. *Cell* **173**, 1426-1438 e1411, doi:10.1016/j.cell.2018.03.038 (2018).

30 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

31 McCarthy, E. F. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. *Iowa Orthop J* **26**, 154-158 (2006).

32 Coley, W. B. The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clin Orthop Relat Res*, 3-11 (1991).

33 Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565-1570, doi:10.1126/science.1203486 (2011).

34 von Boehmer, L. *et al.* NY-ESO-1-specific immunological pressure and escape in a patient with metastatic melanoma. *Cancer Immun* **13**, 12 (2013).

35 Nicholaou, T. *et al.* Immunoediting and persistence of antigen-specific immunity in patients who have previously been vaccinated with NY-ESO-1 protein formulated in ISCOMATRIX. *Cancer Immunol Immunother* **60**, 1625-1637, doi:10.1007/s00262-011-1041-3 (2011).

36 Baumeister, S. H., Freeman, G. J., Dranoff, G. & Sharpe, A. H. Coinhibitory Pathways in Immunotherapy for Cancer. *Annu Rev Immunol* **34**, 539-573, doi:10.1146/annurev-immunol-032414-112049 (2016).

37 Freeman, G. J. *et al.* Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J Exp Med* **192**, 1027-1034 (2000).

38 Wei, F. *et al.* Strength of PD-1 signaling differentially affects T-cell effector functions. *Proc Natl Acad Sci U S A* **110**, E2480-2489, doi:10.1073/pnas.1305394110 (2013).

39 Hui, E. *et al.* T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition. *Science* **355**, 1428-1433, doi:10.1126/science.aaf1292 (2017).

40 Keir, M. E., Butte, M. J., Freeman, G. J. & Sharpe, A. H. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol* **26**, 677-704, doi:10.1146/annurev.immunol.26.021607.090331 (2008).

41 Nishimura, H., Nose, M., Hiai, H., Minato, N. & Honjo, T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity* **11**, 141-151 (1999).

42 Prokunina, L. *et al.* A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* **32**, 666-669, doi:10.1038/ng1020 (2002).

43 Hamid, O. *et al.* Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N Engl J Med* **369**, 134-144, doi:10.1056/NEJMoa1305133 (2013).

44 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, doi:10.1016/S0140-6736(14)60958-2 (2014).

45 Michot, J. M. *et al.* Immune-related adverse events with immune checkpoint blockade: a comprehensive review. *Eur J Cancer* **54**, 139-148, doi:10.1016/j.ejca.2015.11.016 (2016).

46 Koyama, S. *et al.* Adaptive resistance to therapeutic PD-1 blockade is associated with upregulation of alternative immune checkpoints. *Nat Commun* **7**, 10501, doi:10.1038/ncomms10501 (2016).

47 Schumann, K. *et al.* Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc Natl Acad Sci U S A* **112**, 10437-10442, doi:10.1073/pnas.1512503112 (2015).

48 Rupp, L. J. *et al.* CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci Rep* **7**, 737, doi:10.1038/s41598-017-00462-8 (2017).

49 Odorizzi, P. M., Pauken, K. E., Paley, M. A., Sharpe, A. & Wherry, E. J. Genetic absence of PD-1 promotes accumulation of terminally differentiated exhausted CD8+ T cells. *J Exp Med* **212**, 1125-1137, doi:10.1084/jem.20142237 (2015).

50 Rosenberg, S. A. & Dudley, M. E. Adoptive cell therapy for the treatment of patients with metastatic melanoma. *Curr Opin Immunol* **21**, 233-240, doi:10.1016/j.coi.2009.03.002 (2009).

51 Coulie, P. G., Van den Eynde, B. J., van der Bruggen, P. & Boon, T. Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat Rev Cancer* **14**, 135-146, doi:10.1038/nrc3670 (2014).

52 Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512 (1987).

53 Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498, doi:10.1038/35078107 (2001).

54 Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180-2196, doi:10.1038/nprot.2013.132 (2013).

55 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

56 Jeppsson, M., Johansson, B., Jensen, P. R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M. F. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. *Yeast* **20**, 1263-1272, doi:10.1002/yea.1043 (2003).

57 Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**, 12678-12683, doi:10.1073/pnas.0504604102 (2005).

58 Ligr, M., Siddharthan, R., Cross, F. R. & Siggia, E. D. Gene expression from random libraries of yeast promoters. *Genetics* **172**, 2113-2122, doi:10.1534/genetics.105.052688 (2006).

59 Blount, B. A., Weenink, T., Vasylechko, S. & Ellis, T. Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One* **7**, e33279, doi:10.1371/journal.pone.0033279 (2012).

60 Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282-1294 e1218, doi:10.1016/j.cell.2016.07.024 (2016).

61 Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950, doi:10.1038/nbt.1568 (2009).

62 Na, D., Lee, S. & Lee, D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Syst Biol* **4**, 71, doi:10.1186/1752-0509-4-71 (2010).

63 Egbert, R. G. & Klavins, E. Fine-tuning gene networks using simple sequence repeats. *Proc Natl Acad Sci U S A* **109**, 16817-16822, doi:10.1073/pnas.1205693109 (2012).

64 Nolan, P. M. *et al.* Implementation of a large-scale ENU mutagenesis program: towards increasing the mouse mutant resource. *Mamm Genome* **11**, 500-506 (2000).

65 Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508-514, doi:10.1038/nature19356 (2016).

66 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

67 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20-36, doi:10.1016/j.cell.2016.10.044 (2017).

68 Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918, doi:10.1016/j.cell.2013.04.025 (2013).

69 Rao, D. D., Vorhies, J. S., Senzer, N. & Nemunaitis, J. siRNA vs. shRNA: similarities and differences. *Adv Drug Deliv Rev* **61**, 746-759, doi:10.1016/j.addr.2009.04.004 (2009).

70 Hsieh, A. C. *et al.* A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res* **32**, 893-901, doi:10.1093/nar/gkh238 (2004).

71 Tafer, H. *et al.* The impact of target site accessibility on the design of effective siRNAs. *Nat Biotechnol* **26**, 578-583, doi:10.1038/nbt1404 (2008).

72 Larsson, E., Sander, C. & Marks, D. mRNA turnover rate limits siRNA and microRNA efficacy. *Mol Syst Biol* **6**, 433, doi:10.1038/msb.2010.89 (2010).

73 Jackson, A. L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179-1187, doi:10.1261/rna.25706 (2006).

74 Charlesworth, C. T. *et al.* Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. *bioRxiv*, doi:10.1101/243345 (2018).

75 Jeggo, P. A. DNA breakage and repair. *Adv Genet* **38**, 185-218 (1998).

76 Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: the present and the future. *Anal Bioanal Chem* **397**, 3173-3178, doi:10.1007/s00216-010-3821-6 (2010).

77 Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *Annu Rev Biomed Eng* **17**, 63-89, doi:10.1146/annurev-bioeng-071813-104938 (2015).

78 Brown, B. D. *et al.* Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat Biotechnol* **25**, 1457-1467, doi:10.1038/nbt1372 (2007).

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

80 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).

81 Ahler, E. *et al.* Doxycycline alters metabolism and proliferation of human cell lines. *PLoS One* **8**, e64561, doi:10.1371/journal.pone.0064561 (2013).

82 Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).

83 Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661, doi:10.1016/j.cell.2014.09.029 (2014).

84 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).

85 Babiskin, A. H. & Smolke, C. D. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol Syst Biol* **7**, 471, doi:10.1038/msb.2011.4 (2011).

86 Babiskin, A. H. & Smolke, C. D. Synthetic RNA modules for fine-tuning gene expression levels in yeast by modulating RNase III activity. *Nucleic Acids Res* **39**, 8651-8664, doi:10.1093/nar/gkr445 (2011).

87 Kittleson, J. T., Cheung, S. & Anderson, J. C. Rapid optimization of gene dosage in E. coli using DIAL strains. *J Biol Eng* **5**, 10, doi:10.1186/1754-1611-5-10 (2011).

88 Brewster, R. C., Jones, D. L. & Phillips, R. Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. *PLoS Comput Biol* **8**, e1002811, doi:10.1371/journal.pcbi.1002811 (2012).

89 Ferreira, J. P., Overton, K. W. & Wang, C. L. Tuning gene expression with synthetic upstream open reading frames. *Proc Natl Acad Sci U S A* **110**, 11284-11289, doi:10.1073/pnas.1305590110 (2013).

90 Curran, K. A. *et al.* Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat Commun* **5**, 4002, doi:10.1038/ncomms5002 (2014).

91 Yofe, I. *et al.* Accurate, model-based tuning of synthetic gene expression using introns in S. cerevisiae. *PLoS Genet* **10**, e1004407, doi:10.1371/journal.pgen.1004407 (2014).

92 Pasini, M. *et al.* Using promoter libraries to reduce metabolic burden due to plasmid-encoded proteins in recombinant Escherichia coli. *N Biotechnol* **33**, 78-90, doi:10.1016/j.nbt.2015.08.003 (2016).

93 Maniatis, T., Goodbourn, S. & Fischer, J. A. Regulation of inducible and tissue-specific gene expression. *Science* **236**, 1237-1245 (1987).

94 Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20-51, doi:10.1016/j.cell.2018.03.006 (2018).

95 Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-1414, doi:10.1016/j.cell.2007.04.040 (2007).

96 Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol* **3**, e85, doi:10.1371/journal.pbio.0030085 (2005).

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

98 Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-4670 (2002).

99 Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419, doi:10.1038/nature01957 (2003).

100 Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-3016, doi:10.1101/gad.1158803 (2003).

101 Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57-68, doi:10.1016/j.cell.2004.06.017 (2004).

102 Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **162**, 96-107, doi:10.1016/j.cell.2015.06.032 (2015).

103 Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-2060, doi:10.1126/science.1073827 (2002).

104 Rehwinkel, J., Behm-Ansmant, I., Gatfield, D. & Izaurralde, E. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11**, 1640-1647, doi:10.1261/rna.2191905 (2005).

105 Behm-Ansmant, I. *et al.* mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20**, 1885-1898, doi:10.1101/gad.1424106 (2006).

106 Chu, C. Y. & Rana, T. M. Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol* **4**, e210, doi:10.1371/journal.pbio.0040210 (2006).

107 Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840, doi:10.1038/nature09267 (2010).

108 Eichhorn, S. W. *et al.* mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* **56**, 104-115, doi:10.1016/j.molcel.2014.08.028 (2014).

109 Djuranovic, S., Nahvi, A. & Green, R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**, 237-240, doi:10.1126/science.1215691 (2012).

110 Meijer, H. A. *et al.* Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science* **340**, 82-85, doi:10.1126/science.1231197 (2013).

111 Tat, T. T., Maroney, P. A., Chamnongpol, S., Coller, J. & Nilsen, T. W. Cotranslational microRNA mediated messenger RNA destabilization. *Elife* **5**, doi:10.7554/eLife.12880 (2016).

112 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, doi:10.1126/science.1205527 (2011).

113 Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods* **9**, 840-846, doi:10.1038/nmeth.2078 (2012).

114 Arvey, A., Larsson, E., Sander, C., Leslie, C. S. & Marks, D. S. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol* **6**, 363, doi:10.1038/msb.2010.24 (2010).

115 Bosson, A. D., Zamudio, J. R. & Sharp, P. A. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell* **56**, 347-359, doi:10.1016/j.molcel.2014.09.018 (2014).

116 Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**, 1278-1284, doi:10.1038/ng2135 (2007).

117 Vainberg Slutskin, I., Weingarten-Gabbay, S., Nir, R., Weinberger, A. & Segal, E. Unraveling the determinants of microRNA mediated regulation using a massively parallel reporter assay. *Nat Commun* **9**, 529, doi:10.1038/s41467-018-02980-z (2018).

118 Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055-1067, doi:10.1016/j.cell.2012.10.036 (2012).

119 Ferry, Q. R., Lyutova, R. & Fulga, T. A. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat Commun* **8**, 14633, doi:10.1038/ncomms14633 (2017).

120 Rowe, H. M. *et al.* Immunization with a lentiviral vector stimulates both CD4 and CD8 T cell responses to an ovalbumin transgene. *Mol Ther* **13**, 310-319, doi:10.1016/j.ymthe.2005.08.025 (2006).

121 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

122 Yu, D. *et al.* Rational design of a monomeric and photostable far-red fluorescent protein for fluorescence imaging in vivo. *Protein Sci* **25**, 308-315, doi:10.1002/pro.2843 (2016).

123 Zhang, M., Pickart, C. M. & Coffino, P. Determinants of proteasome recognition of ornithine decarboxylase, a ubiquitin-independent substrate. *EMBO J* **22**, 1488-1496, doi:10.1093/emboj/cdg158 (2003).

124 Medina, P. P., Nolde, M. & Slack, F. J. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature* **467**, 86-90, doi:10.1038/nature09284 (2010).

125 Sharma, A. *et al.* Posttranscriptional regulation of interleukin-10 expression by hsa-miR-106a. *Proc Natl Acad Sci U S A* **106**, 5761-5766, doi:10.1073/pnas.0808743106 (2009).

126 Lukk, M. *et al.* A global map of human gene expression. *Nat Biotechnol* **28**, 322-324, doi:10.1038/nbt0410-322 (2010).

127 Denzler, R. *et al.* Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Mol Cell* **64**, 565-579, doi:10.1016/j.molcel.2016.09.027 (2016).

128 Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854-859, doi:10.1038/ng.905 (2011).

129 Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837-840, doi:10.1126/science.aae0568 (2016).

130 Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037-1040, doi:10.1126/science.1221551 (2012).

131 Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921-926, doi:10.1038/nature07666 (2008).

132 Prywes, N., Michaels, Y. S., Pal, A., Oh, S. S. & Szostak, J. W. Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying. *Chem Commun (Camb)* **52**, 6529-6532, doi:10.1039/c6cc02692c (2016).

133 Mogilyansky, E. & Rigoutsos, I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* **20**, 1603-1614, doi:10.1038/cdd.2013.125 (2013).

134 Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397-401, doi:10.1038/nature17995 (2016).

135 Vartanian, J. P., Henry, M. & Wain-Hobson, S. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res* **24**, 2627-2631 (1996).

136 Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* **43**, e37, doi:10.1093/nar/gku1341 (2015).

137 Andrews, S. *FastQC A Quality Control tool for High Throughput Sequence Data* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2012).

138 Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233, doi:10.1016/j.cell.2009.01.002 (2009).

139 Muljo, S. A., Kanellopoulou, C. & Aravind, L. MicroRNA targeting in mammalian genomes: genes and mechanisms. *Wiley Interdiscip Rev Syst Biol Med* **2**, 148-161, doi:10.1002/wsbm.53 (2010).

140 Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**, 185-197, doi:10.1016/j.molcel.2004.07.007 (2004).

141 Teteloshvili, N. *et al.* T-cell Activation Induces Dynamic Changes in miRNA Expression Patterns in CD4 and CD8 T-cell Subsets. *Microrna* **4**, 117-122 (2015).

142 Schmitt, T. M. *et al.* Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro. *Nat Immunol* **5**, 410-417, doi:10.1038/ni1055 (2004).

143 Vizcardo, R. *et al.* Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8(+) T cells. *Cell Stem Cell* **12**, 31-36, doi:10.1016/j.stem.2012.12.006 (2013).

144 Clarke, S. R. *et al.* Characterization of the ovalbumin-specific TCR transgenic line OT-I: MHC elements for positive and negative selection. *Immunol Cell Biol* **78**, 110-117, doi:10.1046/j.1440-1711.2000.00889.x (2000).

145 Balch, C. M. *et al.* Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol* **19**, 3635-3648, doi:10.1200/JCO.2001.19.16.3635 (2001).

146 Breslow, A. Prognostic factors in the treatment of cutaneous melanoma. *J Cutan Pathol* **6**, 208-212 (1979).

147 Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9

ribonucleoproteins. *Genome Res* **24**, 1012-1019, doi:10.1101/gr.171322.113 (2014).

148 Knott, S. R. V. *et al.* A computational algorithm to predict shRNA potency. *Mol Cell* **56**, 796-807, doi:10.1016/j.molcel.2014.10.025 (2014).

149 Ui-Tei, K. *et al.* Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* **32**, 936-948, doi:10.1093/nar/gkh247 (2004).

150 Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol* **10**, 81, doi:10.1186/1472-6750-10-81 (2010).

151 Chicaybam, L., Sodre, A. L., Curzio, B. A. & Bonamino, M. H. An efficient low cost method for gene transfer to T lymphocytes. *PLoS One* **8**, e60298, doi:10.1371/journal.pone.0060298 (2013).

152 Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J Biotechnol* **241**, 136-146, doi:10.1016/j.jbiotec.2016.11.011 (2017).

153 Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344, doi:10.1038/nbt.3481 (2016).

154 Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**, 538-542, doi:10.1038/nbt.3190 (2015).

155 Bak, R. O. *et al.* Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *Elife* **6**, doi:10.7554/eLife.27873 (2017).

156 Eyquem, J. *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113-117, doi:10.1038/nature21405 (2017).

157 June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S. & Milone, M. C. CAR T cell immunotherapy for human cancer. *Science* **359**, 1361-1365, doi:10.1126/science.aar6711 (2018).

158 Brown, S. D. *et al.* Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res* **24**, 743-750, doi:10.1101/gr.165985.113 (2014).

159 Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* **12**, 1007-1017, doi:10.1110/ps.0239403 (2003).

160 Bui, H. H. *et al.* Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics* **57**, 304-314, doi:10.1007/s00251-005-0798-y (2005).

161 Lundegaard, C., Lund, O. & Nielsen, M. Prediction of epitopes using neural network based methods. *J Immunol Methods* **374**, 26-34, doi:10.1016/j.jim.2010.10.011 (2011).

162 Buus, S. *et al.* Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens* **62**, 378-384 (2003).

163 Lever, M. *et al.* Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose. *Proc Natl Acad Sci U S A* **113**, E6630-E6638, doi:10.1073/pnas.1608820113 (2016).

164 Huang, J. *et al.* A single peptide-major histocompatibility complex ligand triggers digital cytokine secretion in CD4(+) T cells. *Immunity* **39**, 846-857, doi:10.1016/j.immuni.2013.08.036 (2013).

165 Schroder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521-529 (2002).

166 Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17-27 (1994).

167 DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F. & Jacks, T. Expression of tumour-specific antigens underlies cancer immunoediting. *Nature* **482**, 405-409, doi:10.1038/nature10803 (2012).

168 Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61, doi:10.1016/j.cell.2014.12.033 (2015).

169 Raz, A. *et al.* Cell surface properties of B16 melanoma variants with differing metastatic potential. *Cancer Res* **40**, 1645-1651 (1980).

170 Porgador, A., Yewdell, J. W., Deng, Y., Bennink, J. R. & Germain, R. N. Localization, quantitation, and in situ detection of specific peptide-MHC class I complexes using a monoclonal antibody. *Immunity* **6**, 715-726 (1997).

171 Zehn, D., Lee, S. Y. & Bevan, M. J. Complete but curtailed T-cell response to very low-affinity antigen. *Nature* **458**, 211-214, doi:10.1038/nature07657 (2009).

172 Martin, S. D. *et al.* A library-based screening method identifies neoantigen-reactive T cells in peripheral blood prior to relapse of ovarian cancer. *Oncoimmunology* **7**, e1371895, doi:10.1080/2162402X.2017.1371895 (2017).

173 Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69-74, doi:10.1126/science.aaa4971 (2015).

174 Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-128, doi:10.1126/science.aaa1348 (2015).

175 Schlabach, M. R., Hu, J. K., Li, M. & Elledge, S. J. Synthetic design of strong promoters. *Proc Natl Acad Sci U S A* **107**, 2538-2543, doi:10.1073/pnas.0914803107 (2010).

176 Bellahsene, A. & Forsgren, A. Effect of doxycycline on immune response in mice. *Infect Immun* **48**, 556-559 (1985).

177 Fife, R. S., Sledge, G. W., Jr., Roth, B. J. & Proctor, C. Effects of doxycycline on human prostate cancer cells in vitro. *Cancer Lett* **127**, 37-41 (1998).

178 Redelsperger, I. M. *et al.* Stability of Doxycycline in Feed and Water and Minimal Effective Doses in Tetracycline-Inducible Systems. *J Am Assoc Lab Anim Sci* **55**, 467-474 (2016).

179 Kuwana, Y. *et al.* Expression of chimeric receptor composed of immunoglobulin-derived V regions and T-cell receptor-derived C regions. *Biochem Biophys Res Commun* **149**, 960-968 (1987).

180 Gross, G., Waks, T. & Eshhar, Z. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. *Proc Natl Acad Sci U S A* **86**, 10024-10028 (1989).

181 Kochenderfer, J. N. *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, doi:10.1182/blood-2010-04-281931 (2010).

182 Morgan, R. A. *et al.* Case report of a serious adverse event following the administration of T cells transduced with a chimeric antigen receptor recognizing ERBB2. *Mol Ther* **18**, 843-851, doi:10.1038/mt.2010.24 (2010).

183 Kochenderfer, J. N. *et al.* B-cell depletion and remissions of malignancy along with cytokine-associated toxicity in a clinical trial of anti-CD19 chimeric-antigen-receptor-transduced T cells. *Blood* **119**, 2709-2720, doi:10.1182/blood-2011-10-384388 (2012).

184 Rapoport, A. P. *et al.* NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. *Nat Med* **21**, 914-921, doi:10.1038/nm.3910 (2015).

185 Guo, X. *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* **24**, 978-985, doi:10.1038/s41591-018-0045-3 (2018).

186 Gornalusse, G. G. *et al.* HLA-E-expressing pluripotent stem cells escape allogeneic responses and lysis by NK cells. *Nat Biotechnol* **35**, 765-772, doi:10.1038/nbt.3860 (2017).

187 Torikai, H. *et al.* A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR. *Blood* **119**, 5697-5705, doi:10.1182/blood-2012-01-405365 (2012).

188 Shukla, S. *et al.* Progenitor T-cell differentiation from hematopoietic stem cells using Delta-like-4 and VCAM-1. *Nat Methods* **14**, 531-538, doi:10.1038/nmeth.4258 (2017).

Appendix

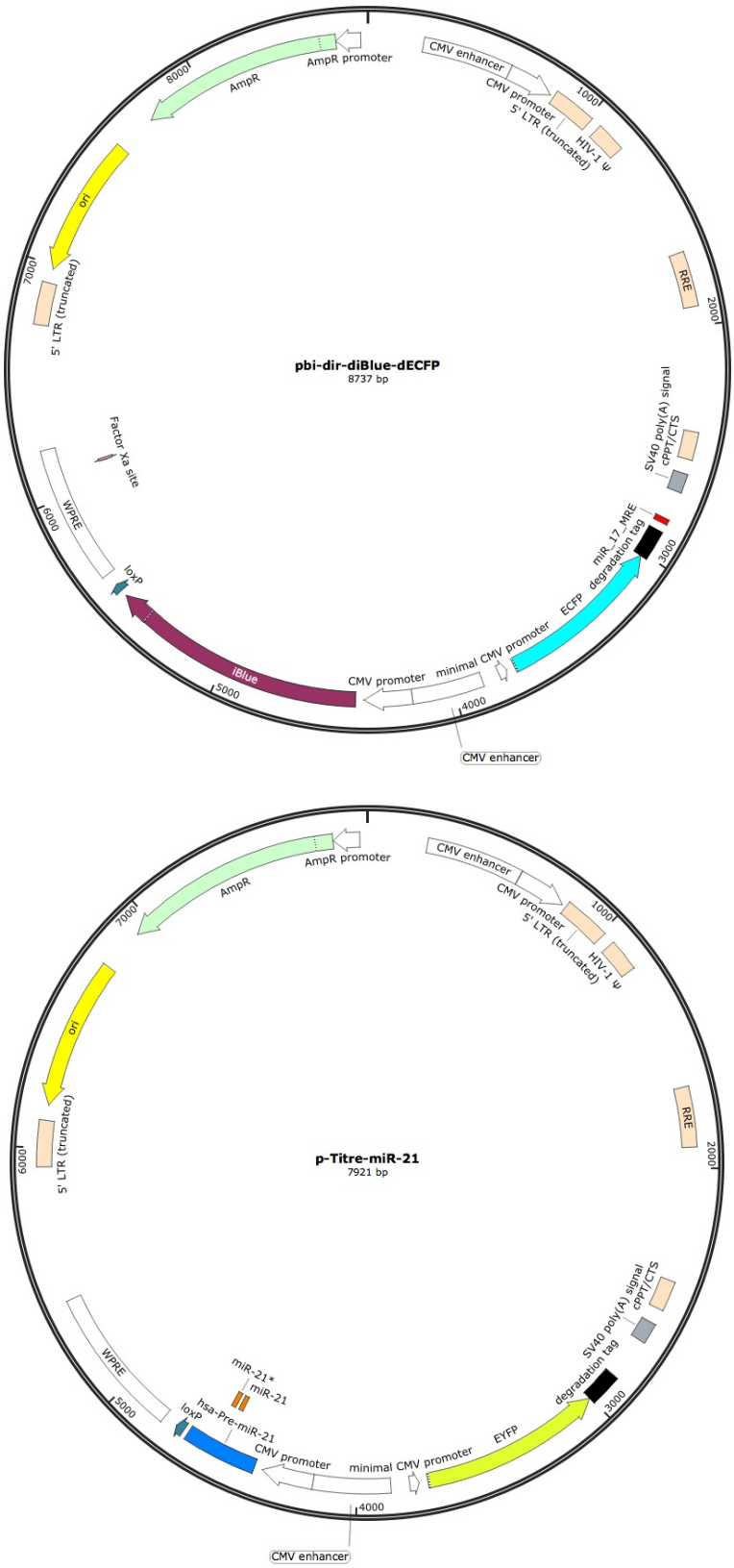

Figure S1. Vectors used for assessing MRE repression kinetics in mammalian cells

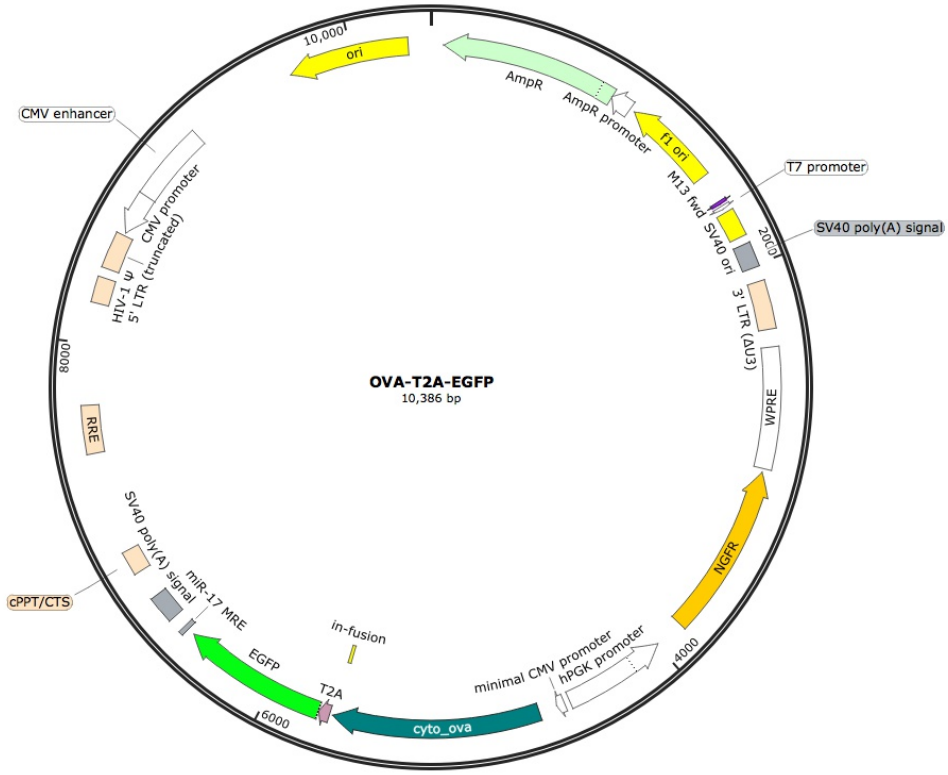

Figure S2. Vectors used for tuning transgene expression

Oligo Name	Oligo Sequence
miR17_Lib_Gen_F	GGTAGACGTCTCTCACCGCATAGCTGCGTATG
miR17_Lib_Gen_R	GATGTCCGTCTCCAAACTACACGGGTCAACATACTAACAC
MRE_Lib_seq_1	CAATTCTGCAGACAAATGGC
MRE_Lib_seq_2	AAAGACCCCAACGAGAAGC
miR_17_mutant_lib	CCGCATAGCTGCGTATGGAGGAAGACAAAGTGCTTACAGTGCAGGTAGTTGCGTGTTAGTATG TTGACCCGTGTAGT
miR_21_mutant_lib	CCGCATAGCTGCGTATGGAGGAAGATAGCTTATCAGACTGATGTTGATTGCGTGTTAGTATGT TGACCCGTGTAGT
bi-dir-Miseq-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGTTTCAGGTTCAGGGGGA
bi-dir-Miseq-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGTGAGTGCTAGCTGACAGG
MRE_qPCR_F	GTGGTATGGCTGATTATGATCCG
MRE_qPCR_R	CGTGAGTGCTAGCTGACAGG
iBlue_qPCR_F	GGTAGTACGTTTCGATGAAGAGG
iBlue_qPCR_R	CAGGCGAACACGGTTATGT
Ova-in-fusion-F	TAGAAGACACCGGAGCTCC
Ova_in_fusion_R2	TCAGGGGAAACACATCTGCC
GFP_in_fusion_F2	GGCAGATGTGTTTCCCCTGAATTCGGCAGTGGAGAG
EGFP-in-fusion-R	CCTAGAAAGAGTAGAGCTAGCACTCACGTGCGAGCTCTAGTTAGAATTCCTTG
PD1_Lenti_Shuttle_F	AGCTCCTGCAGGTGCCGCCATGCAGATC
PDL1_Lenti_Shuttle_R	CTGTCAGCTAGCACTCACGTGC
BBBdir-seq-2	GATCTCAAATCCCTCGGAAG
bi-dir-MRE-seq-1	ATGTTTCAGGTTCAGGGGGA
EXFP_Cloning_Fwd	TCGCCTGGAGCGGCCATGGTGAGCAAGGGCGAG
EXFP_Cloning_Rev	TAAGCTTCTTGTACAGCTCGTCCATGCC
Deg_tag_cloning_Fwd	TGTACAAGAAGCTTAGCCATGGCTT
Deg_tag_cloning_Rev	ACTCACGTGCTACCCCCTACACATTGATCCTAGCAGAAGCACAG
pre_miR21_BiDir_Cloning_Fwd	AGATCCACCGGTCGCCAGTTCGATCTTAACAGGCCAGA
pre_miR21_BiDir_Cloning_Rev	TCGACGAATTCGCGACCAGAGTTTCTGATTATAAACAATGA
sgMMU_PD1_A_F	CACCGGTGGGAAATTCGTAGACTG
sgMMU_PD1_A_R	AAACCAGTCTACGAATTTCCCACC
sgMMU_PD1_B_F	CACCGTGGGAAATTCGTAGACTGG
sgMMU_PD1_B_R	AAACCCAGTCTACGAATTTCCCAC
sgMMU_PD1_D_F	CACCGCTCCAGCTAGCTCAGACCCT
sgMMU_PD1_D_R	AAACAGGGTCTGAGCTAGCTGGAGC
sgMMU_PD1_E_F	CACCGCAGTTCAGCCAAGGCTCCCA
sgMMU_PD1_E_R	AAACTGGGAGCCTTGGCTGAACTGC
sgMMU_PD1_F_F	CACCGCTCCCAGGGTCTGAGCTAGC
sgMMU_PD1_F_R	AAACGCTAGCTCAGACCCTGGGAGC
MMU_PD1_Seq1	CTGTGAGGCAACCCCAA
MMU_PD1_Seq4	ATAACGGGCTTAAACCAAAGG
MMU_PD1_Seq10	TGGCCTCTTTGACCAGATTC
MMU_PD1_Seq11	GTCCCTAGAAGTGCCCAACA
sgIVT_R	AAAAGCACCGACTCGGTGCC
sgIVT-MMU-PD1-D	TAATACGACTCACTATAGGGCTCCAGCTAGCTCAGACCCT

Table S1. Oligonucleotides used in this study

Oligo Name	Oligo Sequence
miSFIT MRE primers	
miR-17-12-C-F	CTAGCCTACCTGCACTCTAAGCACTTTGA
miR-17-12-C-R	CCGGTCAAAGTGCTTAGAGTGCAGGTAGG
miR-17-17-T-F	CTAGCCTACCTGCACTGTAAGTACTTTGA
miR-17-17-T-R	CCGGTCAAAGTACTTACAGTGCAGGTAGG
miR-17-8-G-F	CTAGCCTACCTGGACTGTAAGCACTTTGA
miR-17-8-G-R	CCGGTCAAAGTGCTTACAGTCCAGGTAGG
miR-17-19-G-F	CTAGCCTACCTGCACTGTAAGCAGTTTGA
miR-17-19-G-R	CCGGTCAAACTGCTTACAGTGCAGGTAGG
miR-17-2-perfect-F	CTAGCCTACCTGCACTGTAAGCACTTTGTATCTACCTGCACTGTAAGCACTTTGA
miR-17-2-perfect-R	CCGGTCAAAGTGCTTACAGTGCAGGTAGATACAAAGTGCTTACAGTGCAGGTAGG
BBBdiR-miR-17-F	CTAGCCTACCTGCACTGTAAGCACTTTGA
BBBdiR-miR-17-R	CCGGTCAAAGTGCTTACAGTGCAGGTAGG
BBBdiR-Ce67-F	CTAGCTCTACTCTTTCTAGGAGGTTGTGAA
BBBdiR-Ce67-R	CCGGTTCACAACCTCCTAGAAAGAGTAGAG
Illumina sequencing primers:	
TruSeq D501	AATGATACGGCGACCACCGAGATCTACACtatagcctACACTCTTTCCCTACACGAC
TruSeq D502	AATGATACGGCGACCACCGAGATCTACACatagaggcACACTCTTTCCCTACACGAC
TruSeq D503	AATGATACGGCGACCACCGAGATCTACACcctatcctACACTCTTTCCCTACACGAC
TruSeq D504	AATGATACGGCGACCACCGAGATCTACACggctctgaACACTCTTTCCCTACACGAC
TruSeq D505	AATGATACGGCGACCACCGAGATCTACACaggcgaagACACTCTTTCCCTACACGAC
TruSeq D506	AATGATACGGCGACCACCGAGATCTACACtaatcttaACACTCTTTCCCTACACGAC
TruSeq D507	AATGATACGGCGACCACCGAGATCTACACcaggacgtACACTCTTTCCCTACACGAC
TruSeq D508	AATGATACGGCGACCACCGAGATCTACACgtactgacACACTCTTTCCCTACACGAC
TruSeq D701	CAAGCAGAAGACGGCATACGAGATcgagtaatGTGACTGGAGTTCAGACGTGT
TruSeq D702	CAAGCAGAAGACGGCATACGAGATtctccggaGTGACTGGAGTTCAGACGTGT
TruSeq D703	CAAGCAGAAGACGGCATACGAGATaatgagcgGTGACTGGAGTTCAGACGTGT
TruSeq D704	CAAGCAGAAGACGGCATACGAGATggaatctcGTGACTGGAGTTCAGACGTGT
TruSeq D705	CAAGCAGAAGACGGCATACGAGATttctgaatGTGACTGGAGTTCAGACGTGT
TruSeq D706	CAAGCAGAAGACGGCATACGAGATacgaattcGTGACTGGAGTTCAGACGTGT
TruSeq D707	CAAGCAGAAGACGGCATACGAGATagcttcagGTGACTGGAGTTCAGACGTGT
TruSeq D708	CAAGCAGAAGACGGCATACGAGATgcgcattaGTGACTGGAGTTCAGACGTGT
TruSeq D709	CAAGCAGAAGACGGCATACGAGATcatagccgGTGACTGGAGTTCAGACGTGT
TruSeq D710	CAAGCAGAAGACGGCATACGAGATttcgcggaGTGACTGGAGTTCAGACGTGT
TruSeq D711	CAAGCAGAAGACGGCATACGAGATgcgcgagaGTGACTGGAGTTCAGACGTGT
TruSeq D712	CAAGCAGAAGACGGCATACGAGATctatcgctGTGACTGGAGTTCAGACGTGT

Table S1. Continued

Reagent	Fluorophore	Supplier	Clone	Catalogue number
TruStain fcX (anti CD16/32)	None	BioLegend	93	101320
Live-dead	NIR	BioLegend		423105
CD4	BUV805	BD	GK1.5	564922
CD8	BV711	BioLegend	53-6.7	100747
TCRb	APC	BioLegend	H57-597	109212
CD45.1	PE	BioLegend	A20	110707
CD45.2	PerCP-Cy5.5	BioLegend	104	109828
CD69	BUV737	BD	H1.2F3	564684
CD25	BUV395	BD	Pc61	564022
CD44	AF594	BioLegend	IM7	103054
CD62L	FITC	BioLegend	MEL-14	104405
CTLA4	BV421	BioLegend	UC10-4B9	106311
PD1	PE-Cy7	BioLegend	29F.1A12	135216
Human PD1	PE	eBioscience (Invitrogen)	J105	12-2799-41
NGFR (CD271)	V450	BD	C40-1457	562123
Precision counting beads	None	BioLegend		424902

Table S2. Flow cytometry reagents used in this thesis