# Constructing RNA based genetic circuits for the detection of endogenous microRNAs




**Timothy Rajakumar**

Weatherall Institute of Molecular Medicine

University of Oxford

This dissertation is submitted for the degree of

*Doctor of Philosophy*

St Peter's College

2016

# Abstract

**Constructing RNA based genetic circuits for the detection of endogenous microRNAs**

**Timothy Rajakumar, St Peter's College, DPhil, Michaelmas 2016**

Synthetic biology is a rapidly evolving field enabling the rational design of new biological functionality. RNA is particularly amenable to this rational design processes due to relatively straightforward structural modelling principles and its ability to perform diverse roles, from that of simple messenger RNAs to complex catalytic ribozymes. Consequently, the synthetic design and assembly of RNA building blocks into functional circuits and networks can far exceed the complexity of what could be achieved using protein components. MicroRNAs (miRNAs) are fascinating short non-coding RNA molecules (~22nt long) which have the capacity to bind and tune the expression of thousands of mRNAs, functioning as molecular rheostats during development, physiology and disease. The ability to detect miRNAs, via synthetic biosensor circuits would therefore offer a unique insight into cellular behavior in addition to the ability to link miRNA expression signatures to an output expression programme of choice.

In this thesis, I have developed novel synthetic circuit architectures for the detection of miRNA activity. Firstly, protein and RNA modules were used in concert to develop a biosensor circuit consisting of two serially linked NOT gates that could activate output expression in response to a specific input miRNA. Secondly, methods were devised for the construction of a single molecule biosensor device, built entirely using RNA. At the heart of this device were novel miRNA responsive allosteric ribozymes which have been created through two independent *in vitro* selection methodologies devised for this study. Such engineered modules now constitute part of the synthetic biology toolbox, increasing the versatility with which miRNAs can serve as inputs to artificial genetic circuits.

# Acknowledgements

First I'd like to thank Tudor for taking a chance on me as his first student and nurturing and encouraging my passion for science. He was always there, day or night, to share a scientific success (or more commonly failure) and plan the next steps. Throughout my time in the lab Tudor has been a most inspiring mentor. This work was only possible with the enormous support from my lab colleagues and more importantly friends who have always been there for me both inside and outside the lab. I'm lucky to have shared my PhD journey with the other other students of the lab, Bruno, Quentin, Toni, Yale and Alice. Thanks to the Circulus team (Bruno, Quentin and Toni) for the most successful side project I could have imagined and some of the highlights of my time in the lab. Thanks to Yale and Alice for listening to my endless ribozyme talk and their input to my projects. I thank the post-docs, Markus, Qianxin, Aron and David, for their scientific and life advice. It was a well earnt honour to be accepted into the post-doc coffee club. I'm grateful to have had the opportunity to supervise some incredible students, Paola, Amara and Jacob with whom I was able to do exciting work on ribozymes and circular RNA. Outside the lab, I'll always remember the overly competitive badminton games with Quentin and climbing with the WIMM climbing group. Additionally I need to thank the support from the wider scientific community in the WIMM, including the Sauka-Spengler and Milne labs.

Finally, thanks to my family who have supported me throughout. This probably all started with the microscopes and science kits I was given to play with as a kid. I'm glad to have been encouraged down this path that I have found so interesting and enjoyable. More recently, my family provided the much needed motivation, proof-reading and input that made this thesis possible.

# Contents

**Contents** iii

**List of Figures** ix

**Nomenclature** xiv

**1 Introduction** 1

1.1 Genetic circuits and synthetic biology . . . 1

1.2 Naturally occurring functional RNA species . . . 2

1.2.1 Temperature sensitive RNA . . . 5

1.2.2 Riboswitches . . . 5

1.2.3 Ribozymes . . . 7

1.3 Artificially engineered functional RNA . . . 8

1.3.1 Rational design of RNA regulatory devices . . . 8

1.3.2 *In vitro* selection of RNA aptamers . . . 9

1.3.3 Ribozyme selection . . . 11

1.3.4 Allosteric ribozyme selection . . . 11

1.4 Synthetically engineered, short oligonucleotide responsive, allosteric ribozymes 12

1.4.1 An oligonucleotide regulated ribozyme switch can function within cells 14

1.4.2 Bioinformatic tools for the generation of custom ribozyme logic gates 16

1.4.3 A rational design approach for the construction of oligonucleotide regulated allosteric ribozymes . . . 17

1.5 miRNAs . . . 18

1.5.1 miRNA biogenesis . . . 18

1.5.2 miRNA biochemistry . . . 21

1.5.3 miRNA mediated repression . . . 22

1.5.3.1 Translational inhibition . . . 22

1.5.3.2 mRNA destabilisation . . . 24

1.5.3.3 miRNA repression is mediated by both translational repression and mRNA destabilisation in a time dependent manner 25
1.5.4 miRNAs as potential biomarkers . . . 26
1.5.5 Specific miRNA biomarkers used in this thesis . . . 27
1.6 Sensors for the detection of endogenous miRNA expression . . . 28
1.6.1 Negative miRNA sensors . . . 30
1.6.2 Positive miRNA sensors . . . 31
1.7 Thesis framework . . . 31
1.7.1 Aims and objectives . . . 33

**2 Materials and methods 35**
2.1 DNA work . . . 35
2.1.1 Standard PCR . . . 35
2.1.2 Ribozyme construction by assembly PCR . . . 35
2.1.3 Standard Cloning . . . 35
2.1.4 High efficiency cloning of random ribozyme pools . . . 36
2.1.5 Common expression vectors . . . 37
2.2 RNA work . . . 37
2.2.1 RNA extraction . . . 37
2.2.2 RT-PCR . . . 38
2.2.3 Gene specific RT-qPCR . . . 38
2.2.4 miRNA RT-qPCR . . . 38
2.3 Tissue culture work . . . 39
2.3.1 Cell culture . . . 39
2.3.2 Transfection . . . 39
2.3.3 Flow cytometry . . . 40
2.4 Sequencing . . . 40
2.4.1 MiSeq library preparation . . . 40
2.4.2 RNAseq . . . 42
2.4.3 Sequencing data analysis . . . 42

**3 Construction of a novel double-negative gate miRNA sensor circuit 43**
3.1 Introduction . . . 43
3.2 Results . . . 46
3.2.1 Flow cytometry-based quantification is a valid and readily applicable assay for output reporter protein expression . . . 46

3.2.2 Investigation of the first NOT gate reveals dose-dependent miR-122-mediated repression of L7Ae expression in a model-predicted target site-dependent manner . . . 52
3.2.3 Exploitation of the L7Ae/Kink-turn RNA motif interaction allows for highly regulated repression of the output signal from the second NOT gate . . . 56
3.2.4 The assembled double-negative gate circuit exhibits a markedly reduced ON:OFF ratio than was expected based on the performance of its component NOT gates . . . 60
3.2.5 Simultaneous positive and negative circuit outputs . . . 63
3.3 Discussion . . . 64
3.3.1 Persisting Kt-interaction-related inhibitory levels of L7Ae under conditions of miR-122-mediated repression contribute to the observed suboptimal ON:OFF ratio . . . 64
3.3.2 Exploration of alternative regulator proteins and/or architectures comprising more complex orthogonal circuits . . . 67
3.3.3 Future strategies for the further destabilisation of intermediate protein modules and associated characterisations . . . 67

**4 RT-PCR based selection of miRNA-responsive allosteric ribozymes 69**
4.1 Introduction . . . 69
4.2 Results . . . 73
4.2.1 The wild-type NOT-1 hammerhead ribozyme that is proposed as the basis of NOT gate-type oligonucleotide-responsive allosteric ribozymes functions within the eukaryotic cellular environment . . . 73
4.2.2 Introduction of two ribozyme-independent modifications into the sensor construct led to a marked enhancement of the ON:OFF ratio . . . 74
4.2.3 Introduction of a pair of tandemly located NOT-1 type ribozymes into the sensor construct led to an impairment of the ON:OFF ratio . . . 78
4.2.4 The location of the ribozyme within the mRNA transcript influences the formers inhibitory effect . . . 79
4.2.5 Rationale underlying the design and construction of the diverse RzN8 ribozyme pool intended for *in vitro* selection . . . 82
4.2.6 Selection of allosteric ribozymes via an RNA-based RT-PCR strategy 83

4.2.6.1 OFF state selection for highly active self-cleaving ribozymes yields strong evidence of the desired selection but candidate selected ribozymes exhibit lower than parental levels of cleavage activity . . . 83
4.2.6.2 ON state selection for ribozymes capable of miRNA-mediated allosteric inhibition of cleavage activity was confounded by the large-scale generation and enrichment of indel-containing "cheater" ribozymes . . . 94
4.2.6.3 Additional OFF state selection as a means to filter out indel-containing ribozymes . . . 98
4.2.7 Functional analysis of the evolutionarily-selected candidate Rz13.1 allosteric ribozyme reveals a paradoxical apparent miR-21-associated activation of its cleavage activity . . . 99
4.2.8 Tracking of the selection landscape by high-throughput sequencing provides useful insights into the nature and outcomes associated with the applied selection pressures . . . 104
4.2.9 Functional analysis of the most enriched ribozyme sequence present at the end of the ON state selection fails to reveal detectable evidence of allosteric modulation . . . 111
4.3 Discussion . . . 116
4.3.1 The two selected ribozymes display prerequisite autocleavage activity but do not exhibit detectable miRNA-mediated allosteric modulation . 116
4.3.2 The catalytic activities of ribozyme variants seemingly selected for enhanced cleavage rates by RT-PCR were lower than that of the parent wild-type ribozyme . . . 118
4.3.3 Theoretical additional mechanisms and associated modules by which miRNA-mediated repression could potentially be minimised . . . 120
4.3.4 Lentiviral transduction and cell viability selection pressure . . . 121

**5 RNAseq-based *in vitro* selection for the construction of miRNA-responsive allosteric ribozymes 122**
5.1 Introduction . . . 122
5.2 Results . . . 123
5.2.1 Rationale underlying the design of the three ribozyme pools intended for RNAseq selection . . . 123
5.2.2 Population level analysis of ribozyme and/or miRNA related effects on dEGFP-coding constructs bearing individual ribozyme pools . . . 126

5.2.3 Development of experimental platforms to select for ribozyme activity in the specific presence or absence of miR-21 . . . . . 130
5.2.4 RNAseq selection by way of depletion screening facilitated by quantification of uncleaved ribozymes . . . . . 131
5.2.4.1 Pilot RNAseq selection using RzSeq1 pool ribozymes . . . 134
5.2.4.2 Characterisation of candidate allosteric variants within the RzSeq1 and RzSeq2 ribozyme pools revealed the likely occurance of both miR-21 mediated inhibition of ribozyme cleavage and repression of output expression from the associated ribozyme-bearing transcript . . . . . 139
5.2.5 RNAseq enrichment screen for miRNA-responsive allosteric ribozymes 146
5.2.5.1 Application of an experimental and analytical digital RNAseq protocol reduced technical noise associated with the read count output of a ribozyme enrichment screen . . . . . 147
5.2.5.2 Application of a digital RNAseq-supported enrichment screen of the RzSeq3 pool led to identification of 12 candidate allosteric ribozymes meeting the nominated cut-off thresholds 149
5.3 Discussion . . . . . 152
5.3.1 mRNA-mediated inhibition has the potential to contribute to global reduction in ribozyme expression across a randomised pool and consequent normalisation-related masking and/or indentification of false hits . . . . . 152
5.3.2 Functional allosteric ribozymes may be still hidden amongst the technical noise even when a noise reduction digital RNAseq protocol is utilised . . . . . 155
5.3.3 Parallel DNA/RNA sequencing and potential future innovations supported by massively expanded sequencing offers powerful further opportunities for identification of allosteric ribozymes . . . . . 156
5.3.4 Use of a modified reverse transcription procedure minimises exposure to post RNA extraction *in vitro* conditions favouring ribozyme autocleavage . . . . . 157
5.3.5 Apparent lack of target miRNA binding to the majority of RzSeq2 pool randomised ribozymes potentially accounts for the absence of a miR-21-related allosteric effect . . . . . 158
5.3.6 Using barcodes to enable unrestricted randomisation within the ribozyme sequence . . . . . 159

**6 Conclusions** **160**

6.1 Achievements of the current study: new tools, novel strategies and a more compact miRNA-responsive positive sensor circuit . . . . . 160

6.2 Future development . . . . . 164

6.2.1 Scaling up circuit complexity to allow for multiple inputs, Boolean logic of choice, compact design and/or memory of an input signal . . 164

6.2.2 miRNA-directed apoptosis: an example of a therapeutically relevant outcome . . . . . 165

**References** **166**

**A Plasmid maps** **186**

# List of Figures


1.1 Natural and synthetic RNA based regulators of gene expression . . . . . 4
1.2 Schematic of the systematic evolution of ligands by exponential enrichment (SELEX) process . . . . . 10
1.3 A method for *in vivo* selection of allosteric ribozymes . . . . . 13
1.4 Oligonucleotide regulated allosteric ribozyme . . . . . 15
1.5 miRNA biogenesis and mechanisms of repressive action . . . . . 19
1.6 Schematic representation of both negative and positive miRNA sensor circuits 29
1.7 Circuit architectures capable of positive regulation of output in response to miRNA input (double negative gate circuits) . . . . . 32

3.1 Schematic representation of Boolean NOT gates and genetic circuits that can recapitulate their function . . . . . 44
3.2 Schematic diagram of a biological double-negative gate circuit for the positive detection of miRNA activity . . . . . 47
3.3 Gating of flow cytometry data from HEK293T cells . . . . . 49
3.4 Standard analysis pipeline for generation and validation of flow cytometry data as a means to measure protein and/or gene expression levels . . . . . 51
3.5 Investigation of miR-122 mediated inhibition of a L7Ae regulator . . . . . 53
3.6 Prediction of 3' UTR-located miRNA target sites within the relevant regions of the ECFP-L7Ae ± 122T constructs . . . . . 57
3.7 Investigation of L7Ae mediated translational inhibition . . . . . 58
3.8 Investigation of the function of the fully assembled miRNA positive sensor circuit . . . . . 61
3.9 Simultaneous measurement of the individual output from each of the two constituent NOT gates in the miRNA positive sensor circuit . . . . . 65

4.1 Schematic diagram of a ribozyme-regulated sensor construct . . . . . 71
4.2 An allosteric ribozyme makes for a more direct and condensed circuit . . . . 72

4.3 The NOT-1 hammerhead ribozyme undergoes self-cleavage within a cellular environment . . . 75
4.4 Exploration of ribozyme-independent measures to optimise the NOT-1-mediated ON:OFF ratio . . . 77
4.5 Investigating the ribozyme ON:OFF ratio as a function of the location of the ribozyme within the output mRNA transcript . . . 81
4.6 Putative secondary structure of the parent ribozyme used to generate the RzN8 ribozyme pool comprising variants carrying a nine nucleotide region of randomised sequence . . . 84
4.7 Overview of the ribozyme autocleavage activity-focused selection scheme that was performed as part of this study . . . 85
4.8 Detailed OFF state selection scheme for the selection of active ribozymes . . . 87
4.9 Tracking the selection journey via Sanger sequencing . . . 91
4.10 Nine randomly chosen RzN8 ribozymes present following the 5th selection cycle are capable of inhibiting dEGFP expression . . . 93
4.11 Detailed ON state selection scheme for the enrichment of allosterically inhibited ribozymes . . . 95
4.12 The unintended accumulation of ribozyme-inactivating indels over the ON state selection cycles confounded discovery of allosterically-inhibited ribozyme variants . . . 97
4.13 Further cycles of OFF state selection eliminate the indel-carrying ribozymes and converge on a single ribozyme . . . 100
4.14 Characterisation of the effect of a 3' UTR-located Rz13.1 sequence on transcript abundance and corresponding reporter protein-associated phenotype . . . 102
4.15 The sequences of the ribozyme variants within the pool undergo dramatic and progressive degradation as consequence of the emergence of indels throughout the five cycles of ON state selection . . . 107
4.16 Alignment of high-throughput sequencing data broadly identifies the locations and sizes of indels that have emerged during the ON state selection cycles . . . 108
4.17 Bioinformatic based identification of the most highly enriched indel-free ribozyme sequences present at the end of the $10^{th}$ selection cycle . . . 112
4.18 Characterisation of the effect of a 3' UTR-located Rz5.7 sequence on reporter protein-associated phenotype . . . 114

5.1 Predicted secondary structures of indicative representatives of the RzSeq1, RzSeq2 and RzSeq3 ribozyme pools showing miR-21 binding sites and randomised regions . . . 125

5.2 Population level analysis of individual ribozyme pools in response to the cognate miR-21 . . . 128
5.3 Titration to identify the minimum dose of antagomiR-21 required for inhibition of its cognate miRNA . . . 132
5.4 Schematic of the RNAseq depletion screening strategy utilised in this study . 133
5.5 Schematic of the RNAseq data analysis pipeline implemented in this study . . 135
5.6 Overview of the pilot RNAseq depletion screening strategy pursued in this study 137
5.7 Correlation of data between technical replicates of the pilot RNAseq selection experiment . . . 138
5.8 Analysis of differential expression levels of individual ribozyme variants in the presence and absence of miR-21 reveals candidate allosteric variants . . . 140
5.9 Use of Levenshtein edit distance clustering and heatmap visualisation to explore potential relationships between the sequence of the 6 nt randomised region of RzSeq1 and the observed ribozyme function . . . 142
5.10 Characterisation of RzSeq1-sourced individual candidate allosteric ribozymes identified by depletion screening analysis . . . 145
5.11 Schematic outline of the UMI based digital RNAseq analysis pipeline utilised in this study . . . 148
5.12 Overview of the RNAseq enrichment screening protocol to which theRzSeq3 ribozyme pool was subjected . . . 150
5.13 Digital RNAseq reduces technical noise and thus increases the correlation between data obtained using replicate samples . . . 151
5.14 Differential expression analysis of RzSeq3 ribozyme variants in the presence and absence of miR-21 reveals candidate allosteric variants responsive to this miRNA . . . 153
5.15 Characterisation of individual candidate allosteric RzSeq3 ribozymes identified by enrichment screening . . . 154

# Nomenclature

**Acronyms / Abbreviations**

3-AT 3-amino-1,2,3-triazol

5-FOA 5-fluorooratic acid

Ago Agonaute

antagomiR Antisense competitive inhibitor

CrPV Cricket Paralysis Virus

DAPI 4',6-diamidino-2-phenylindole

DCP2 decapping protein 2

DMSO Dimethyl sulfoxide

DTA Diphtheria toxin A

EBV Epstein-Barr virus

EMCV Encephalomyocarditis Virus

FACS fluorescence assisted cell sorting

FBS fetal bovine serum

GlcN6P glucosamine-6-phosphate

HCV Hepatitis C Virus

HSV-tk herpes simplex virus thymidine kinase

HTS high-throughput sequencing

indel insertion and/or deletion

IRES Internal ribosome entry site

Kt Kink-turn

LNA locked nucleic acid

MALAT1 metastasis-associated lung adenocarcinoma transcript 1

MCS multiple cloning site

miRISC miRNA induced silencing complex

miRNA microRNA

MODC Mouse ornithine decarboxylase

MZT maternal to zygotic transition

NGS next generation sequencing

nt nucleotide

OBS oligonucleotide binding site

oncomiR oncogenic miRNA

PBS phosphate-buffered saline

PEI polyethylenimine

pre-miRNA precursor-miRNA

pri-miRNA primary miRNA

pSILAC pulse labelled stable isotope labelling with amino acids in cell culture

RBS ribosome binding site

RISC RNAi induced silencing complex

RT reverse transcription

rtTA reverse tetracycline-controlled transactivator

Rz ribozyme

SAM S-adenosylmethionine

SELEX systematic evolution of ligands by exponential enrichment

Shld1 Sheild-1

TPP thiamine pyrophosphate

TRE tetracycline response element

UAS upstream activating sequence

UMI unique molecular index

XNA xeno nucleic acid

XRN1 exoribonuclease 1

# Chapter 1

# Introduction

## 1.1 Genetic circuits and synthetic biology

Life is encoded in DNA and governed by the manner in which this information is recalled, via RNA and proteins, in order to exert its genetic programme upon the cell. The control and regulation of this informational flow is thus the ultimate determinant of all living function. Although our understanding is still incomplete, we are beginning to unravel some of the biological tools that comprise the multi-layered regulatory systems including epigenetic, transcriptional, post-transcriptional, translational and finally post-translational control mechanisms. The scale of these regulatory mechanisms has been hinted at by recent data from the ENCODE project showing that despite only 2% of the human genome encoding proteins, up to 80% is transcribed and thus may be involved in non-coding functionalities [1]. A correlation between the quantity of regulatory non-coding DNA and organismal complexity highlights the importance of this machinery and may explain the otherwise paradoxical relationship between complexity and the number of protein coding genes [2]. Whilst ongoing efforts are directed at a finer dissection of these naturally occurring genetic circuit architectures, synthetic biology offers an alternative approach whereby the construction and characterisation of artificial biological parts and circuits can offer an insight into the operation of endogenous counterparts. At the same time, synthetically constructed circuits offer promise in the fields of biological engineering, bio-manufacturing and bio-fuels in addition to constituting potential diagnostic and therapeutic tools [3].

Despite the traditional view that proteins were the ultimate cellular effector molecules, RNA may be a competitor for this title given the emergence of vast quantities of non-coding RNA species that far exceeds that of protein coding transcripts [4]. From a synthetic standpoint, RNA is also a favoured tool given that is possesses a number of qualities including well defined intramolecular base pair interactions that contribute to secondary structure, amenabil-

ity to computational structural prediction, ease of manipulation using basic molecular biology tools and an ability to perform in signalling, structural and catalytic roles. Given these properties, RNA has already been well exploited by nature and found to perform a diverse range of roles within the cell [5]. These naturally occurring functional RNA modules are the first and foremost source for the harvesting of potential parts for use in synthetic circuits however powerful tools have also been developed for the *de novo* design of functional RNA molecules [6–8].

Such engineered functional modules can be wired together to establish synthetic biosensor circuits, which can generate a chosen output signal in response to the detection of a specific input trigger. The implementation of basic biosensor circuits has already proved to be an important tool in scientific research enabling an indirect means to measure the levels of a multitude of biologically important molecules. In this thesis, the construction of synthetic biosensor circuits for the detection of miRNAs has been described. miRNAs were chosen as the input to this system for a number of reasons including their importance in the processes of health and disease and therefore their potential as potent biomarkers, the ability to engineer binding sites with high confidence through the use of base pair complementarity, and the potential to integrate information from multiple input miRNA signals for higher order signal processing. Ultimately such tools may have utility in the differentiation of cell types and cellular states based on the detection of characteristic miRNA expression signatures.

## 1.2 Naturally occurring functional RNA species

According to the RNA world hypothesis, RNA was once the sole carrier of genetic information as well as the ultimate cellular effector molecule, performing catalytic, structural and signalling roles [9]. RNA has since been superseded by DNA regarding its role in information storage, and by proteins as the biological catalysts in most major forms of life. RNA still performs multiple vital roles in the cell, including the transfer of genetic information from DNA to protein via polypeptide synthesis catalysed by RNA in the ribosome.

Importantly, biological systems have evolved a multitude of regulatory building blocks comprised of RNAs that fall within endogenous gene expression networks. These include riboswitches, which can exert their regulatory effect on either transcription or translation in response to specific input signals [10]. Allosteric self-cleaving ribozymes can be used to elicit similar regulatory responses [10]. Eukaryotic organisms have also evolved complex regulatory systems for the post-transcriptional repression of target transcripts by short RNA guided protein complexes [11]. A subset of these RNA based regulatory devices are comprised entirely of *cis*-acting RNA. Their mechanistic simplicity has made them prime candidates

for the adaptation and incorporation into synthetically engineered genetic circuits. Following below is a brief discussion of the molecular mechanisms by which some of these RNA devices can function (Fig. 1.1).

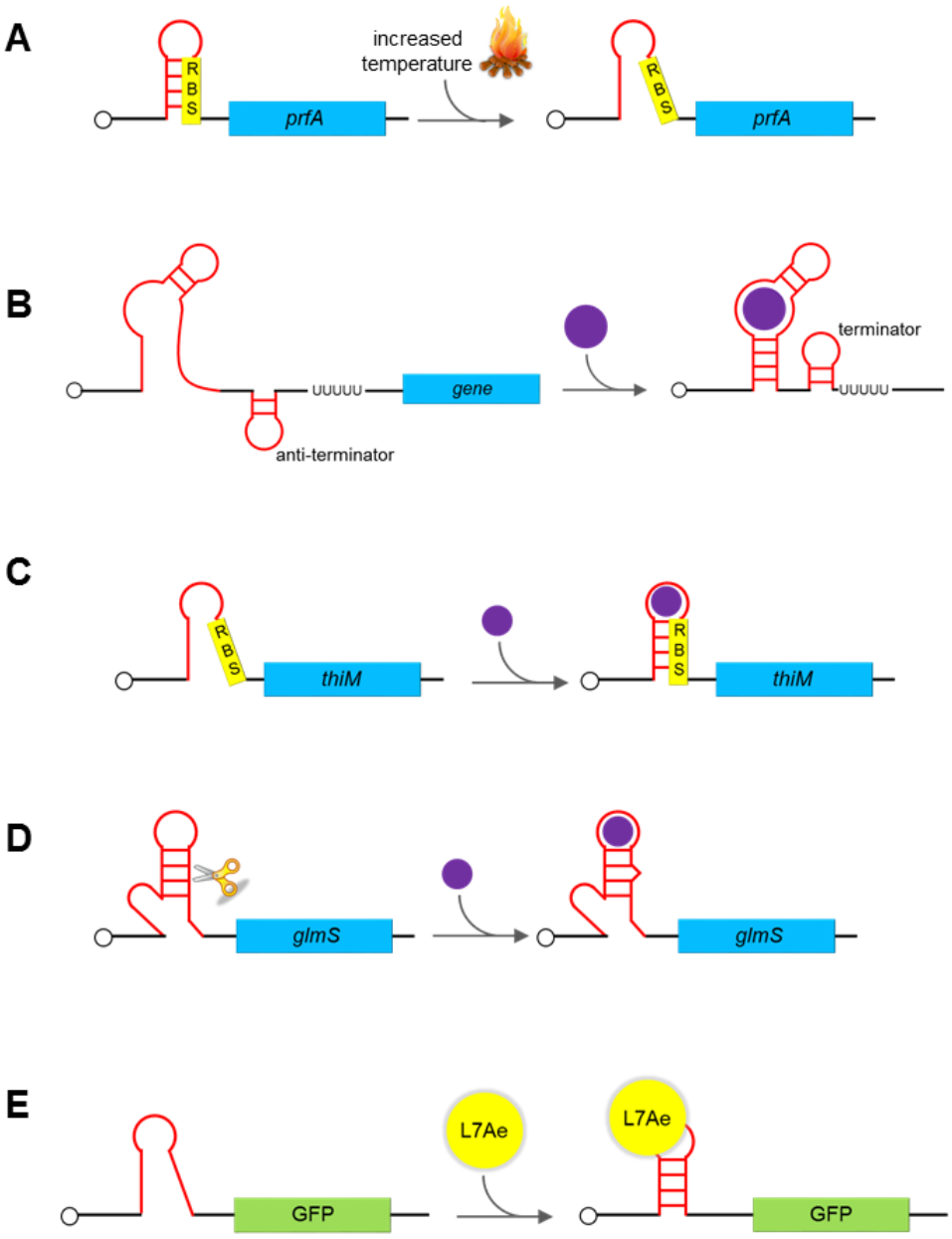



Figure 1.1: Natural and synthetic RNA based regulators of gene expression
*(A) A 5' UTR located thermosensor functions to inhibit gene expression at low temperatures through sequestration of the ribosome binding site (RBS) and inhibition of translation. The structure of the thermosensor becomes destabilised upon an increase in temperature, freeing the RBS and permitting translation. (B) Riboswitches can function through the regulation of transcription termination. Here the mRNA leader sequence (5' UTR) contains a riboswitch which adopts an anti-terminator structure in the absence of ligand, permitting the transcription of the downstream coding sequence. A small molecule ligand can bind to the riboswitch and induce a conformational change to form a terminator. Now transcription is prematurely terminated and expression of the downstream gene is inhibited. (C) Riboswitches can also function to regulate translation. Here a 5' UTR riboswitch is of insufficient stability to sequester the RBS (or in eukaryotic systems, to impede the scanning ribosome) and permits translation. Continued...*

### 1.2.1 Temperature sensitive RNA

The secondary structure of RNAs is highly dependent on temperature, a fact exploited by the simplest class of RNA-based regulatory devices. These "RNA thermosensors" undergo defined structural alterations upon changes in temperature, mediated by melting of intramolecular Watson-Crick base pairs, which leads to the modulation of gene expression [14]. This can be achieved in prokaryotic organisms with 5' UTR RNA thermosensor stem-loop structures that include the ribosome binding site (RBS) and translation initiation codon. At low temperatures, this thermosensor will adopt a highly stable conformation in which the RBS and translation initiation codon are unavailable for ribosome recruitment and translation. Upon an increase in temperature, melting of the stem-loop may occur resulting in the release of RBS and start codon, permitting translation initiation (Fig. 1.1A).

Such a strategy has naturally evolved to regulate the expression of virulence genes in the bacterium *Listeria monocytogenes* through a thermosensor residing in the 5' UTR of the *prfA* transcript, which is only expressed in the warm environment of an animal host [15] (Fig. 1.1A). RNA-based thermosensors have been shown to act via a similar mechanism to regulate expression of the *hspA* heat shock gene [16]. More recently the Hartig laboratory have exploited this principle to generate artificial allosteric ribozymes that undergo modulation of cleavage activity in response to temperature changes [17].

### 1.2.2 Riboswitches

Riboswitches are a diverse class of RNA based *cis*-acting regulatory devices. Encoded in the untranslated regions of host genes, riboswitches have been found to be involved in the regulation of transcription and translation [10]. The regulation occurs in response to small molecules or metabolite ligands, which bind to a pocket in the riboswitch thus altering its structure and its gene tuning activity. Frequently riboswitches are found as components in

*Figure 1.1 continued: The riboswitch undergoes structural stabilisation upon binding a small molecule ligand and now sequesters the RBS (or impedes the scanning ribosome) and thus inhibits translation. (D) A 5' UTR allosteric ribozyme will catalyse transcript cleavage resulting in inhibition of gene expression. The ribozyme will undergo allosteric modulation upon binding of a small molecule ligand, attenuating its cleavage activity and thus activating expression of the downstream gene. (E) This synthetic riboswitch has been developed to regulate translation in eukaryotic cells [12, 13]. The riboswitch alone is of insufficient stability to inhibit progression of the scanning ribosome through the 5' UTR and therefore permits translation. The structure of the riboswitch becomes stabilised upon addition of an cognate RNA binding protein and adopts a structural conformation that prevents passage of the ribosome, thus inhibiting translation.*

negative feedback circuits that regulate the homeostatic maintenance of metabolic processing pathways [18]. To this end, a riboswitch responsive to a specific metabolic ligand is used to negatively regulate the expression of biosynthetic enzymes used in the production of the same specific metabolite. Consequently, increased metabolite concentrations result in reduced rates of biosynthesis, and *vice versa*.

Riboswitches have been implicated in the regulation of transcription termination. Prokaryotes utilise a mechanism of Rho-independent transcription termination in which the RNA polymerase stalls upon encountering a stable GC rich terminator stem-loop followed by a string of uracil residues [19]. Riboswitches have evolved to regulate this process through the incorporation of anti-terminator sequences, which can interact with the terminator stem-loop and alter its structure in a manner that no longer sustains transcriptional termination [10]. Metabolite binding aptamer sequences have been found to precede a number of anti-terminator sequences to create riboswitches that can regulate transcription termination in a metabolite-dependent manner (Fig. 1.1B). These have been found to function both as positive and negative regulators of transcription termination, as described for the adenine [20] and a S-adenosylmethionine (SAM) [21] riboswitches respectively.

Other riboswitches function to regulate gene expression at the level of translation. These are frequently located in the 5' UTR of genes where they function to regulate accessibility to the RBS and start codon and thus modulate the efficiency of translation initiation. The thiamine pyrophosphate (TPP) riboswitch was found through the identification of conserved 5' UTR sequences in the *thiM* transcript that functioned to regulate gene expression in a protein-independent manner. When bound by TPP, this riboswitch adopted a structural conformation in which the RBS was sequestered thus inhibiting the initiation of translation in a TPP-dependent manner [22] (Fig. 1.1C).

More recently riboswitches have been identified in eukaryotic organisms where they have been proposed to regulate gene expression through the modulation of mRNA splicing. A TPP riboswitch has been found in the *NMT1* gene (involved in TPP metabolism) in the fungus *Neurospora crassa* where it functions to mediate a structural switch between states that direct the generation of alternatively spliced transcript isoforms [23]. Splicing of the functional protein coding isoform is inhibited by the riboswitch in the presence of TPP thus reducing the expression of *NMT1* and forming a negative feedback circuit.

Remarkably, naturally occurring riboswitches have been found which utilise multiple ligand binding aptamer domains and/or regulatory domains to enable more complex information processing. A riboswitch from *Vibrio cholerae* has been identified containing two glycine aptamers which function cooperatively to regulate the formation of a transcriptional terminator stem-loop [24]. This was proposed to facilitate significantly enhanced sensitivity to the

input glycine ligand. A similar riboswitch has been found in *Bacillus anthracis*, consisting of tandem TPP aptamers that function in a non-cooperative manner to reduce the ligand concentration changes required to induce switching behaviour [25]. Riboswitches containing two different aptamer domains enable the processing of two independent input signals and are thus capable of processing Boolean logic functions. A riboswitch from *Bacillus clausii* can function as a molecular NOR gate in response to input signals from SAM and coenzyme B12 [26]. These independent SAM and coenzyme B12 aptamer domains, both within the 5' UTR of the *metE* transcript, are each associated with terminator stem-loops which become more stable in the presence of their cognate ligands. Consequently, increasing levels of either SAM or coenzyme B12 will inhibit expression of the *metE* gene through an increased rate of premature transcriptional termination.

### 1.2.3 Ribozymes

Ribozymes are RNA molecules that catalyse a chemical reaction in an equivalent manner to protein enzymes. Originally met with great scepticism, the seminal work in the Cech and Altman laboratories during the 1980s demonstrated the ability of RNA to serve as a catalyst [27, 28]. Self-cleaving ribozymes were originally discovered in viroids and satellite RNAs, where they were found to be involved in the process of rolling circle amplification [29]. Subsequently, such ribozymes have been found in all kingdoms of life. The process by which ribozymes can be identified has been greatly aided in recent years by comparative analysis of whole genome sequencing data leading to the discovery of new ribozyme classes and knowledge of their widespread occurrence [30–32]. Within the group of self-cleaving ribozymes, cleavage typically occurs through a mechanism of phosphodiester isomerisation whereby an internal 2'-OH attack results in the generation of a 2',3'-cyclic phosphate and a 3' hydroxyl terminus. First described in 1986, the hammerhead ribozyme is one of the most studied self-cleaving ribozymes and as such has been heavily exploited for the generation of synthetic devices [29].

To function as ligand inducible gene expression switch, ribozyme self-cleavage activity must be susceptible to modulation by a small molecule ligand in an allosteric manner. To date, the *glmS* ribozyme from *Bacillus subtilis* is the only naturally occurring allosteric ribozyme that has been shown to function in this manner [33, 34]. This ribozyme is located immediately upstream of the protein coding sequence in the 5' UTR of the *glmS* gene, the enzyme product of which catalyses the synthesis of glucosamine-6-phosphate (GlcN6P) [35]. The *glmS* ribozyme functions in a negative feedback circuit to regulate levels of GlcN6P. GlcN6P binds to and allosterically activates the self-cleavage activity of the *glmS* ribozyme leading to the generation of a transcript incapable of directing protein translation (Fig. 1.1D).

## 1.3 Artificially engineered functional RNA

Particularly attractive from an engineering standpoint has been the ease with which RNA molecules capable of demonstrating novel functionality can be constructed. The relative simplicity of RNA molecules, comprised of combinations of only four different monomers, compared to the twenty different amino acids that may occur within proteins, has aided this process of design. RNA functionality is largely determined by secondary structure, which itself is determined by the well-defined rules of the canonical Watson-Crick base pairs and the GU wobble pair. Thus the development of algorithms for the prediction of secondary structure, tertiary structure and folding kinetics has made RNA amenable to a rational design process [36–40]. Additionally, the wealth of biochemical tools at our disposal for the manipulation of RNA in an *in vitro* setting further enhance its ability to be engineered. The interconversion between DNA and RNA by RNA polymerase and reverse transcriptase enzymes enables the production and amplification of synthetic variants.

### 1.3.1 Rational design of RNA regulatory devices

Synthetic RNA regulatory devices can be designed to mimic the naturally occurring architecture in which a ligand binding aptamer domain is used in conjunction with a gene regulatory module, for example an anti-terminator, RBS or self-cleaving ribozyme. This modularity has enabled the mix and match of naturally occurring as well as artificially designed components for the generation of novel functionality.

The most conceptually simple class of artificial RNA regulatory device has been designed for the regulation of translation in eukaryotic organisms. Recruitment of the small ribosomal subunit, an initiating step in translation, occurs at the 5' cap from which point it must scan the 5' UTR until it encounters an ATG start codon. RNA secondary structures of sufficient stability can impede the scanning ribosome and thus reduce the efficiency of translation. In order to exploit this situation, Werstuck and Green have inserted an RNA aptamer for a Hoechst dye into the 5' UTR of a reporter gene [41]. In the absence of ligand, the cognate Hoechst aptamer lacked sufficient stability to hinder the scanning ribosome, therefore enabling the efficient translation of the reporter gene. However in the presence of Hoechst, the RNA aptamer undergoes structural stabilisation and consequently inhibits the progression of the nascent ribosome complex and thus translation. Expanding this concept, Saito *et al.* have located a protein binding aptamer within the 5' UTR of reporter transcripts to achieve protein dependent control of translational efficiency [12, 13]. The naturally occurring RNA binding protein (L7Ae) and its cognate aptamer (Kink-turn [Kt] motif) were appropriated for use in this system. The Kt motif alone lacks sufficient stability to obstruct translation, yet when bound by L7Ae poses enough

of a hurdle to block the scanning ribosome (Fig. 1.1E).

More complex RNA regulatory devices can be constructed through the engineering of ribozymes, capable of allosteric modulation. A wealth of self-cleaving ribozymes have been identified and isolated from nature in addition to a diverse group of ligand binding RNA aptamers. It was therefore hypothesised that the rational fusion of an aptamer and ribozyme may result in the generation of an allosteric ribozyme, in which catalytic activity could be modulated by a specific input ligand. Such a strategy was employed by Tang and Breaker, who fused an ATP binding aptamer to stem II of a minimal hammerhead ribozyme [42]. The resulting ribozyme demonstrated allosteric modulation, in which its catalytic rate was reduced approximately 180 fold in the presence of its cognate ATP ligand. Since this initial demonstration, similar principles have been applied to rationally design allosteric ribozymes responsive to a diverse set of ligands including flavin mononucleotide, theophylline and others [43].

### 1.3.2 *In vitro* selection of RNA aptamers

RNA has proved to be particularly amenable to *in vitro* selection and evolution techniques as a result of the direct link between phenotype and genotype. This process has been aided by our ability to chemically synthesise oligonucleotides of random sequence and amplify these using the well-established tools of molecular biology. Application of these techniques has enabled the isolation of RNA sequences capable of performing basic functions.

Aptamers are short RNA molecules capable of binding to a specific ligand with high affinity. Abundant examples of aptamers have been found to occur in natural biological systems which can be incorporated into synthetic systems. More powerful still is the ability to select aptamers against theoretically any ligand through the process of systematic evolution of ligands by exponential enrichment (SELEX), a technique developed simultaneously by the Gold and Szostak laboratories [6, 7]. To generate a pool of random oligonucleotide sequences suitable for SELEX, a random sequence was flanked by known sequences at the 5' and 3' termini to enable amplification by PCR. SELEX was performed through a cyclical process involving the capture of random oligonucleotides by immobilised ligands, washing to remove any weak or non-specific binding followed by elution to recover an enriched pool of potential aptamer sequences. These selected sequences were exponentially amplified by RT-PCR and used as templates for *in vitro* transcription to regenerate a pool of RNA from which aptamer sequences can once again be selected for (Fig. 1.2). At each cycle of selection it is possible to increase the stringency of selection based on the binding and washing conditions. Following the initial description of SELEX, aptamers have been selected with affinity to a range of targets including metal ions [44], short peptides [45], the ribosome [46], intact virions [47], and red blood cells [48].

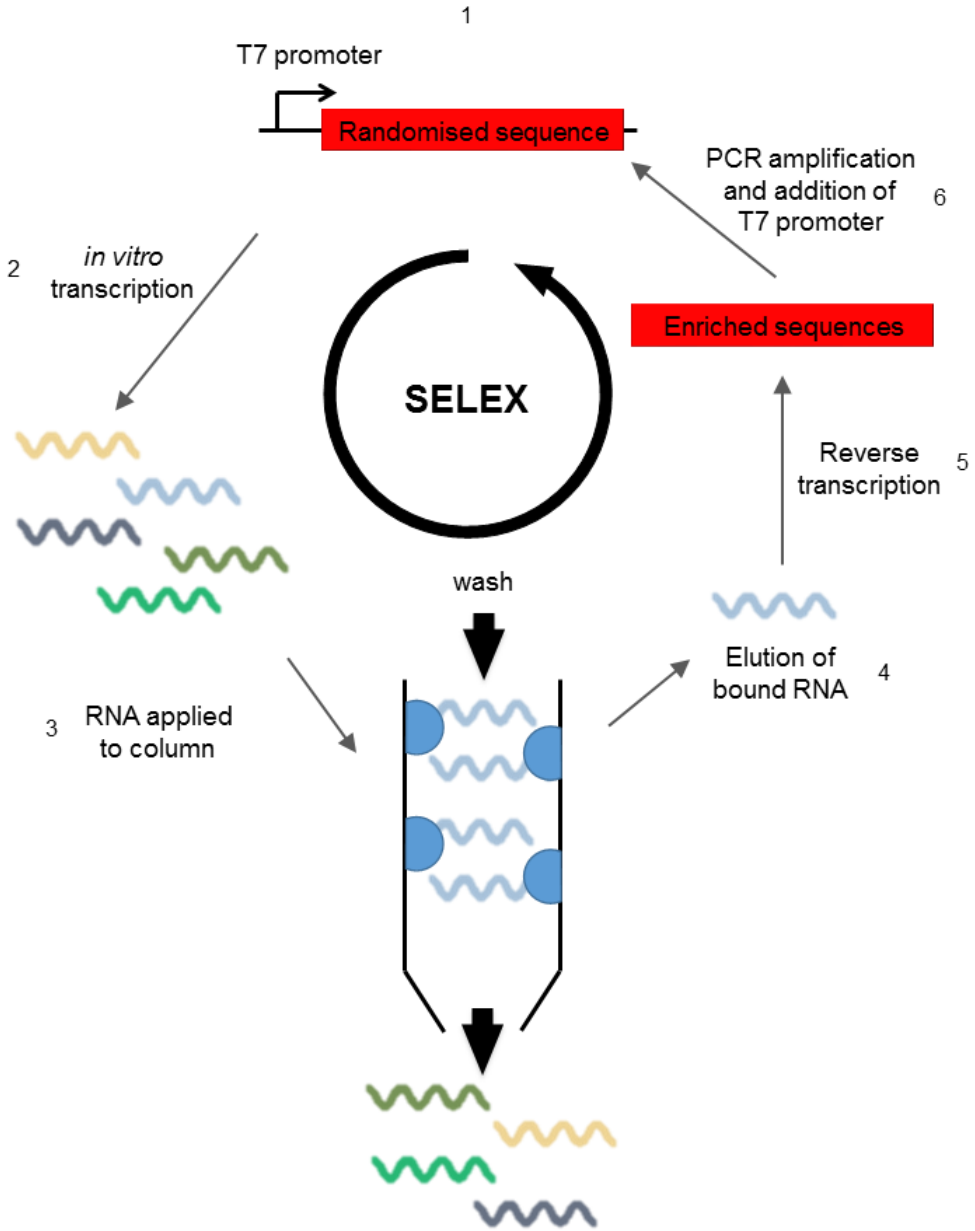



Figure 1.2: Schematic of the systematic evolution of ligands by exponential enrichment (SELEX) process
*(1) A pool library of randomised DNA oligonucleotides is generated downstream of a T7 promoter. (2) In vitro transcription with the T7 RNA polymerase is used to generate a pool of randomised RNA species. (3) The RNA is applied to a column containing a ligand of interest. Potential RNA aptamers with affinity for the cognate ligand will remain trapped in the column, whilst non-binding species will pass through. (4) Bound RNA is eluted, (5) converted to cDNA by reverse transcription, (6) and extended and amplified by PCR to generate an enriched pool of random species, upon which to start the entire process again.*

### 1.3.3 Ribozyme selection

Another breakthrough came when Bartel and Szostak showed that new catalytic function could be identified within a large pool of random RNA sequences by a process of selection. In their landmark paper [8] they selected for ribozymes capable of catalysing a ligation reaction from a pool of $1.6 \times 10^{15}$ unique sequences. Members of this random pool consisted of a fixed sequence at the 5' and 3' ends surrounding an internal stretch of 220 random nucleotides. The sequence of the 5' region was designed such that it could interact with the 3' end of a fixed substrate anchor oligonucleotide and position it in such a way as to orient the 3' OH group at the 3' terminus of the substrate in close proximity to the 5' PPP group at the 5' terminus of the random pool oligonucleotides to which it may ligate. Following an incubation period, RT-PCR amplification was performed with primers targeted to the 5' end of the anchor oligo and the 3' fixed region of the random oligonucleotide. Thus amplification was only possible from a template produced by ligation of an anchor to random oligonucleotide. These captured species were converted back to RNA through *in vitro* transcription and the process was restarted. By the end of 10 selection cycles, ribozymes had been isolated with the capability to catalyse their ligation at a rate 7 million fold greater than the uncatalysed reaction.

This initial proof of principle was rapidly proceeded by further work demonstrating the ability to select artificial ribozymes capable of catalysing a varied range of chemical reactions [49–54]. These techniques were not limited to RNA, as it was later shown that random sequences of DNA could be selected for the possession of catalytic activity [55, 56]. Furthermore, this concept has since been extended to apply to xeno nucleic acid (XNA) molecules and even oligonucleotides incorporating non-natural bases [57–59].

### 1.3.4 Allosteric ribozyme selection

The simple fusion of an aptamer and ribozyme is not always sufficient for the engineering of allosteric ribozymes, thus a more universally applicable approach involves the *in vitro* selection of a linker sequence joining the two functional domains. This involves the sequential selection, in the presence and absence of a specific ligand, for either highly active or inactive catalysis, depending on the desire for allosteric activation or inhibition.

Klauser *et al.* elegantly applied this principle to develop an *in vivo* selection method for small molecule-responsive allosteric ribozymes in yeast [60]. A pool of candidate ribozymes was created, in which a neomycin aptamer [61, 62] was fused to stem III of a hammerhead ribozyme via a random 5 nucleotide (nt) linker. These ribozymes were cloned into the 3' UTR of the yeast GAL4 transcriptional activator, expression of which would inversely correlate with the catalytic activity of its paired ribozyme. In coupling their ribozyme to the expression of

the GAL4 transcriptional activator, Klauser *et al.* created a versatile system which enabled the indirect linkage between ribozyme activity and any gene downstream of a GAL4 responsive, upstream activating sequence (UAS) element [60]. For the first time, this system enabled the dual positive and negative selection of ribozyme cleavage activity from a single library (Fig. 1.3).

Positive selection was achieved through the transformation of the ribozyme regulated library into a yeast strain in which the essential HIS3 gene was under UAS regulated transcriptional control. As a result, only inactive ribozymes, which permitted the expression of GAL4, would result in viable yeast cells expressing HIS3. Furthermore, the drug 3-amino-1,2,3-triazol (3-AT), an inhibitor of the HIS3 gene product could be used to increase the stringency of positive selection such that only the cells that expressed the highest levels of HIS3, corresponding to those harbouring the least active ribozymes, could survive.

Conversely, negative selection was performed by transforming the ribozyme library into a yeast strain containing a UAS regulated URA3 gene. Exogenously delivered 5-fluorooratic acid (5-FOA) was converted to a toxic product by the URA3 enzyme, therefore only cells containing ribozymes that possessed sufficient catalytic activity to silence GAL4 expression were able to survive. Once again exogenous control over selection stringency was governed by the concentration of 5-FOA applied to the cells. This system is amenable for use in selecting allosteric ribozymes whose catalytic activity can be either activated or inhibited by ligand binding depending on whether the ligand is supplied during the positive or negative selection steps. Using this system Klauser *et al.* [60] were able to select a neomycin responsive allosteric hammerhead ribozyme (neo-B2). Finally, the activity of neo-B2 was assessed by using the ribozyme regulated GAL4 to control the expression of a LacZ reporter gene. The activity of the neo-B2 allosteric ribozyme in the absence or presence of neomycin (ON:OFF ratio) was estimated to be approximately 0.04 based on an 25 fold reduction in β-galactosidase activity.

## 1.4 Synthetically engineered, short oligonucleotide responsive, allosteric ribozymes

Short oligonucleotides are particularly attractive input ligands for allosteric modulation of ribozyme cleavage activity, due to the relative ease with which binding sites can be engineered into a ribozyme based on Watson-Crick base pair complementarity. In contrast, designing small molecule binding sites represents a much more complex task, which often requires using naturally occurring and evolutionarily optimised aptamers or a process of *in vitro* selection.

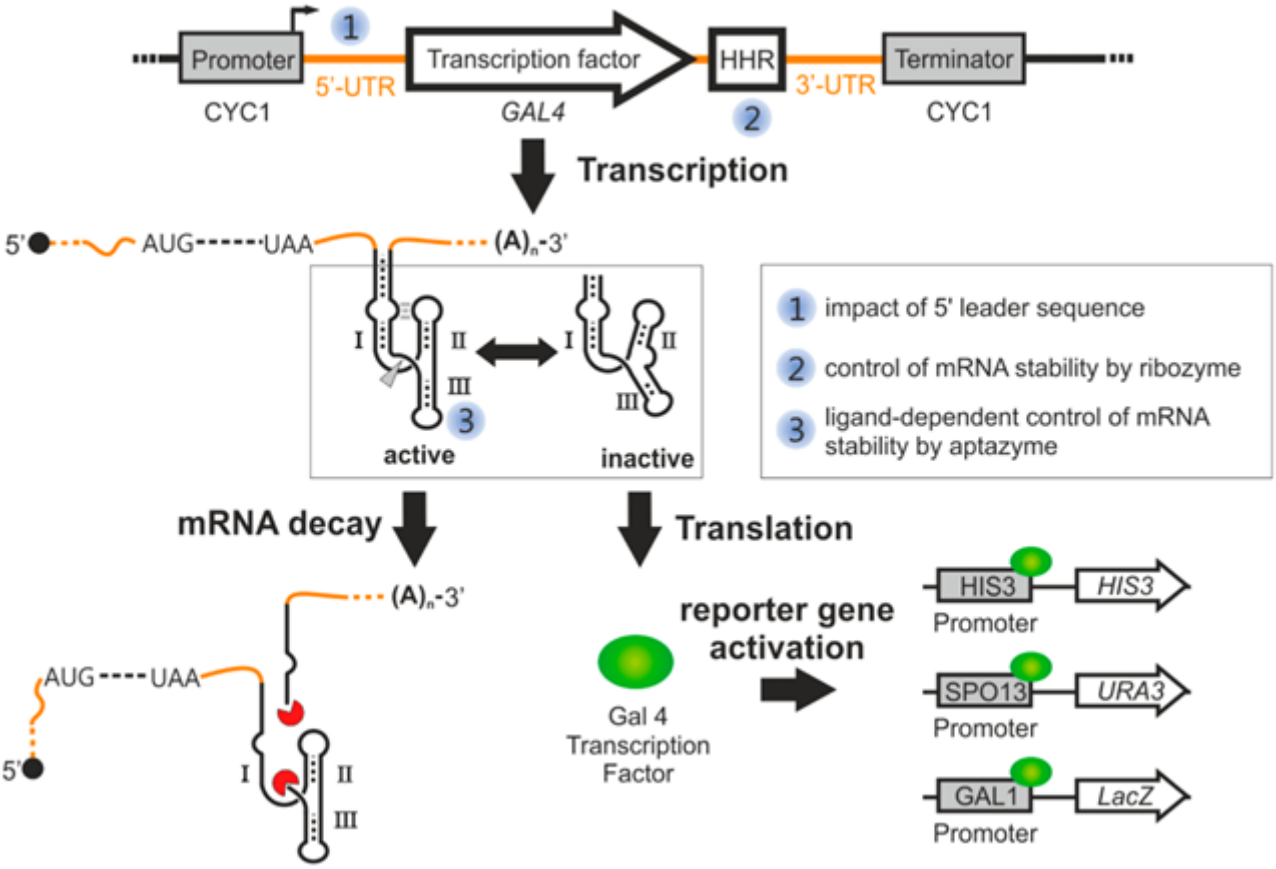



Figure 1.3: A method for *in vivo* selection of allosteric ribozymes

*Klauser et al. have described the above method in which a pool of randomised ribozyme sequences have been cloned into the 3' UTR of a GAL4 transcription factor gene. The expression of GAL4 is thus dependent on (1) the 5' leader which should impact all ribozyme clones in an approximately equal manner, (2) the stability of the mRNA transcript as determined by the cleavage activity of the particular ribozyme clone which it harbours, and (3) any potential ligand induced allosteric modulation of ribozyme cleavage activity. Finally, selection based on cell viability is performed through the transformation of the ribozyme regulated GAL4 library into mutant strains in which either the HIS3 (positive selection) or URA3 (negative selection) gene have been placed under GAL4 mediated transcriptional control. Figure copied from Klauser et al. [60].*

Furthermore, the consequences of oligonucleotide binding interactions lead to predictable outcomes in which the nucleotides contained within the binding site are sequestered by the ligand and thus switch to a state which is unavailable for intramolecular interactions. Since the first description of a short oligonucleotide allosteric ribozyme in 2004 a number of landmark papers have expanded upon this concept to further increase the versatility of sequence inputs, ribozyme types that can be programmed, and the mechanisms by which gene expression can be regulated [63–66] (Fig. 1.4).

### 1.4.1 An oligonucleotide regulated ribozyme switch can function within cells

Yen *et al.* [63] provided the first description of a short oligonucleotide responsive ribozyme. To achieve this, the intracellular cleavage activity of candidate hammerhead ribozymes was first optimised through a process of rational modifications and activity screening. A large number of ribozymes (and their catalytically inactive controls) were cloned into the 5' UTR of a *LacZ* reporter gene in a mammalian expression vector, transfected into HEK293T cells and assayed for their ability to inhibit the expression of β-galactosidase (LacZ). Of all candidates tested, the naturally occurring Pst-3 and Sm1 hammerhead ribozymes, derived from the *Dolichopoda* cave cricket [67] and *Schistosoma mansoni* [68] respectively, were found to be the only two with detectable cleavage activity in mammalian cells (HEK293T). In hindsight this result can most likely be attributed to the fact that the non-functional ribozymes were present in the minimal conformation, now known to lack important tertiary interactions that are required for catalytic activity under intracellular conditions of low magnesium ion concentration [69, 70]. Further modifications were made to the Sm1 ribozyme, including a single base substitution within the catalytic core (C to U at nucleotide position 7) and the extension of the loop at the end of stem III to create ribozyme variants with greater intracellular activity. Finally the position of the ribozyme within the 5' UTR was moved from immediately distal to the promoter to a location just proximal to the protein coding sequence. Together these modifications resulted in the generation of the N79 ribozyme with an increased ON:OFF ratio of approximately 225 compared to the parental Sm1 ribozyme with an ON:OFF ratio of 19 [63].

Following the demonstration that the presence of N79 ribozyme within the 5' UTR of a LacZ transcript is capable of a potent repression of protein expression, the ability to inhibit ribozyme cleavage activity by a short oligonucleotide ligand was tested [63]. Oligonucleotides complementary to regions of the hammerhead stems I and III and traversing the scissile phosphodiester bond were designed and synthesised using various modified nucleic acid back-

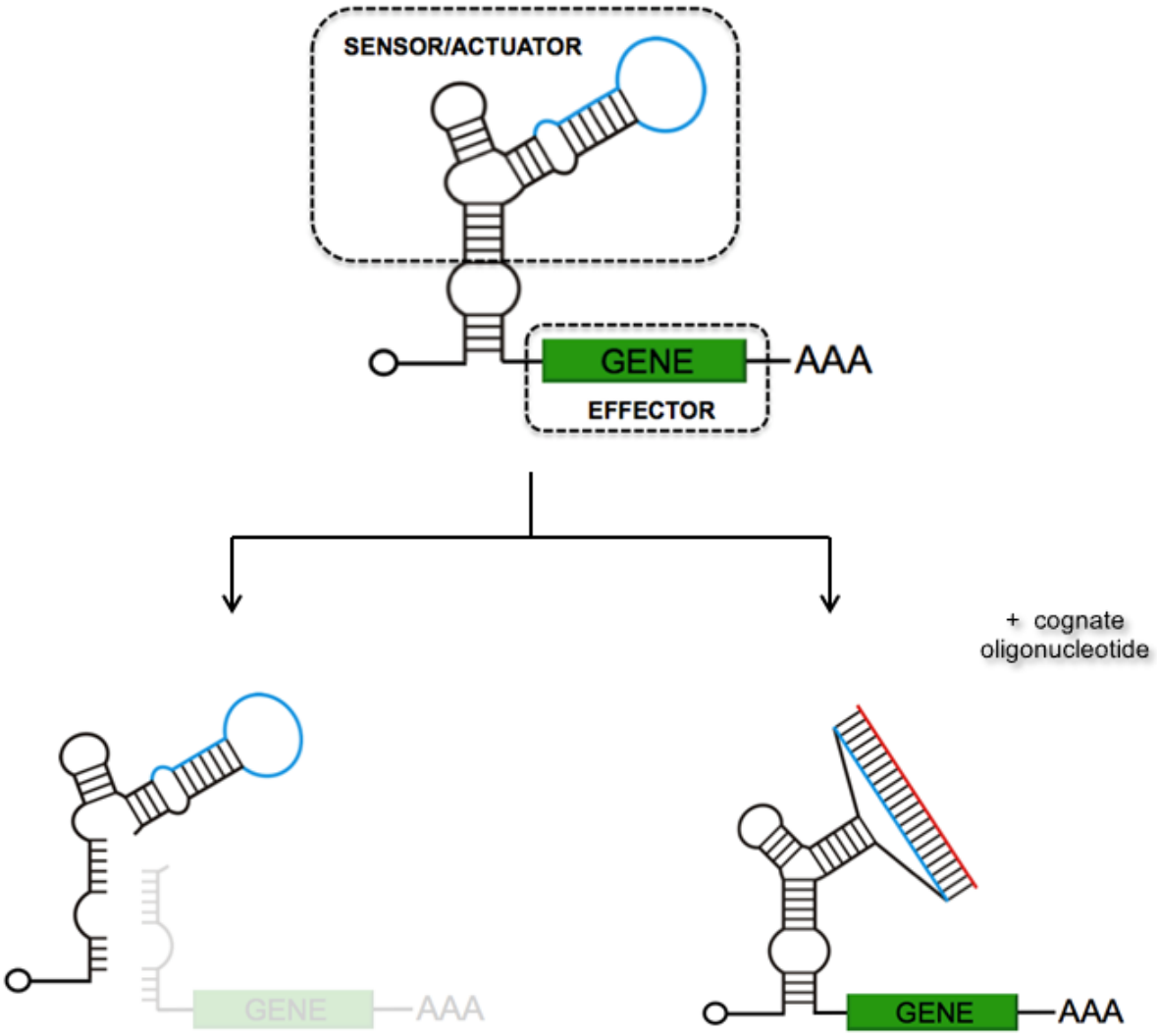



Figure 1.4: Oligonucleotide regulated allosteric ribozyme
*These allosteric ribozymes (sensor/actuator) consist of an oligonucleotide binding site (OBS) located distant from the catalytic core. In the absence of oligonucleotide ligand (left hand side) the ribozyme folds into an active conformation and undergoes autocatalytic cleavage. In the presence of a cognate oligonucleotide ligand (right hand side), the OBS becomes occupied transmitting structural changes to the catalytic core and inhibiting cleavage activity. Such ribozymes located in the non-coding regions of mRNA transcripts (5' or 3' UTRs) can be used to regulate gene expression.*

bones. HEK293T cells transfected with the N79 containing reporter gene and an antisense morpholino oligonucleotide demonstrated increased expression of the LacZ reporter by a factor of up to 2000 fold. This result was attributed to oligonucleotide-mediated inhibition of ribozyme cleavage activity. Surprisingly, this effect was much less pronounced when using oligonucleotides synthesised with phosphorothiolate and 2'-O-metyl modifications and could not be recapitulated with peptide nucleic acids or locked nucleic acids (LNA). It was unlikely this result was due to differences in the relative affinities of these modified oligonucleotides for their cognate RNA targets given that the non-functional LNA based oligonucleotides are regarded as having the highest base-pairing affinity [71]. Instead, it was proposed that the observed differences might be related to unequal transfection efficiencies and/or various levels of cellular toxicity. Despite these caveats, this study successfully introduced and demonstrated the concept of short oligonucleotide regulated ribozymes [63].

### 1.4.2 Bioinformatic tools for the generation of custom ribozyme logic gates

As opposed to the strategy used by Yen *et al.* [63] in which an oligonucleotide was designed to modulate a specific ribozyme sequence, Penchovsky and Breaker have created tools to enable the design of custom ribozymes that undergo allosteric modulation by user defined input oligonucleotide sequences [64]. Ribozymes produced with these tools contain oligonucleotide binding sites (OBSs) in either stem II or III, locations known to be amenable to engineering without loss of catalytic activity. In order to engineer allostery, a large number of ribozymes ($10^7$) containing randomised OBS sequences were computationally modelled, in the presence or absence of their respective cognate input oligonucleotide ligands, using a partition function algorithm to model the ensemble of secondary structures that may exist at equilibrium [72]. Subsequently, it was possible to identify potential ribozymes that underwent a change in dominant secondary structure, from catalytically active to inactive or *vice versa*, in the presence of their cognate oligonucleotide input/s.

Using these computational tools, Penchovsky and Breaker created ribozymes that could be used to implement the Boolean logic functions AND, OR, NOT and YES. The cleavage state of the ribozyme (uncleaved or cleaved) represented the binary on/off output of these molecular logic gates. The state in which a ribozyme existed was determined as a function of the supplied oligonucleotide input signals. *In vitro* ribozyme cleavage assays were used to validate the function of these bioinformatically designed logic gates. Finally, it was shown that the output of one ribozyme logic gate could serve as the input to a second ribozyme logic gate thus creating a connected circuit.

The successful implementation of this strategy highlighted the power of computational structural prediction algorithms for the rational design of functional RNA species. This work has been particularly instrumental given that it was the first demonstration of true allosteric modulation of ribozyme cleavage activity by a short oligonucleotide. The mechanism by which N79 cleavage activity was regulated more closely resembled competitive inhibition, mediated through oligonucleotide binding to the ribozyme catalytic site. The ribozyme architecture described by Penchovsky and Breaker was distinct in that it separated the OBS from the catalytic core of the ribozyme. In this case, the state of the OBS occupancy elicits structural alterations in the ribozyme that are transmitted to the distant catalytic site to allosterically regulate its cleavage activity. This strategy was essential to enable unrestricted modification of OBS sequence to accommodate different input ligands whilst maintaining the integrity of the conserved sequences surrounding the catalytic core.

### 1.4.3 A rational design approach for the construction of oligonucleotide regulated allosteric ribozymes

Klauser and Hartig provide the first description of a short oligonucleotide responsive ribozyme used to regulate translation in prokaryotic cells thus further expanding the functional repertoire of this class of synthetic regulatory devices [66]. Relying on the established mechanism of allosteric regulation of hammerhead ribozymes, Klauser and Hartig engineered an OBS into stem III which would transmit structural changes to the catalytic core resulting in inhibition of cleavage activity, following binding of a cognate oligonucleotide ligand. In order to convert this to a detectable output signal, ribozyme cleavage was coupled to the exposure of a reporter transcript RBS, a mechanism already exploited for the ribozyme regulated control of reporter protein expression [73]. This was achieved through the construction of a 5' UTR-encoded hammerhead ribozyme, in which stem I has been extended to incorporate an RBS sequestered by intramolecular interactions. Under conditions favouring active ribozyme catalysis, self-cleavage results in the separation of the two halves of the hammerhead stem I, thus exposing the RBS and permitting translation. Conversely, translation of protein would be inhibited under conditions of low ribozyme cleavage activity, following allosteric inhibition by a cognate short oligonucleotide ligand. Based on this mechanism of regulation, Klauser and Hartig achieved a 10 fold inhibition of reporter expression upon transformation of bacteria with a ribozyme regulated GFP transcript in the presence of its cognate oligonucleotide ligand.

A further innovative adaptation introduced by Klauser and Hartig is the use of the naturally occurring YUNR RNA structural motif to enhance the OBS accessibility to its cognate

oligonucleotide ligand [74]. The YUNR motif, engineered to lie at the apical loop of stem III, results in the adoption of a structural conformation in which the bases are displayed along the outer perimeter of the loop. This conformation provides the opportunity to initiate a kissing interaction between the ribozyme OBS and complementary oligonucleotide ligands, upon which binding can propagate along the full length ribozyme sequence.

## 1.5 miRNAs

microRNAs (miRNAs) are short non-coding RNA molecules (~22nt long) found within almost all metazoan genomes, from worms to mammals. miRNAs have the capacity to bind to and tune the expression of target mRNAs, bearing sequence complementarity. They function as molecular rheostats exerting control over many essential gene expression networks during development, normal physiology and importantly, human disease [11]. The pattern of miRNA expression may provide an important diagnostic tool and thus the key aspect of this project was to seek to create synthetic circuits capable of integrating input information from specific miRNAs, in order to regulate the expression of various user-defined cellular responses.

### 1.5.1 miRNA biogenesis

The biogenesis of a mature miRNA is a multi-step process involving transcription, processing and cleavage by two RNase III enzymes and finally loading into the multi-protein miRNA induced silencing complex (miRISC) [75]. This process is conserved throughout metazoans and performed by homologous enzymes; here we have focussed specifically on the mammalian processing pathway (Fig. 1.5A). The first step in this process is the transcription, by RNA polymerase II, of a primary miRNA (pri-miRNA) transcript containing a stable stem-loop region encoding the mature miRNA sequence. In mammalian cells, pri-miRNAs can be encoded in intergenic regions or within the introns or exons of both coding and noncoding genes [76]. Multiple miRNA genes can be clustered in close proximity, resulting in their inclusion in polycistronic pri-miRNA transcripts upon which individual miRNAs are released following subsequent processing of constituent stem-loop structures [75].

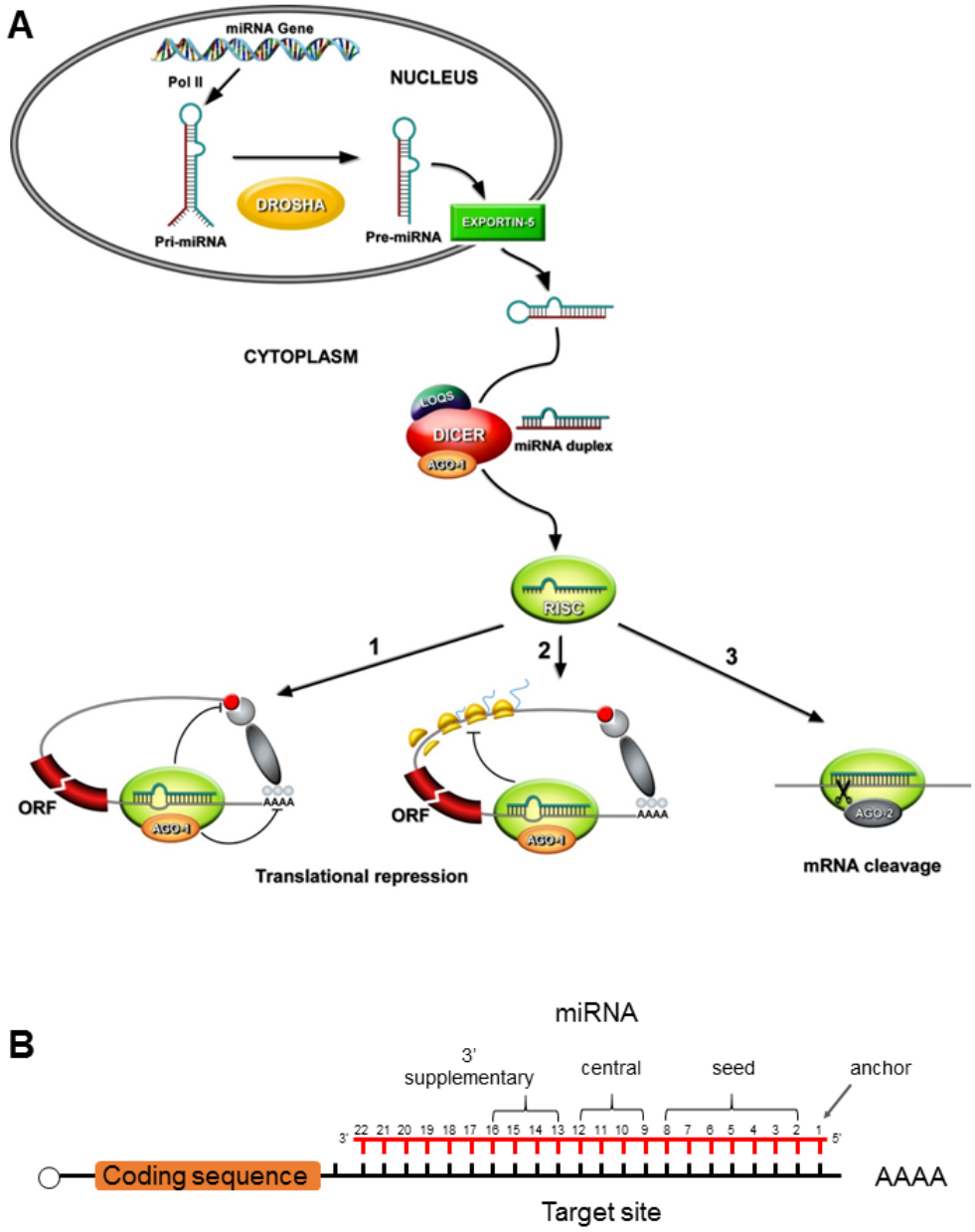



Figure 1.5: miRNA biogenesis and mechanisms of repressive action
*(A) Primary miRNA (pri-miRNA) are transcribed by RNA polymerase II before nuclear processing by Drosha to release a precursor-miRNA (pre-miRNA) hairpin. The pre-miRNA is transported to the cytoplasm via exportin-5 where is subject to a final processing event by Dicer to generate the mature miRNA duplex. The guide strand is loaded into an Argonaute protein within the RNAi induced silencing complex (RISC) which it directs to complementary binding sites within target mRNA transcripts. A miRNA bound transcript may undergo repression via, (1) enhanced deadenylation, decapping and mRNA destabilisation, (2) translational repression, or (3) Ago2 catalysed cleavage (which is dependent on perfect complementarity between miRNA and target). Continued...*

The first processing step occurs in the nucleus and involves the cleavage of the miRNA containing stem-loop from its pri-miRNA transcript. This reaction is catalysed by a heterotrimeric complex comprised of the RNase III enzyme Drosha and two copies of its cofactor DGCR8. DGCR8 interacts with the pri-miRNA so as to precisely position the two RNAse domains of Drosha ~11 nt distal from the junction between single and double stranded RNA of the hairpin [77]. Drosha catalyses an asymmetric double stranded cleavage reaction to generate a hairpin precursor miRNA (pre-miRNA) with a two nucleotide long 3' overhang [78–81], as is common to the cleavage products of RNAse III class enzymes [82]. The resulting ~65 nt pre-miRNA hairpin is bound by a transport complex made up of Exportin-5 and a GTP bound Ran protein that specifically recognises the short 3' overhang associated with dsRNA stems greater than 14 bp [83, 84]. The pre-miRNA containing complex is then shuttled through the nuclear membrane pore, with subsequent GTP hydrolysis resulting in complex disassembly and cytoplasmic deposition of free pre-miRNA.

A second cleavage event occurs in the cytoplasm, catalysed by the RNase III enzyme Dicer, which removes the terminal loop and generates a short dsRNA duplex containing the mature ~22 nt miRNA paired to a partially complementary passenger strand [85–87]. Dicer contains binding pockets which accommodate and anchor both the two nucleotide 3' end overhang and the phosphorylated 5' end of the pre-miRNA hairpin. This process enables precise cleavage of the pre-miRNA hairpin ~22 nt from the 5' termini [88]. Finally, the miRISC is assembled through the loading of the mature miRNA into one of the four mammalian Agonaute (Ago) proteins. Crystal structures have revealed that human Ago2 exists in a conformation comprised of an N-terminal lobe containing the PAZ domain and a C-terminal lobe containing the MID and PIWI domains [89]. The 5' phosphate group of the miRNA is tightly anchored between the MID and PIWI domains whilst the 3' terminus is bound by the PAZ domain, positioning the miRNA along a basic groove located in the MID and PIWI lobe. The PIWI domain demonstrates similarity to that of RNase H, however only human Ago2 retains the conserved catalytic triad of amino acids that are required for endonucleolytic cleavage of targets [90, 91]. The process of miRISC assembly with the mature miRNA strand as opposed to its passenger strand demonstrates a remarkable degree of specificity mediated by preferential loading of the strand demonstrating 5' end duplex instability [92, 93]. The fully assembled RISC is now

*Figure 1.6 continued: (B) The contribution of individual base positions within a miRNA to target binding and repression differs across its length and has resulted in the identification of several functionally distinct regions. Binding affinity is determined largely by seed sequence complementarity to the target. Pairing of the 3' supplementary sequences can increase this affinity. Pairing within the central region does not contribute to binding affinity but is essential to mediate Ago2 catalysed cleavage of targets.*

competent to seek out target mRNAs, guided by the loaded mature miRNA.

### 1.5.2 miRNA biochemistry

Biochemical investigations into the properties of miRNAs have contributed to our understanding of both their target binding affinity and their rate and propensity to cleave bound targets. The direct interaction between miRNA and target is mediated by Watson-Crick base pairing. The affinity of this interaction is however influenced by the protein components of the RISC which are also responsible for the catalysis of target cleavage.

The determinants of miRNA binding affinity are not equal along its length but appear to be concentrated within the first eight nucleotides at the 5' end of a miRNA known as the seed region [94]. A finer dissection of the biochemical properties of a selected miRNA within an Ago2 containing RISC resulted in the identification of more discrete functional elements comprising the anchor, seed, central, and 3' supplementary regions in a 5' to 3' direction, each contributing uniquely to the overall target affinity [95] (Fig. 1.5B). The anchor region consists solely of the first nucleotide at the 5' end of the miRNA and is characterised by having no effect on the binding affinity to target mRNAs. This is consistent with structural data revealing the tight association of the 5' phosphate to the Ago2 protein preventing the formation of a Watson-Crick base pair interaction with the corresponding target site position, regardless of its identity. The importance of nucleotides 2-8, comprising the seed sequence, towards determining the binding affinity may be inferred from structural data in which they are observed to be displayed and stacked in an accessible conformation, upon which the interaction may nucleate [96, 97]. Wee *et al.* observed no statistically significant difference in binding affinities between a miRNA and either a perfectly complementary or seed only complementary target site, further supporting the predominant importance of the seed region [95].

Once bound, several factors determine whether a miRISC is capable of catalysing target mRNA cleavage. Of the four mammalian Ago proteins (Ago1-4) which may bind a miRNA within the RISC only Ago2 has functional RNAse activity. This is a result of Ago2 having maintained the conserved catalytic DDH amino acid triad within its ribonuclease H domain, that is required for target RNA cleavage [98]. Furthermore, the rate at which an Ago2 miRISC may cleave its targets is influenced by the Watson-Crick base pair complementarity between miRNA and target. Wee *et al.* investigated this process through the *in vitro* measurement of the $k_{cat}$ for miRNA mediated cleavage of a panel of dinucleotide mismatched target sites [95]. This study revealed a dramatic decrease in $k_{cat}$ for miRNA target sites containing central mismatches (dinucleotide mismatches between positions 8-12). This would be consistent with a disruption near to the active site as revealed by the modelled structure in which the scissile bond lies between positions 11-12 within the active RNAse H domain of the Ago2 PIWI

domain [90]. Information regarding the *in vivo* cleavage activity of the Ago2 miRISC was gleaned from experiments performed in HeLa cell cytoplasmic extracts [99]. Here cleavage activity was only demonstrable against target sites bearing perfect Watson-Crick base pair complementarity to their cognate miRNA.

### 1.5.3 miRNA mediated repression

miRNAs can direct the repression of their target mRNAs through a number of distinct mechanisms, including cleavage, mRNA destabilisation and decay or translational inhibition. A number of factors are known to influence the choice of repressive mechanisms affecting a given target. A clear distinction is made between target sites that display perfect complementarity to a miRNA, and are thus susceptible to cleavage, and imperfect target sites which are repressed via other mechanisms. In animals, the majority of endogenous mRNA/miRNA target pairs are defined by imperfect complementarity and are thus the targets of mRNA destabilisation and translational inhibition [100]. The relative contribution of these two repressive mechanisms has been the subject of debate, resulting in the generation of sometimes contradictory data in support of each particular silencing pathway. Recent high-throughput analysis of paired proteomic and transcriptomic data sets has suggested that the majority of miRNA mediated repression can be explained by a cleavage independent mRNA destabilisation mechanism [101]. Finally, two papers have shown a time dependent change in the miRNA mediated inhibitory mechanism from initial translational repression to later mRNA destabilisation [102, 103].

#### 1.5.3.1 Translational inhibition

Translational repression as a mechanism for miRNA induced target silencing was proposed following early findings that expression of targets was inhibited at the protein level without a reduction in abundance of the encoding mRNA transcript [104–107]. Mechanistically, this could be mediated by inhibition of any one or more of the numerous steps involved in the process of translation, from initiation through to elongation and finally termination. A number of studies have attempted to address this question through the utilisation of techniques that explore different aspects the translation process.

Polysome profiling involves the separation of ribosome/mRNA complexes using sucrose gradient sedimentation in order to quantify and resolve mRNA species into an unbound fraction or fractions associated with increasing numbers of ribosomes [108]. This assay can provide a readout for translational competency of a particular transcript, which when actively translated may be found in the heavier polysome fractions. Conversely, under conditions in

which translational initiation may be inhibited, the majority of mRNAs may shift to a lighter fraction associated with fewer ribosomes. The technique of polysome profiling was therefore used to investigate the translational state of miRNA targeted transcripts. This showed that certain miRNA-repressed target mRNAs were associated with polysomes [109, 110], and therefore actively translated, however other targets were found to shift to lighter fractions suggestive of an initiation blockade [111]. Petersen *et al.* proposed the premature termination of translating ribosomes as a mechanism of post-initiation inhibition, based on their observation that polysomes bound to a miRNA-repressed transcript were faster to dissociate following treatment with an inhibitor of translation initiation [110]. Given the contradictory data generated by this work alternative investigations were required to provide clarity on the translational inhibitory mechanisms that may be employed by miRNAs.

Internal ribosome entry sites (IRESs) are used by certain viruses to enable the non-canonical initiation of translation from viral mRNA transcripts by substituting to varying extents for components of the ordered canonical process [112]. As a result they have been exploited for the investigation of miRNA dependent inhibition of translational initiation. The canonical translation initiation process begins with the recognition of the 5' cap by eIF4E, a component of the multi-protein eIF4F complex also containing the RNA helicase eIF4A and the scaffold protein eIF4G. Circularisation of the mRNA is promoted through interactions between eIF4G and the PABPC which coats the poly(A) tail at the 3' terminus of transcripts. This structured complex is required for the recruitment of the 43S pre-initiation complex, comprised of the 40S ribosomal subunit, eIF2 ternary complex, eIF5 and eIF3. Subsequently, the 43S complex scans the 5' UTR in a 5' to 3' direction until it encounters a start codon upon which translation is initiated. Non-canonical initiation may occur from an IRES which is able to bypass certain canonical initiation factors and promote translation through a more direct pathway. Naturally occurring IRESs have been classified depending on the canonical factors that they bypass and the mechanism by which they promote translational initiation. Type 1 and 2 IRESs are capable of directly binding to an eIF4G and eIF4A complex, therefore bypassing the requirement for 5' cap binding by eIF4E. Type 3 IRESs are capable of directly binding to and recruiting the 43S pre-initiation complex. Finally, type 4 IRESs require no initiation factors but rather directly recruits the 40S ribosomal subunit, circumventing the need for the initiator tRNA through an IRES domain that structurally mimics the codon-anticodon interaction. The extent to which proteins expressed from these functionally gradated IRES-dependent transcripts are subject to miRNA-mediated inhibition can be used to dissect, at high resolution, the particular silencing method employed. To this end, Meijer *et el.* [113] generated IRES bearing reporter constructs to show that type 2 (Encephalomyocarditis Virus [EMCV] IRES) but not type 3 (Hepatitis C Virus [HCV] IRES) or type 4 (Cricket Paralysis Virus [CrPV] IRES) IRESs were

susceptible to miRNA mediated repression. This finding pinpointed the stage at which miRNA mediated repression occurs and implicated either eIF4G or eIF4A as the mechanistic target. This was followed by followed by knockdown of both potential factors and the revelation that eIF4A2 was essential for miRNA mediated inhibition of translational initiation. This is however contradictory to the results of Pillai *et al.* [111] in which translation driven by the EMCV IRES was immune to miRNA-mediated repression thus suggesting an involvement of eIF4E and cap recognition. Furthermore, Petersen *et al.* demonstrated that even translation from a CrPV IRES is susceptible to inhibition, arguing against a mechanism involving translational initiation [110].

Together, this evidence shows that miRNAs are capable of inhibiting translation of targets by a number of mechanisms, including inhibiting translational inhibition (via eIF4A2) and the promotion of premature termination of translation. The multiple inhibitory mechanisms identified in different studies may represent true biological diversity in the mode of miRNA action in different at different target types. Alternatively, it is possible that miRNA mediated repression proceeds via an ordered process and the different experimental techniques used above have probed this pathway at differing points along the spectrum of miRNA mediated repression.

#### 1.5.3.2 mRNA destabilisation

mRNA destabilisation is an alternative mechanism by which miRNAs have been suggested to inhibit expression of their targets. This involves an enhancement in the rate at which targeted mRNAs undergo degradation by the general cellular machinery promoting decapping, deadenylation and exonucleolytic decay. This mechanism is supported by several studies showing the miRNA dependent degradation of target transcripts [114, 115], and the finding that the abundance of predicted miRNA targets increases in a genetic background where components of the miRNA pathway have been knocked out [115–117]. More comprehensive evidence in support of mRNA destabilisation arose from the Bartel laboratory when they used parallel RNA sequencing and high throughput ribosome profiling to simultaneously investigate the genome wide effects of a miRNA on the abundance and translation status of target transcripts [101]. The ribosome profiling employed by the Bartel laboratory involved the RNAse I digestion of cellular mRNAs (with bound ribosomes) to generate ribosome protected fragments, their purification on a sucrose gradient, and finally sequencing. These two parallel profiling approaches were performed in HeLa cells transfected with miR-155 (not normally expressed in HeLa cells). These conditions were chosen to enable comparison to available proteomic datasets generated using pulse labelled stable isotope labelling with amino acids in cell culture (pSILAC) to measure changes in protein abundance in response to miR-155 in HeLa

cells [118]. It was found that mRNAs containing at least one 3' UTR miR-155 target site were underrepresented in sequencing libraries from both RNA sequencing and high throughput ribosome profiling. The ribosome profiling data were normalised to mRNA abundance to study translational repression without the confounding influence of mRNA destabilisation. This revealed a minor degree of translational repression at miR-155 targets. The majority of repression at the protein level however could be accounted for by mRNA destabilisation.

Structural studies have shed more light on the molecular mechanisms underpinning the process of miRNA mediated mRNA destabilisation. Biochemical studies have identified a direct interaction between Argonaute proteins and GW182 [119], a scaffold protein involved in the recruitment of the protein mediators of decay. The CCR4-NOT deadenylation complex is recruited through a direct interaction with effector domains located in both the N-terminal and C-terminal domains of GW182 [120]. Additionally, a second deadenylation complex PAN2-PAN-3 has been shown to directly interact with the GW182 scaffold [121]. miRNA mediated destabilisation and general mRNA decay appear to converge at this point of deadenylation and proceed via a common pathway involving removal of the 5' cap by decapping protein 2 (DCP2) and 5' to 3' degradation by exoribonuclease 1 (XRN1) [122].

#### 1.5.3.3 miRNA repression is mediated by both translational repression and mRNA destabilisation in a time dependent manner

The high throughput data from the Bartel laboratory appears to firmly suggest that target mRNA destabilisation is the dominant mechanism of miRNA mediated repression at steady state [101]. This however fails to explain the strong evidence demonstrating miRNA mediated translational inhibition [104–107]. This problem is compounded by the close interrelation of these two processes in which it is difficult to resolve cause and effect. For example, experiments have shown that inhibition of translational initiation via mutation of initiation factors can promote the mRNA decay via decapping and deadenylation [123]. To further disentangle the contribution of these two mechanisms to the mode of miRNA action further investigation was required.

To address these questions the Giraldez and Green laboratories have independently conducted experiments to investigate the mechanism of miRNA mediated repression as a function of time. The Giraldez laboratory used a Zebrafish embryo developmental model in which the expression of miR-430 is turned on during the maternal to zygotic transition (MZT) between 2-4 hours post fertilisation [103]. Ribosome profiling was used to investigate the translation status of miR-430 targets before and at two time points (4 and 6 hours post fertilisation) after the MZT in wild type embryos and *dicer* mutants (that do not express any miRNAs). At the first time point (4 hours) a reduction in ribosomal occupancy on miR-430 targets was detected

before significant deadenylation or destabilisation. In contrast, by 6 hours post fertilisation mRNA destabilisation became the dominant detected mechanism of repression. This evidence would suggest either a temporal switch in the mechanism of miRNA mediated repression from translational repression to mRNA destabilisation or an ordered process in which mRNA destabilisation is dependent on preceding translational repression. The Green laboratory came to similar conclusions through the study of the kinetics of miRNA repressive mechanisms in an inducible Drosophila melanogaster cell culture (S2 cells) system [102]. Inducible luciferase reporter constructs containing 3' UTR miRNA target sites were transfected into S2 cells, their expression was induced over a 90 minute pulse, and the levels of reporter mRNA and luciferase luminescence was measured over a 48 hour period. This experiment confirmed the initial action of miRNAs to repress the translation of their targets followed by a switch to target mRNA destabilisation [102].

### 1.5.4 miRNAs as potential biomarkers

miRNAs have been shown to play essential roles in a wide variety of biological processes in both health and disease. Due to the lax requirement for base pair complementarity between miRNA and target (as few as six complementary base pairs within the seed region may be sufficient for a functional interaction [95]), each miRNA is predicted to target and modulate the expression of many mRNA molecules. Likewise each mRNA is predicted to be targeted by many different miRNAs. Consequently, miRNAs and mRNAs are involved in complex, intertwined and finely balanced pathways governing the fate of each cell. Even subtle deviations from the normal expression profiles of such molecules, may be enough to alter cellular behaviour and promote the onset of disease. Many miRNAs have been identified with a causative role in diseases such as viral infections, neurodegeneration and cancer [124, 125]. In addition, the widespread cellular changes underlying disease pathogenesis may themselves disrupt the normal miRNA expression profile of a cell. For these reasons, miRNAs have emerged as attractive biomarkers for pathological conditions, with implications in prognosis, diagnosis and even therapeutic intervention.

Infection with pathogens such as bacteria and viruses can often be identified by the presence of exogenous material in the form of DNA, RNA or proteins. These same molecules also serve as potential targets for therepeutic intervention. Certain viruses however have evolved ingenious mechanisms to avoid the immune response which often also hinder therapeutic targeting. One such example is provided by members of the Herpesviridae family of DNA viruses that can successfully persist in a host for a lifetime in a latent phase [126]. This phase is characterised by a highly restricted expression of viral proteins insufficient to invoke an immune response and also markedly limiting the range of potential therapeutic targets. Although often

harmless, latency may lead to viral reactivation and disease. Additionally, viral latency itself has also been linked with certain pathogenic conditions. For example, latent Epstein-Barr virus (EBV) can result in transformation of host B cells, a phenomenon that has been linked to the development of several types of lymphoma [127]. Surprisingly viral miRNAs, which are non-immunogenic, have been found to be highly expressed during latent infection and therefore represent novel therapeutic targets [128]. The EBV miR-BART2 plays a direct role in maintenance of latency through inhibition of expression of BALF5, an EBV DNA polymerase [129]. Sixteen other EBV miRNAs have also been found to be expressed in latently infected B cells. The presence of a viral miRNA can serve as a digital biomarker since its presence or absence indicates either an infected or uninfected cell respectively. Therefore a viral miRNA can serve as a clean signal with which to identify cells carrying the latent EBV virus.

Cancer is another human disease that has been heavily associated with miRNAs in both causative and epiphenomenal roles. Genes encoding miRNAs were found to be enriched within regions of the genome amplified, deleted or translocated in various forms of cancer [130]. Evidence has now accumulated suggesting that certain miRNAs, classed as oncogenic miRNAs (oncomiRs) and tumour suppressor miRNAs, can directly contribute to tumourigenesis through their overexpression or silencing respectively. The miR-19-92 cluster is often overexpressed in B cell lymphomas and can cooperate with the oncogene *myc* to promote tumour formation [131]. Additionally other miRNAs have been associated with roles in promoting hallmark features of cancer such as metastasis (miR-10b [132]) and angiogenesis (miR-210 [133]). Regardless of their role, miRNAs can be accurately used as biomarkers for cancer and to differentiate various cancer subtypes from each other. For this purpose, miRNAs have proved to have an even greater predictive power than mRNAs [134]. In contrast to the situation with viral miRNAs, the subtle alterations in the levels of endogenous miRNAs associated with cancer constitute an analogue signal input. Therefore, in order to overcome the intrinsic noise, multiple input signals must be integrated. Computational analysis of data from the MicroRNA Atlas can be used to select a set of highly and lowly expressed miRNAs from a given cell line that sufficiently differentiate it from other similar lines. The resolving power increases as the set of classifier miRNAs increase. Five classifier miRNAs are sufficient to differentiate the HeLa cell line from the closest other cell line with a predicted six fold difference [135]. A similar process could be used to define the minimum set of miRNAs needed to classify any type of endogenous cancer.

### 1.5.5 Specific miRNA biomarkers used in this thesis

We have chosen to use the miRNAs miR-122 and miR-21 as input signals to the sensor circuits developed in this thesis. These miRNAs have been chosen based on their involvement in

different disease processes and their potential as informative biomarkers and druggable targets.

miR-122 is an important miRNA, conserved amongst vertebrate species and highly expressed in the liver (where it may account for up to 72 % of expressed miRNAs [136]). miR-122 has been strongly linked to the replication of HCV through a direct interaction with the viral genome [137]. Contrary to its expected action, miR-122 binding to the HCV RNA genome appears to enhance translation [138]. Furthermore, the stability of the HCV genome appears to be increased when bound by miR-122 [139]. miR-122 has become an attractive drug target given its importance in the regulation of HCV resulting in the generation of antisense inhibitory oligonucleotides [140] that have been shown to be efficacious in clinical trials [141].

miR-21 is an onco-miR found to be overexpressed in many types of solid tumour [142] as well as B cell lymphoma [143, 144]. The oncogenic function of miR-21 is mediated through its inhibition of target phosphatases involved in signal transduction through the AKT and MAPK pathways [144]. The role of miR-21 in oncogenesis was explored in a mouse model system for the conditional overexpression of miR-21 in hematopoietic cells resulting in the development of pre-B-cell malignancy [143]. This phenotype was reversible upon the inhibition of miR-21 leading to a proliferative block and increased levels of apoptosis and further supporting miR-21 as the causative agent in driving the malignant transformation.

Given the importance of both miR-122 and miR-21, the development of sensor circuits responsive to these miRNAs could serve as useful tools in basic scientific research. Further ahead it can be envisaged that synthetic gene expression systems could be created in order to regulate the expression of therapeutically relevant proteins in response to the detection of these disease associated miRNAs.

## 1.6 Sensors for the detection of endogenous miRNA expression

In order to further our understanding of the biological function of miRNAs a number of systems have been devised to enable the detection of their activity within cells and even whole animals. These systems have since been adapted as components of synthetic gene regulatory circuits. miRNA sensor circuits can be separated into two major groups; those that regulate output expression in an anticorrelated manner (negative sensors), or those that regulate output expression in a correlated manner (positive sensors) to input miRNA activity (Fig. 1.6).

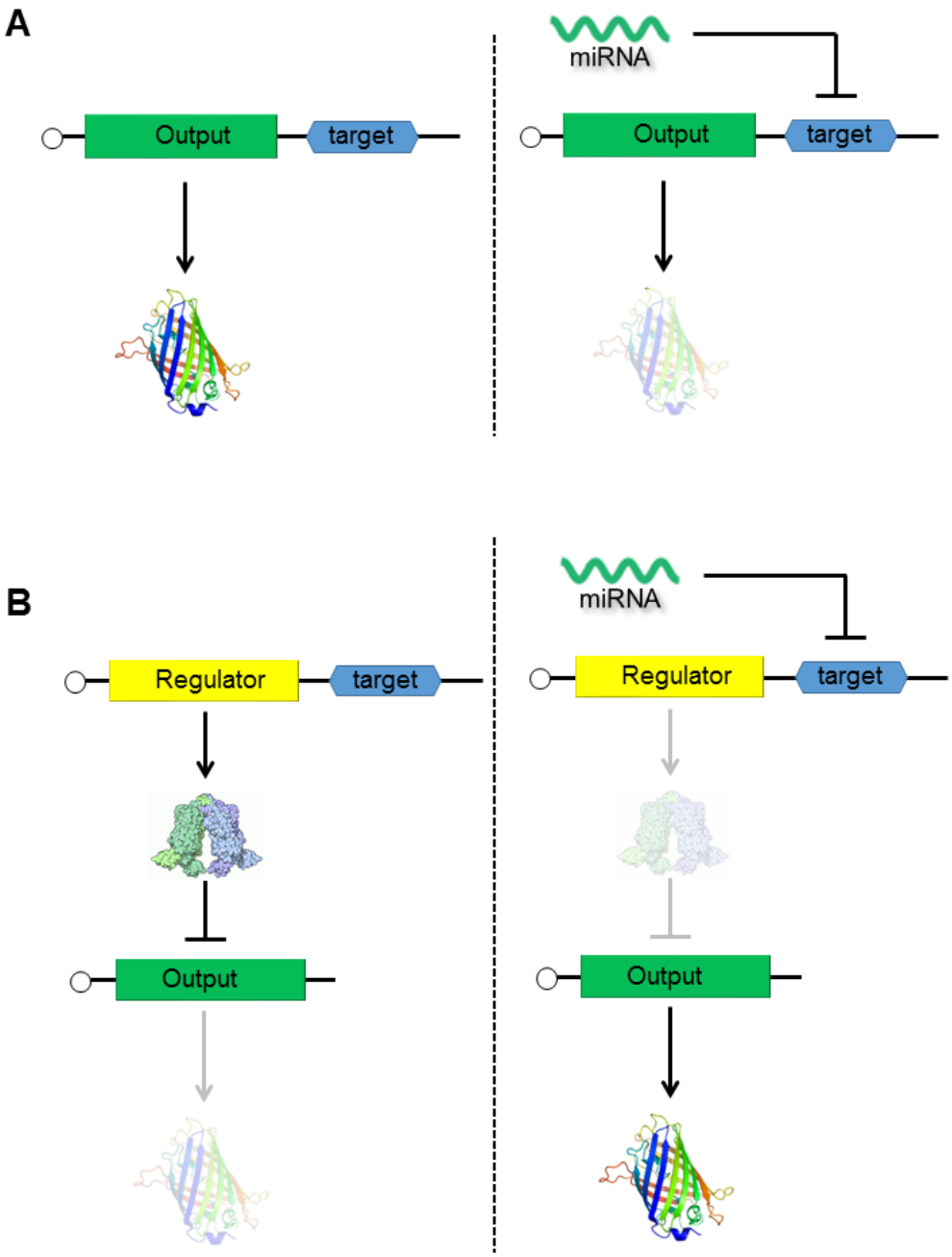



Figure 1.6: Schematic representation of both negative and positive miRNA sensor circuits *(A) In the absence of miRNA (left hand side) the output of a negative sensor is uninhibited and leads to high levels of protein expression. In the presence of miRNA (right hand side) the output transcript is subject to miRNA mediated inhibition and thus the silencing of output protein expression. (B) A positive sensor utilises a more complex circuit architecture to produce the opposit effect to a negative sensor. In this case, in the absence of miRNA (left hand side) expression of a regulator (which functions to repress the output) is unopposed and thus functions to efficiently silence the output. In the presence of miRNA (right hand side) the regulator is subject to repression thus derepressing the output and permitting its expression.*

### 1.6.1 Negative miRNA sensors

The endogenous role of miRNAs as modulators of gene expression lends themselves nicely to the incorporation into synthetic genetic circuits. This is particularly facilitated by the well defined rules dictating target site properties and design [95]. The first and most straightforward implementation of a miRNA controlled synthetic circuit involved the engineering of a system in which a miRNA target site is inserted into the 3' UTR of a reporter gene. Consequently, the expression of the reporter protein will be susceptible to miRNA mediated repression, via one of the mechanisms discussed above. As such, an inverse relationship will be displayed between the levels of reporter protein and cognate miRNA expression, constituting a negative miRNA sensor circuit. Based on knowledge regarding the mechanisms of miRNA repression and the susceptibility of modified target sites to repression it is possible to tune the relationship between miRNA and reporter. A target site that displays full length, perfect, Watson-Crick base pair complementarity to its cognate miRNA will undergo repression via cleavage through the RNAi pathway [99]. Given the multi-turnover cleavage of target mRNA that is catalysed by the cognate miRNA-RISC, circuits of this nature can offer a high sensitivity to their input signal and therefore undergo switching over a narrow dynamic range. This range of cognate miRNA required to switch from an OFF to ON state can be tuned through rational modifications to the target site. Tuning can be achieved through changing the degree of complementarity that a target bears to its cognate miRNA, which will alter both the binding affinity ($k_m$) and the rate of cleavage ($k_{cat}$) [95].

Negative sensor circuits of this nature were first described in 2002 with the insertion of miRNA target sites into the 3' UTR of a luciferase reporter in an *in vitro* based system [99] and a chloramphenicol acetyltransferase reporter in mammalian 293T cells [145]. Subsequently such sensors were used to create transgenic animals (Drosophila and mouse) to assess the expression of miRNAs within whole organisms [105, 146]. More recently, this concept has been adopted for use in high throughput screening methodologies [147]. This was achieved through the creation of lentiviral vectors, each encoding an EGFP gene with a unique 3' UTR target site for 1 of 291 conserved miRNAs. These vectors were pooled and used to create lentivirus for transduction into various human cell lines. Transduced cells were stratified based on EGFP fluorescence by fluorescence assisted cell sorting (FACS) and subject to targeted sequencing of the integrated viral genome. miRNA activity was inferred from the abundance its corresponding sensor (inverse correlation) in each of the EGFP strata.

### 1.6.2 Positive miRNA sensors

Positive miRNA sensors have recently been created in order to expand upon the versatility of downstream functions that can be governed by miRNA regulated synthetic circuits [135, 148, 149]. Circuits of this nature require a more complex multi-component architecture such that expression of their output positively correlates with cognate miRNA levels. This is achieved through targeting the input miRNA to an intermediate repressor module (regulator) which itself acts to inhibit expression of the output signal. Specifically, miRNA target sites were inserted into the 3' UTR of a transcriptional repressor gene (*LacI*) that inhibited transcription of the output reporter gene [135]. In effect, this circuit links together two NOT gates and is thus referred to as a double negative gate circuit.

Whilst such circuits do enable the positive correlation of an output signal with input miRNA expression, they are associated with a number of caveats mainly due to the multitude of biological processes that separate input from output (Fig. 1.6). The first of these results from the fact that the miRNA represses the regulator mRNA transcript, whilst it is the regulator protein which actively inhibits the output. Thus the circuit kinetics are dependent on the stability of the regulator protein, which may persist long after its mRNA transcript has been degraded. Further lag may be introduced even after the regulator protein has been removed, between the transcription of the output gene and its eventual translation. Finally, the delivery of such multicomponent systems into living cells may be problematic given limitations in the size of genetic material that can be carried by some vectors (particularly viral vectors).

## 1.7 Thesis framework

This thesis aims to expand upon existing genetic circuitry for the detection of miRNAs in a more direct manner. Focusing on positive miRNA sensors, we have attempted to condense existing circuit architecture into simplified systems. Our two chosen systems are schematically represented in Figure 1.7. We have utilised RNA as a substrate to create circuit components wherever possible based its potential to function in diverse ways and its amenability to engineering. The use of RNA is further supported by the ease with which it can be engineered to interact with our cognate miRNA input signals. Given the importance of miRNAs in a multitude of processes both in health and disease, such circuits may eventually constitute powerful tools for biological research or the regulation of gene expression in a therapeutic context.

The first objective of this work was the design of a double negative gate circuit based on work published by the Benenson laboratory [135]. The key difference in our circuit being the use of a translational repressor (L7Ae, [12, 13]) instead of a transcriptional repressor (LacI) as the regulator module. The rationale behind this revision is exemplified in Figure

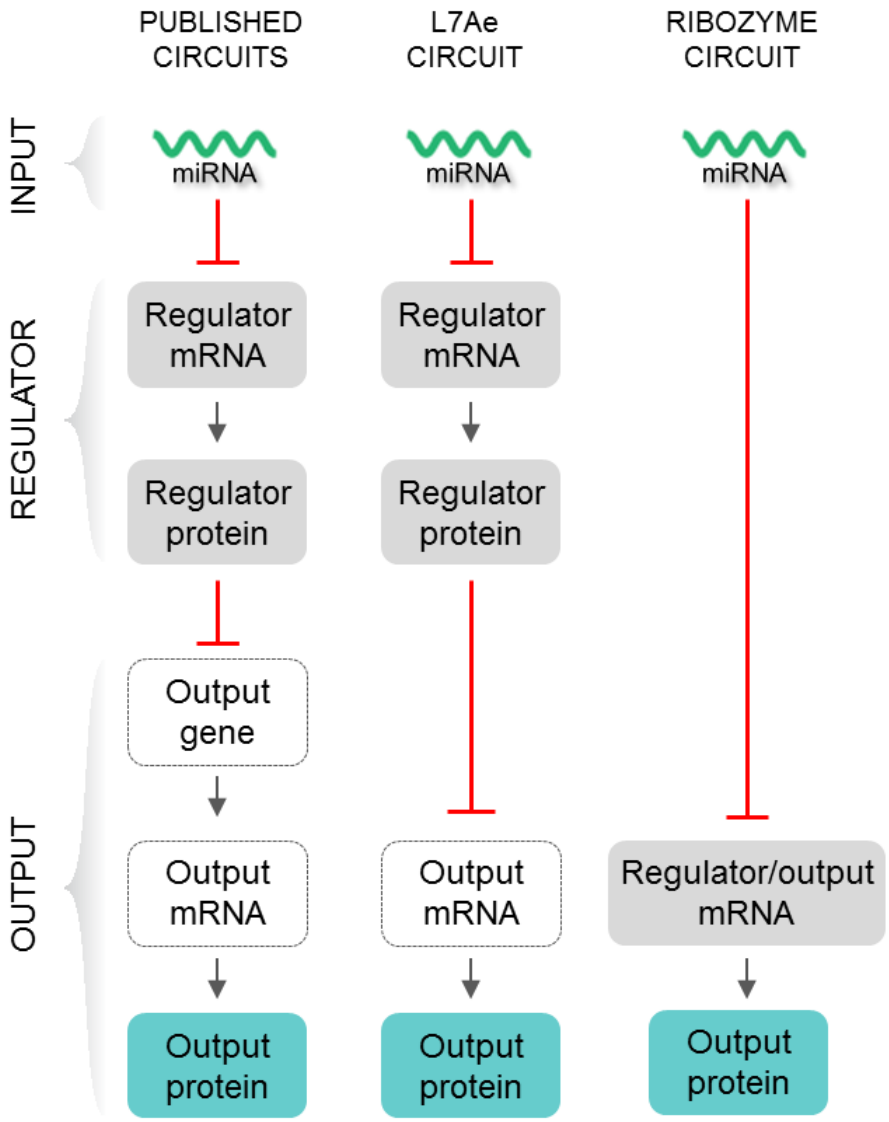



Figure 1.7: Circuit architectures capable of positive regulation of output in response to miRNA input (double negative gate circuits)
*All three circuits function via two linked NOT logic gates; input (miRNA) mediated inhibition of the regulator and regulator mediated inhibition of the output. The published circuits [135] utilise the most complex pathway including the transcriptional regulation of output gene expression. Our proposed L7Ae circuit (Chapter 3) exploits a more compact architecture and functions at an entirely post-transcriptional level. Our proposed ribozyme circuit (Chapters 3 and 4) employ the most compact architecture through the fusion of regulator and output functionality into a single RNA molecule.*

1.7 as the reduction in the number of biological process that must occur between the onset of input (miRNA) signal and the induction of output (fluorescent protein) signal. The more direct link between input and output should bestow greater sensitivity and responsiveness upon the circuit. L7Ae was chosen as the regulator module on the basis of its substantial prior characterisation, both *in vitro* and in living cells [12, 13]. First, an experimental platform was created so that levels of fluorescently tagged circuit components could be indirectly inferred from their levels of fluorescence by flow cytometry. Subsequently, the two individual NOT gates (miRNA repression of regulator, and regulator repression of output), were first tested and optimised. Finally, the entire circuit was delivered to HEK293T cells and its performance was characterised.

Our second objective was pursued in an attempt to create an even more direct and compact circuit architecture for the positive detection of miRNAs. This was facilitated through the fusion of the regulator and output functionalities into a single molecule. Using the principles and techniques of RNA engineering (as introduced earlier) we sought to create miRNA responsive allosteric ribozymes to serve as the regulator module. This ribozyme based regulator could then be inserted in *cis* to the output gene to create a fusion between the regulator and output. A miRNA sensor circuit based around this circuit architecture would represent the most direct link between input and output signals (Fig. 1.7). Putative miRNA responsive allosteric ribozymes were created through a process involving aspects of rational design (fusion of a miRNA binding site to a hammerhead ribozyme via a random linker) and *in vitro* selection (identifying optimal linker sequence for allosteric activity). It was chosen to perform the selection within living mammalian cells (HEK293T and HeLa) based on several successful reports of *in vitro* selection carried out in similar manners [150–152]. This would maximise the chance that any selected allosteric variants would function within the environment they were intended to be deployed in. Selection was performed by two different methods (RT-PCR based [Chapter 4] and RNAseq based [Chapter 5]) to increase the chance of identifying the desired ribozyme variants. Novel methodologies were developed in order to adapt published selection protocols to serve our precise objectives of miRNA responsive allosteric ribozyme selection.

### 1.7.1 Aims and objectives

1. To create a double negative gate miRNA sensor circuit (Chapter 3):

   (a) Validate a flow cytometry based workflow for the indirect measurement of circuit components.

   (b) Characterise both individual negative gates:

i. miRNA mediated repression on L7Ae regulator.

ii. L7Ae mediated repression of fluorescent protein output.

(c) Assemble two NOT gates serially so as to create a circuit that responds in a positive manner to a specific cognate input miRNA.

2. To create a single component miRNA sensor circuit (Chapters 4-5):

(a) Identify active ribozymes that can mediate the cleavage of mRNA transcripts in *cis* within a mammalian cellular environment.

(b) Optimise the inhibition of gene expression occurring as a result of *cis*-acting ribozyme self-cleavage activity.

(c) Design a putative miRNA mediated allosteric ribozyme library through the fusion of a miRNA binding site to a hammerhead ribozyme via a randomised linker region.

(d) Develop methods for the high throughput selection of miRNA responsive allosteric ribozymes (Both PCR and RNAseq selection):

i. Optimise selection strategies for the selection of catalytically active ribozymes.

ii. Optimise selection strategies for the sub-selection (from pre-selected active ribozymes) of ribozymes demonstrating miRNA mediated allosteric inhibition.

# Chapter 2

# Materials and methods

## 2.1 DNA work

### 2.1.1 Standard PCR

PCR was routinely conducted using the 2x Phusion polymerase master mix containing a high GC content buffer (NEB). Primers were designed manually or with the Primer3 web tool (http://primer3.ut.ee/, [153]) and synthesised by Integrated DNA Technologies (IDT). A total reaction volume of 20 µl was used, including 10 µl of master mix, 1 pg – 1 ng of DNA template, forward and reverse primers at a final concentration of 500 nM, and made up to volume with nuclease free water. 35-40 amplification cycles were performed before the final amplicon was either directly purified (Qiagen, PCR cleanup kit) or resolved on an agarose gel and purified (Qiagen, gel extraction kit).

### 2.1.2 Ribozyme construction by assembly PCR

Give that the typical length of ribozymes used in this study exceeded that which could be synthesised economically in a single oligonucleotide, an assembly PCR technique was used [154]. This enabled the generation of a longer double stranded DNA molecule, of any sequence, from the assembly of multiple shorter, overlapping fragments by a PCR based process. Primers were designed using the online web tool (http://www.yorku.ca/pjohnson/AssemblyPCRoligomaker.html) and assembled according to the protocol of Rydzanic et al. [154].

### 2.1.3 Standard Cloning

Vector and inserts were cut by restriction enzyme digest (NEB) and purified by PCR cleanup kit (Qiagen) following PCR or gel extraction kits (Qiagen) following agarose gel electrophore-

sis. Vector DNA was dephosphorylated using Antarctic phosphatase (NEB). Vector and insert fragments were fused using either T4 ligase or Quick ligase (NEB) using a vector to fragment ratio that was empirically determined.

A small aliquot (~5 µl) of the ligation reaction was transformed into 50 µl of *E. coli* DH5α chemically competent cells by a 30 s heat shock at 42 °C. Following transformation, bacterial cells were diluted in 1 ml SOC and incubated at 37 °C for 1 h whist being agitated by 225 RPM shaking. Finally bacteria were concentrated by centrifugation and resuspended in a volume of approximately 200 µl, all of which was spread over a 10 cm LB agar plate containing the appropriate antibiotic selection (pLentiLox and pcDNA based vectors, 100 µg/ml ampicillin; pEM based vectors, 50 µg/ml kanamycin) and incubated overnight at 37 °C.

Plates were removed from the incubator after 12-16 h and stored at 4 °C. A number of colonies were chosen, based on the ratio of colonies found on the ligation plates relative to the vector only control, and grown overnight in 5 ml LB media containing antibiotic selection at 37 °C whist being agitated by 225 RPM shaking. 12-16 h later, bacteria were resuspended by centrifugation and plasmids were extracted using a miniprep kit (Qiagen).

### 2.1.4 High efficiency cloning of random ribozyme pools

A high efficiency cloning method was required to ensure maximal representation of the ribozyme sequence space within the plasmid library. In order to achieve this, the GEKO cloning protocol which described the high efficiency cloning of approximately 80,000 guide RNAs was adapted as follows [155].

An approximately 80 nt oligonucleotide was synthesised (IDT) encoding a putative ribozyme sequence containing a region of six internal degenerate bases. This oligonucleotide was PCR amplified using Phusion polymerase for 10 cycles, following which it was purified using a PCR cleanup kit, and subject to restriction enzyme digest in parallel with the destination vector. Both vector and ribozyme encoding insert were resolved on an agarose gel before gel extraction under long wavelength UV (~365 nm) illumination. Approximately 100 ng of vector and 20 ng of insert were used in a 20 µl ligation reaction with T4 ligase. Four transformation reactions, each containing 5 µl of ligation reaction and 25 µl of High efficiency, *E. coli* 10β electrocompetent cells were prepared on ice and transferred to electroporation cuvettes. Next, electroporation was performed on an Eppendorf Eporator set to 1660 V. 1 ml of warm SOC was immediately used to suspend each batch of transformed bacteria which were incubated at 37 °C for 1 h whist being agitated by 225 RPM shaking. Approximately 200 µl of bacteria were spread over each of approximately 20, 10 cm LB agar plates containing 100 µg/ml ampicillin. After approximately 14 h of growth colonies were scraped and pooled from each plate into LB media, following which they were concentrated into a pellet by centrifuga-

tion. Mixed plasmid DNA was recovered from this pellet using a maxiprep endonuclease free kit (Qiagen).

### 2.1.5 Common expression vectors

All miRNAs are expressed from the pEM plasmid backbone (obtained from the tenOever lab). This contains a strong constitutive CMV promoter that drives the expression of a RFP gene. Contained within the RFP gene is an intron sequence, into which primary miRNA transcripts can be cloned. To avoid fluorescence interference from RFP protein into other channels (eg. GFP channel, 488 nm excitation, 530/40 nm emission filters) during flow cytometry, the RFP gene was inactivated through the insertion of a premature stop codon. pEM-empty and pEM-122 were gifts from the tenOever Laboratory. pEM-21 was constructed by PCR amplification of the miR-21 primary transcript gene, from mammalian cell genomic DNA, and restriction enzyme cloning into the pEM-empty vector at SpeI and EcoRI sites.

A number of plasmids are used for the expression of reporter genes, some of which harbour ribozymes or other regulatory sequences in their 5' or 3' UTRs. The LentiLox 3.7 (pLL3.7, [156]) plasmid was used to express (destabilised)EGFP coupled ribozyme containing transcripts for optimisation experiments. The commercial pcDNA3.1 plasmid (Life Technologies) is frequently used for the expression of reporter genes from the strong CMV promoter. Cloning of reporter genes into these vectors was performed with restriction endonucleases targeting the regions in the multiple cloning site (MCS).

Full annotated maps of all plasmids used in this work can be accessed at:
https://drive.google.com/open?id=0BzqJTxzVRnHGZEloclNkUlk4dFU,
or
http://tinyurl.com/hbkxt4q.

## 2.2 RNA work

### 2.2.1 RNA extraction

Total RNA was extracted from mammalian cells using either the RNeasy or miRNeasy kits (Qiagen). miRNeasy was only used when the recovery of short RNAs, including miRNAs, was required. For all other purposes, RNA was extracted using RNeasy. Extracted RNA was dissolved in RNase free water and stored at either -20 °C for the short term or -80 °C for longer term storage. A small aliquot (~2 µg) of total RNA was run on a 1 % agarose gel to ensure the presence of strong bands corresponding to the 28S and 18 S ribosomal RNAs (at an expected 2:1 intensity ratio) and no signs of smearing which may indicate degradation.

### 2.2.2 RT-PCR

DNA contaminants were first removed from total RNA using TurboDNase (Life Technologies). DNA free total RNA was subject to reverse transcription (RT) with the iScript Reverse Transcription Supermix (Bio-Rad) according to the manufacturer's protocol. The resulting cDNA was then used as a template for standard PCR as described above.

### 2.2.3 Gene specific RT-qPCR

A modified RT-qPCR method was used throughout this study which enabled the amplification from mRNA templates but not contaminating DNA. This was necessary upon the discovery, early in the work, of a substantial lab wide contamination with a GFP encoding plasmid.

This technique was based on the specificity of the reverse transcriptase enzyme for an RNA template and the temperature differential between RT and the primer annealing step of PCR. Chimaeric, gene specific RT primers were designed, which contained approximately 12 nucleotides complementary to their RNA target at the 3' end, and an approximately 12 nucleotide 5' tail of unique sequence. This short region of complementarity at the primer 3' end was sufficient to anneal to its target during RT at 42 °C but not during the PCR annealing step at $\geq$60 °C. The resulting cDNA molecules were thus tagged by a unique sequence at their 5' end. PCR amplification was conducted using a forward primer, fully complementary to both cDNA and contaminating plasmid DNA. The RT primer was used as the reverse PCR primer, and at PCR annealing temperatures could only bind to the hybrid target site created on the cDNA molecules, but not on any DNA contaminants. This technique was performed using the SensiFAST SYBR No-ROX One-Step Kit (Bioline) on a CFX384 PCR system (Bio-Rad) and analysed using the provided CFX manager software (Bio-Rad).

A standard curve was performed upon receipt of each new primer pair in order to optimise PCR conditions. Additionally no-reverse-transcriptase controls were performed to ensure that any DNA contaminants were not subject to amplification. All qPCR reactions were performed in technical triplicates in addition to any biological replicates performed for individual experiments.

### 2.2.4 miRNA RT-qPCR

Performing RT on a miRNA template presents difficulties associated with their very short length. In order to overcome this, a technique has been devised using a structured stem-loop RT primer. This structured stem-loop primer contains an approximately 6 nt single stranded overhang at its 3' end with complementarity to the 3' end of the miRNA. The double stranded

regions elsewhere in the stem-loop primer prevent the spurious binding to internal regions in off-target RNAs through steric hindrance and thus provide the determinant of specificity towards a chosen miRNA.

TaqMan miRNA assay kits (Life Technologies) were ordered for each specific miRNA to be measured by RT-qPCR. These kits contained the structured stem-loop RT primer in addition to both forward and reverse PCR primers and TaqMan probe. RT was performed with the TaqMan miRNA Reverse Transcription Kit (Life Technologies) according to the manufacturer's instructions. The resulting cDNA was then used as an input for a qPCR reaction with the TaqMan Universal Mastermix (Life Technologies).

## 2.3 Tissue culture work

### 2.3.1 Cell culture

HeLa and HEK293T cells were cultured at 37 °C and 5% $CO_2$ in DMEM with 25 mM glucose (Gibco, 41966-029), 10% (v/v) fetal bovine serum (FBS, Gibco, 10270) and 1% (v/v) penicillin-streptomycin (Gibco, 15140-122). These were passaged every two days at dilutions of 1:4 and 1:6 for HeLa or HEK293T cells respectively. Cultures were maintained for approximately 25-30 passages or until phenotypic changes were observed at which point a new vial of early passage cells was thawed.

### 2.3.2 Transfection

Both HeLa and HEK293T cells were transfected using polyethylenimine (PEI) according to the methods of Aricescu *et al.* [157]. Briefly, cells were seeded into 12 or 24 well plates 24 h prior to transfection at a density of ~1x$10^5$ cells/ml such that they would be at a confluence of approximately 80 % at the time of transfection.

On the day of transfection, plasmid DNA and PEI transfection reagent were added to Opti-MEM (Life Technologies) in a ratio of 1:1.5, and incubated at room temperature for about 10 min to enable the formation of DNA/PEI complexes. Transfection mixes destined for each well of HEK293T cells in a 12 well plate were made with approximately 2 µg of plasmid DNA (total DNA which may be comprised of multiple plasmids for a mixed transfection) and 3 µl PEI in 100 µl Opti-MEM. HeLa cells were transfected with approximately half the amounts of DNA and PEI, in the same volume of Opti-MEM, due to a greater observed sensitivity to plasmid DNA toxicity. Transfection mixes were scaled down by a factor of two for the transfection of each well of cells in a 24 well plate.

### 2.3.3 Flow cytometry

Flow cytometric analysis was performed 48 hours post-transfection, using a suspension of single cells. Media from cells was first removed before washing with phosphate-buffered saline (PBS) in order to remove any residual FBS. Subsequently approximately 200 µl of a solution of 0.05% trypsin was added to each well in order to detach the cells. An incubation of approximately 5-10 min at 37 °C, followed by the addition of 1 ml of full media (DMEM with 10 % FBS) was sufficient to enable the gentle detachment of cells by pipetting. Cells were pelleted from this suspension through centrifugation at 500 g for 3 min, after which the media supernatant was removed and cells were resuspended in PBS.

Flow cytometry was performed 48 hours post-transfection, using a CyAn (DakoCytomation) machine. A 1 ml suspension of cells was applied to the machine upon which 50,000 events were recorded per sample. Cells were gated, first on their forward versus side scatter profiles in an attempt to remove any contribution from auto fluorescent cellular debris, and secondly on their pulse width versus side scatter to enable the exclusion of clumps of more than one cell. Fluoresence signal was measured in either the CFP (405 nm excitation, 450/50 nm emission filters), and/or GFP (488 nm excitation, 530/40 nm emission filters), and/or iBlue (642 nm excitation, 665/20 nm emission filters) channels depending on the complement of reporter proteins used in the particular experiment. Details concerning the calculation of mean and normalised fluorescence levels are discussed in the following results chapters.

## 2.4 Sequencing

### 2.4.1 MiSeq library preparation

RNAseq libraries were prepared by a two-step RT-PCR amplification process. P5 and P7 Illumina adapters (corresponding to sequences that are required at the 5' and 3' termini of amplicons to enable their annealing to the Illumina sequencing flowcell), as well as multiplex barcodes, were encoded in the two pairs of forward and reverse primers such that these sequences could be fused to the termini of the PCR amplicons. This enabled the direct application of PCR amplicons to the MiSeq, and avoided the typical process of adapter ligation. RT was carried out in two different ways depending on whether linear or structured stem-loop primers were being used.

In the case of linear primers, RT-PCR was conducted in a single tube using the SensiFAST SYBR No-ROX One-Step Kit (Bioline). A single primer was used to prime both the RT reaction and to serve as the reverse primer in the subsequent PCR amplification. This primer was a chimaeric primer, as discussed above, which ensured the specific amplification only

from RNA templates. Given this specificity, a DNAse digestion step was not performed on the isolated RNA. Structured stem-loop primers were used for RT where it was necessary to specify a precise 3' end from which to initiate the RT reaction, for example when selecting for ribozymes that have undergone self-cleavage. Structured stem-loop primers were designed using a web tool (http://genomics.dote.hu:8080/mirnadesigntool/) according to the methods described by Czimmerer *et al.* [158]. For this purpose, RT was performed using the TaqMan miRNA Reverse Transcription Kit (Life Technologies) according to the manufacturer's instructions. Resulting cDNA was amplified with 2x Phusion polymerase master mix containing a high GC content buffer (NEB) in a first round PCR.

This first PCR was performed on each individual sample derived from a different transfection condition. Initially, all first stage PCR amplifications were performed over 25 cycles, following which an aliquot was resolved by agarose gel electrophoresis. DNA was visualised at a short exposure (0.2 s) on a UV transilluminator to estimate total quantities of PCR amplicon across each sample. The PCR was then repeated and cycle number was adjusted for each product such as to achieve faintly visible bands with as few amplification cycles as possible. The reasoning behind minimising the number of amplification cycles involved the reduction of PCR associated stochastic noise as discussed in greater detail in Chapter 5.

Having obtained the first round PCR product, samples containing different multiplexed barcodes could be pooled before the second PCR. This second reaction was performed for only 10 cycles, following which amplicons were purified by PCR cleanup kit (Qiagen) to remove any PCR primers or dNTPs. Further purification was performed with AMPure XP beads to remove any residual contaminating primers that may otherwise interfere with the downstream sequencing. This was performed according to the manufacturer's instructions with the exception of a modification to the ratio of DNA to beads of 1:1. This AMPure XP purification was based on a size selection to discard DNA strands <100 bp (eg. primers) and enrich longer species (eg. PCR amplicons destined for sequening). The size distribution of amplicon fragments within the library was determined using the TapeStation (Agilent Technologies) to demonstrate the predominance of a single sized PCR amplicon fragment of the expected size and the abscence of any residual primers. Finally, DNA concentration (in ng/µl) was determined using the Qbit dsDNA high sensitivity kit (Life Technologies) according to the manufacturer's instructions. Given the knowledge of the amplicon length this could be converted to a concentration in moles/l as required to ensure loading of the Illumina sequencing flowcell with the correct amount of DNA.

### 2.4.2 RNAseq

The sequencing library (a solution of amplicons diluted to a final total DNA concentration of 4 nM) was loaded onto a MiSeq Reagent Kit v2 (300-cycles) and sequenced to completion. Given the low complexity nature of this PCR amplicon sequencing library, it was necessary to take measures to artificially increase the complexity. This is a result of the fluorescent image processing that detects the identity of the base added at each extension cycle during sequencing at each of the millions of loci spotted on the flowcell. If the library is of low complexity and thus the same base is added at the majority of loci then the ability of the software to correctly resolve two adjacent loci (corresponding to two different DNA species being sequenced in parallel) is diminished. This issue can be addressed by spiking in a high complexity library, such as one derived from the PhiX viral genome however this sacrifices a fraction of sequencing depth. A more attractive method, and the one used here, is the use of staggered PCR primers during library prep. These contain a variable length stagger sequence between the priming site at the 3' end of the primer and the 5' sequence elements corresponding to Ilumina adapter sequences. High throughput sequencing of such a staggered PCR amplicon library will generate multiple reads in different frames such as to mimic a higher level of complexity.

### 2.4.3 Sequencing data analysis

FASTQ data files were extracted following sequencing analysis and pre-processed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). 5' and 3' adapters as well as fixed ribozyme sequences were removed. Where necessary, for digital RNAseq, identical sequences were collapsed during this pre-processing stage.

The pre-processed FASTQ files were then processed via the FASTaptamer toolkit in order to count the unique occurrences of each sequence combination within the region of randomisation and generate an output FASTA data file containing this information [159].

Finally, R was used to tabulate the occurrences of each unique sequence and perform statistical analysis on this data with the help of an expert bioinformatician (Quentin Ferry).

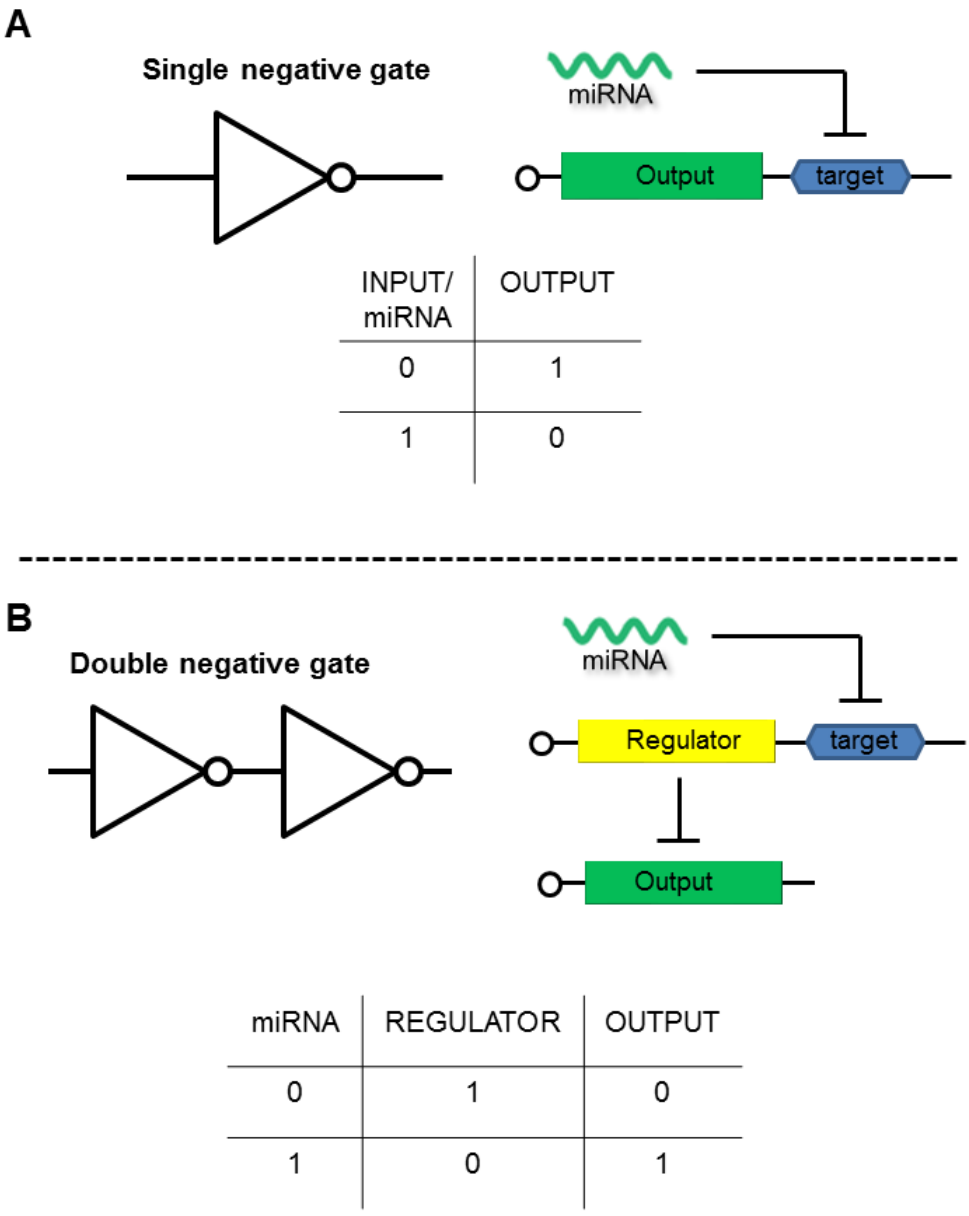
# Chapter 3

# Construction of a novel double-negative gate miRNA sensor circuit

## 3.1 Introduction

miRNAs are capable of highly specific and efficient inhibition of target transcripts to which they bear perfect complementarity. The interaction between a miRNA and its target results in an action that can be compared to that of a NOT gate, which behaves according to the truth table detailed in Fig. 3.1A. This circuit architecture has previously been exploited for the generation of miRNA sensors as tools for basic scientific research, in which the expression of a reporter gene is negatively regulated by a miRNA target site(s) within the 3' UTR of this gene [145]. In order to enhance the versatility of miRNA-controlled circuits we sought to implement a double-negative gate circuit, capable of generating a positive output signal in response to the presence of a specific input miRNA signal (Fig. 3.1B). Such a genetic circuit is comprised of two serially linked NOT gates, in which the input signal from a specific miRNA represses expression of a regulator protein (regulator module) that in turn negatively controls the expression of an output gene of interest. Using this strategy it is possible to coerce a specific miRNA to induce an OFF-to-ON state transition in a genetic circuit.

The above circuit architecture has previously been successfully implemented in several studies using a transcriptional repressor as a regulator module [135, 148, 149]. However, these circuits suffer from the disadvantage that there is considerable lag between the onset of a miRNA input signal and the induction of the output response. This occurs as a result of numerous cellular processes that must ensue in order to couple the initial input signal to the eventual expression of the output protein. Such processes include degradation of transcriptional repressor mRNA, decay of the corresponding repressor protein governing induction of




| INPUT/ miRNA | OUTPUT |
|---|---|
| 0 | 1 |
| 1 | 0 |

| miRNA | REGULATOR | OUTPUT |
|---|---|---|
| 0 | 1 | 0 |
| 1 | 0 | 1 |

Figure 3.1: Schematic representation of Boolean NOT gates and genetic circuits that can recapitulate their function
*(A) Schematic of a single NOT gate that functions as a signal inverter as observed in the truth table. An equivalent genetic circuit can be created with a miRNA (input) acting via a target site to repress the expression of an output. (B) Schematic of double NOT gates coupling two signal inverters in series such that the input and output signals are concordant as demonstrated by the truth table. An equivalent genetic circuit can be created with a miRNA (input) acting to repress a regulator which in turn acts to repress the output.*

transcription of the output gene, and finally translation of the output mRNA into its cognate protein. Most of this lag can likely be attributed to the turnover rate of the regulator protein, which may persist after its encoding mRNA has been destroyed by a targeting input miRNA.

In this study, an alternative strategy was pursued that aimed to couple the miRNA input signal to induction of the output reporter protein with fewer intervening biological steps. This was achieved through a regulator protein which modulates output expression at the post-transcriptional level, as opposed to transcriptional level as was the case in previous implementations [135, 148, 149]. Conceptually, in this system, the output gene is continuously transcribed from a constitutive promoter, such as the CMV promoter ($P_{CMV}$), allowing the accumulation of its mRNA to steady state levels, while translation from this template is inhibited post-transcriptionally by the regulator protein. Upon inhibition of expression of the regulator protein by a miRNA input signal, the existing pool of output mRNA should become immediately available for translation, thus enabling the prompt production of output reporter protein. However, a short lag between the onset of the input signal and the induction of output expression may still arise due to transient persistence of the regulator protein, even after its encoding mRNA has been silenced by its matching input miRNA. To address this issue we explored several strategies aimed at reducing the half-life of the regulator module.

The presence of a stable secondary structure within the 5' UTR of mammalian transcripts is known to interfere with ribosome scanning and thus reduce the efficiency of translation initiation [160]. A system that enables the conditional formation and/or stabilisation of such 5' UTR secondary structures could therefore form the basis of a translational negative regulatory switch. Such systems have been constructed by placing RNA aptamer sequences within 5' UTRs to facilitate secondary structure stabilisation upon binding of corresponding cognate ligands [12, 13, 161]. In the absence of ligand, the aptamer structure displays insufficient stability to block progression of ribosome scaninning. However, following ligand binding, the resulting stabilised structure has the ability to block ribosome scanning thus efficiently inhibit translation. Saito *et al.* utilised the Archaea-derived Box C/D Kink-turn (Kt) RNA motif and its cognate protein ligand, L7Ae, as modular components for the construction of such a genetic NOT gate [12]. In their study, the Kt motif, when placed in the 5' UTR of a reporter transcript was found to regulate expression of the reporter in a L7Ae-dependent manner.

The experiments described in this chapter focus on employing and combining these tools to create a double-negative gate circuit for positive detection of target miRNA activity (Fig. 3.2). Both the regulator and output genes are transcribed from constitutive promoters such that all circuit-related regulatory steps occur at the post-transcriptional level. An OFF state is maintained in the absence of target miRNA by the unopposed repressive effect of the regulator protein on the translation of the output (Fig. 3.2A). In the presence of a cognate miRNA how-

ever, the regulator is repressed thus permitting expression of the output protein and switching of the circuit into an ON state (Fig. 3.2B).

## 3.2 Results

### 3.2.1 Flow cytometry-based quantification is a valid and readily applicable assay for output reporter protein expression

Evaluating the performance of the proposed genetic circuits required the ability to quantitatively measure the level of output gene expression. This could be achieved by either measuring output mRNA levels by RT-qPCR or the levels of output protein using quantitative Western blots (direct measurement) or indirectly by flow cytometry (detection of a quantifiable fluorescence signal).

Although flow cytometry is preferable due to its relative experimental simplicity and quantitative output, it is nonetheless an indirect measure of output gene expression. Therefore it was necessary to investigate the degree to which flow cytometry-determined measurements correlate with levels of the corresponding transcript. Furthermore, a series of preliminary experiments were performed in order to determine the conditions and controls required to interpret the data generated from this work.

A standard pipeline was applied to the analysis of all flow cytometry data (Fig. 3.3). This involved the identification of putatively viable cells based on assessment of their size and granularity using the forward and side scatter parameters (Fig. 3.3A). A further filtering step was then applied in order to select only single cells as opposed to clusters, which may otherwise display multiplied fluorescence signals (Fig. 3.3B). Next, the fluorescence signals from single viable cells were subjected to analysis. The majority of flow cytometry experiments involved the measurement in two fluorescence channels, corresponding to a fluorescent reporter protein of interest and a transfection control corresponding to a second fluorescent protein. Given that the predominant reporter protein used throughout this thesis was a GFP derivative, it was important to select for the transfection control a fluorescent protein with sufficiently different spectral properties such that its signal could be readily differentiated from that of GFP. For this reason a transfection control plasmid (piBlue-N1, a gift from Michael Davidson (Addgene plasmid # 54781)) encoding the iBlue protein was chosen. iBlue is a fluorescent protein with excitation and emission wavelengths near the far-red end of the visible light spectrum (648 nm excitation, 667 nm emission) and thus the emission of iBlue was not expected to overlap to any significant degree with that from GFP. In order to confrm the ability to accurately discriminate signals from GFP and iBlue, cells were transiently transfected with

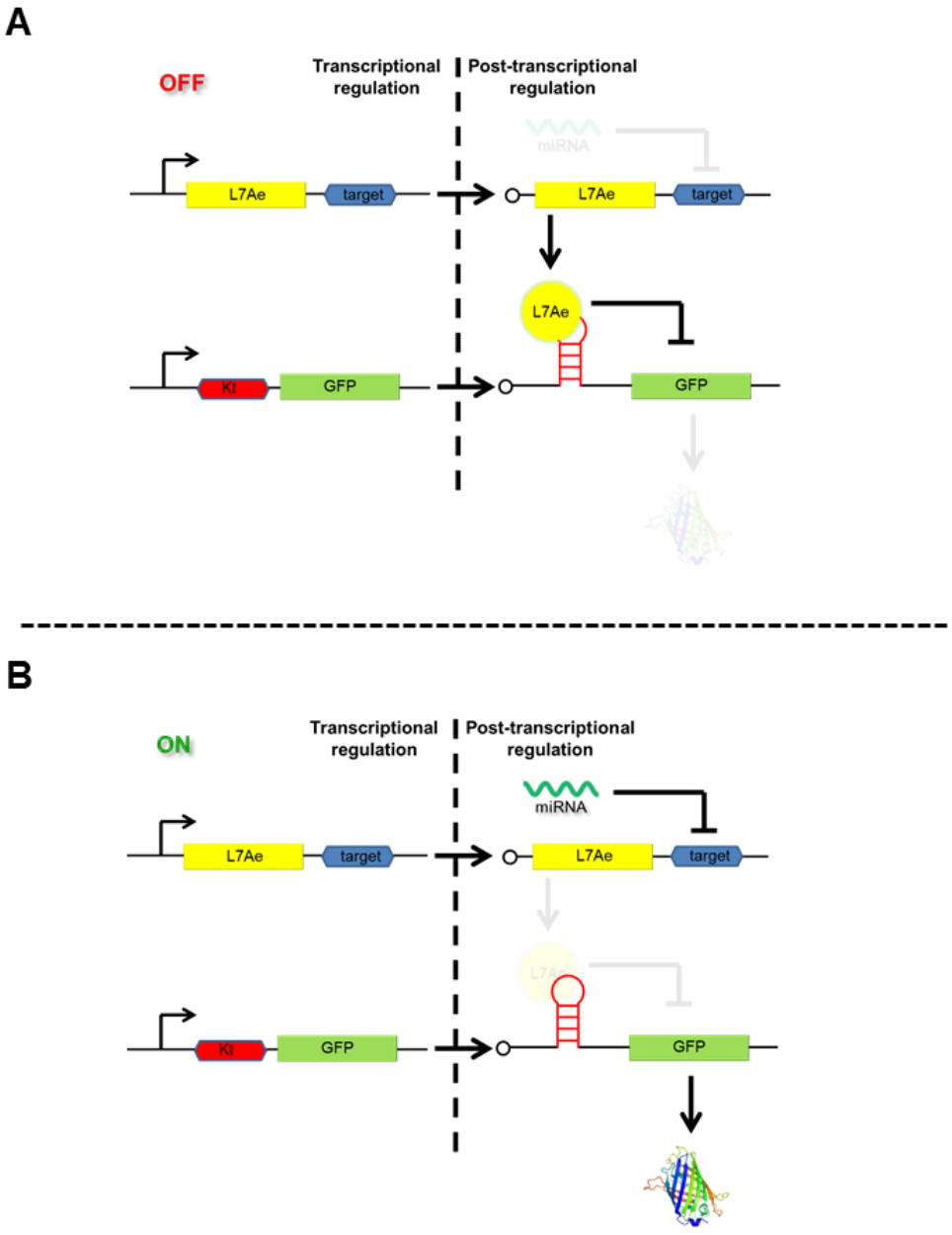



Figure 3.2: Schematic diagram of a biological double-negative gate circuit for the positive detection of miRNA activity

*This circuit operates in the manner shown in Fig. 3.1B. The components of this circuit are constitutively transcribed (left hand side, transcriptional regulation), with miRNA-dependent regulation occurring at the post-transcriptional level (right hand side, post-transcriptional regulation). The input (miRNA) acts to repress expression of the regulator (L7Ae) in the first NOT gate, which in turn is coupled to the second NOT gate in which the regulator (L7Ae) represses expression of the output (GFP). In the absence of miRNA (A), the unopposed L7Ae regulator functions to silence the output and hold the circuit in an OFF state. However, in the presence of a specific miRNA (B), the L7Ae regulator is subject to repression thus enabling the expression of GFP output and a circuit ON state.*

constructs coding for one or the other fluorescent protein alone before analysis by flow cytometry and measurement of the emitted signal using both the GFP (488 nm excitation, 530/40 nm emission filters) and iBlue detection channels (642 nm excitation, 665/20 nm emission filters). Cells transfected with a dEGFP expression vector (pcDNA-dEGFP) demonstrated a high intensity fluorescence signal specifically in the GFP channel with no iBlue fluorescence detected above the background level seen in non-transfected cells (Fig. 3.3, D compared to C); the inverse was observed in cells transfected with an iBlue expression vector (piBlue-N1, Fig. 3.3, E compared to C). This clear separation in fluorescence signals from the dEGFP and iBlue proteins made them an ideal pairing for the circuit output and transfection control reporter proteins, respectively. In several subsequent experiments in this thesis, further circuit elements were tagged with different fluorescent reporter proteins (ECFP, RFP). The emission spectra from these two fluorescent proteins are not as distinct and easily discernible by flow cytometry from that of GFP. To control for this, all such experiments were performed with a panel of cells transfected with plasmids expressing the individual single fluorescent proteins to enable colour compensation [162].

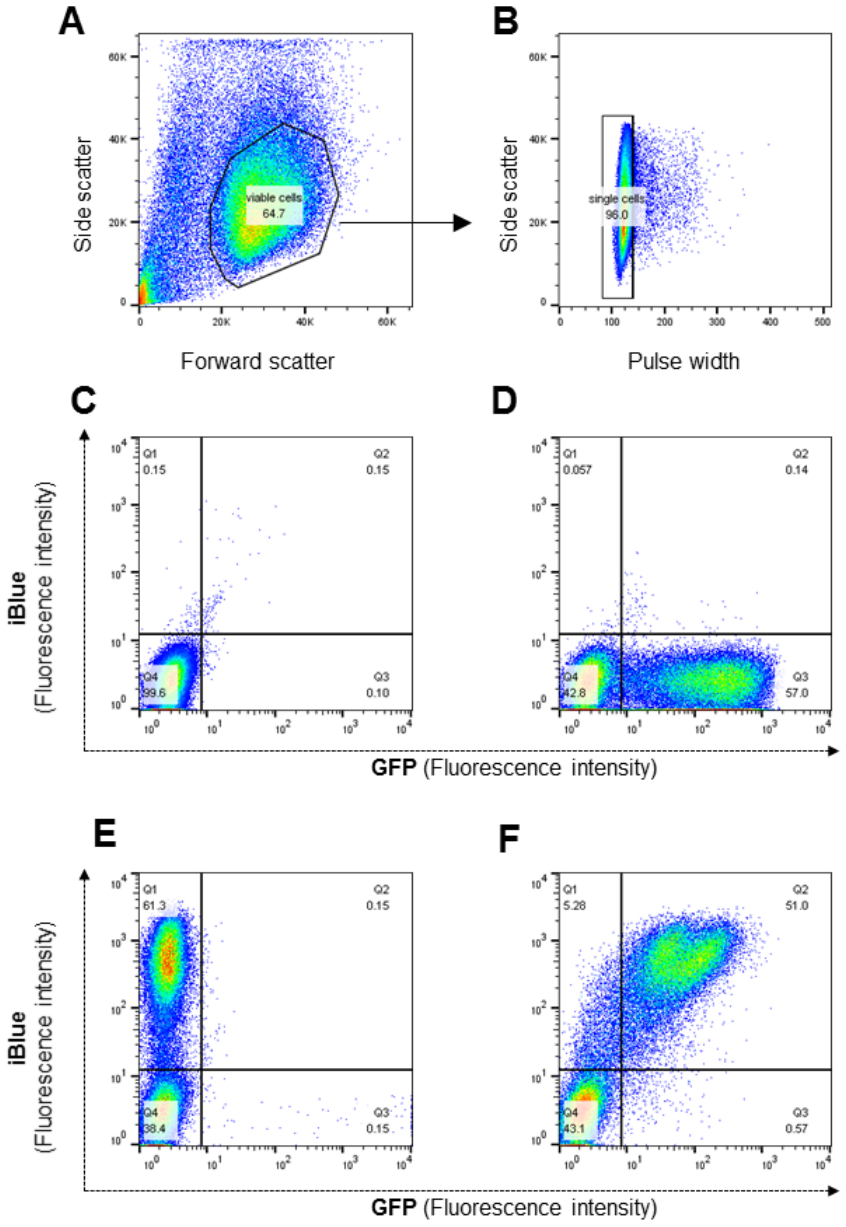



Figure 3.3: Gating of flow cytometry data from HEK293T cells

*(A) Putatively viable cells were gated on forward scatter and side scatter. (B) Viable cells were further classified by isolating single cells from doublets through gating on pulse width and side scatter. (C - F) The fluorescence signals of single, viable cells were visualised on a scatter plot charting GFP fluorescence against iBlue fluorescence. (C) Non-transfected cells served as a control on which to define gates based on the autofluorescence of cells. Gates Q1, Q2, Q3 and Q4 corresponded to iBlue positive, GFP/iBlue double-positive, GFP positive and non-transfected cells, respectively. (D) HEK293T cells transfected with 1 µg pcDNA-dEGFP alone. 57.2 % of cells demonstrated a fluorescence signal above background (gates Q1 + Q2 + Q3), while the vast majority (99.7 %) of these lay within gate Q3 representing cells positive for GFP alone. Minimal signal was detected in the iBlue channel confirming an absence of cross-talk. Continued...*

The final analysis step aimed to assess the relative fluorescence intensities of reporter proteins (Fig. 3.4). Based on the plot charting reporter fluorescence (GFP) against that of the transfection control (iBlue), a gate was defined above the threshold fluorescence level of iBlue to encompass all transfected cells (Fig. 3.4A). This gate was designated to lie immediately above the background levels of iBlue-like autofluorescence observed in non-transfected cells (Fig. 3.3C). The mean fluorescence intensity of transfected cells, in both the GFP and iBlue channels, was then calculated (Fig. 3.4B). Finally, the relative reporter (usually GFP, but also CFP or RFP) fluorescence was calculated (Fig. 3.4B, [relative reporter fluorescence] = [mean reporter fluorescence] ÷ [mean transfection control fluorescence]).

To validate flow cytometry fluorescence measurement as a suitable method for the assessment of output gene expression the data obtained using this method were directly compared to results from RT-qPCR analysis. Using this strategy, it was possible to determine the correlation between mRNA transcript abundance and the fluorescence intensity of its cognate reporter protein. HEK293T cells were transfected with an expression vector for the fluorescent protein iBlue and 48 h later sorted by FACS into four bins of increasing iBlue fluorescence intensity. Total RNA was extracted from cells contained within each bin and subjected to RT-qPCR measurement of iBlue gene transcript abundance. In accordance with similar data from the Smolke laboratory a strong correlation between mean fluorescence intensity and transcript abundance was demonstrated (Fig. 3.4C) [163]. These results supported the use of flow cytometry based measurements as a suitable means to measure levels of output gene expression from subsequent circuits.

*Figure 3.3 continued: (E) HEK293T cells transfected with 1 µg piBlue-N1. As above, the transfected cells demonstrated high iBlue fluorescence with vanishingly minimal spillover into the GFP chanel. (F) HEK293T cells co-transfected with 500 ng pcDNA-dEGFP and 500 ng piBlue-N1. 56.9 % of all cells were transfected. Of these, 89.6 % lay within gate Q2 and thus represented cells that were expressing both dEGFP and iBlue. This suggests a strong concordance between GFP and iBlue signals and supports the conclusion that the majority of transfected cells harboured both plasmids. Indeed, it is possible that the small population of transfected cells that demonstrated iBlue fluorescence alone represented an artifact given the greater stability of the iBlue protein relative to dEGFP which contains a fused destabilising PEST domain.*

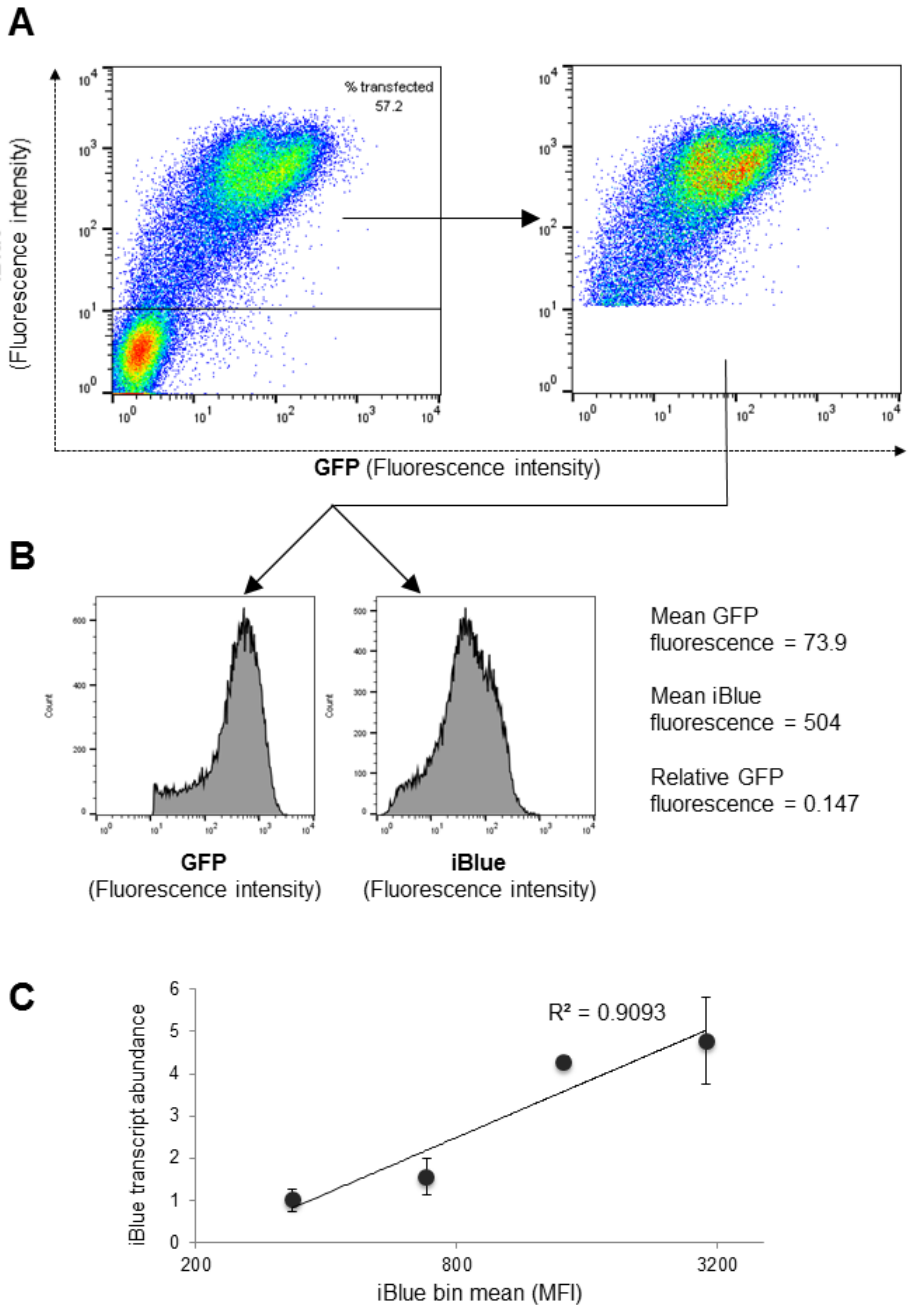



Figure 3.4: Standard analysis pipeline for generation and validation of flow cytometry data as a means to measure protein and/or gene expression levels

*(A) Cells were gated based on an above background level of fluorescence in the transfection control channel (iBlue). (B) The mean fluorescence intensity of both the transfection control (iBlue) and fluorophore of interest (GFP) was calculated. Relative GFP fluorescence was then defined as the ratio of mean GFP to mean iBlue fluorescence. (C) HEK293T cells were transfected with 1 µg of piBlue-N1. 48 h post-transfection, cells were sorted by FACS into four triplicate bins of increasing iBlue fluorescence. RNA was extracted from each sample and subjected to RT-qPCR measurement of expression of the iBlue and GAPDH housekeeping genes. The mean iBlue fluorescence from each bin against the abundance of the iBlue transcript relative to its GAPDH counterpart was then plotted (n = 3, error bars represent standard deviation).*

### 3.2.2 Investigation of the first NOT gate reveals dose-dependent miR-122-mediated repression of L7Ae expression in a model-predicted target site-dependent manner

The proposed circuit was deconstructed into its most basic constituent parts, two NOT gates, for the purposes of optimisation before assembly of the complete circuit. The first logic gate to be tested was that comprising miRNA-mediated repression of the L7Ae regulator module (Fig. 3.5A). The entire L7Ae coding sequence was synthesised as double-stranded DNA (gBlock, IDT) and cloned into the pcDNA3.1 eukaryotic expression vector to generate pcDNA-L7Ae. In order to engineer miRNA-responsiveness into this vector, two perfectly complementary miR-122 target sites were incorporated into the 3' UTR of the L7Ae gene to construct pcDNA-L7Ae-122T. Ago2-catalysed cleavage of the target L7Ae mRNA transcript that would be expected to occur upon binding of a cognate miRNA to these perfectly complementary target sites should ensure maximal repression relative to alternative miRNA-mediated post-transcriptional silencing mechanisms. To experimentally follow the expression levels of the L7Ae regulator, both expression vectors were modified to generate pcDNA-ECFP-L7Ae and pcDNA-ECFP-L7Ae-122T, each of which encoded an identical fusion ECFP-L7Ae fluorescent reporter protein, thus enabling direct analysis by flow cytometry. miR-122 was chosen as the cognate miRNA for this work due to its recognised association with hepatitis C virus (HCV) replication and the substantial disease burden associated with this pathogen [164]. Even though miR-122 is not expressed in HEK293T cells, it can be exogenously expressed from a plasmid to study its regulation of targets. We have used this strategy in subsequent work, in which the sequence coding for a primary miR-122 transcript was cloned into the pEM expression vector (gift from Benjamin tenOever) within the intron of a $P_{CMV}$-driven non-coding transcript to create pEM-122. Thus, following transfection of pEM-122 into HEK293T cells, the pri-miR-122 would be predicted to be subjected to processing by the endogenous Drosha and Dicer cellular machinery before being loaded into the RISC where it would be competent to regulate miR-122 targets. A complementary control vector (pEM-empty) which encodes a matching transcript that lacks a miRNA sequence within its matching intron was also constructed.

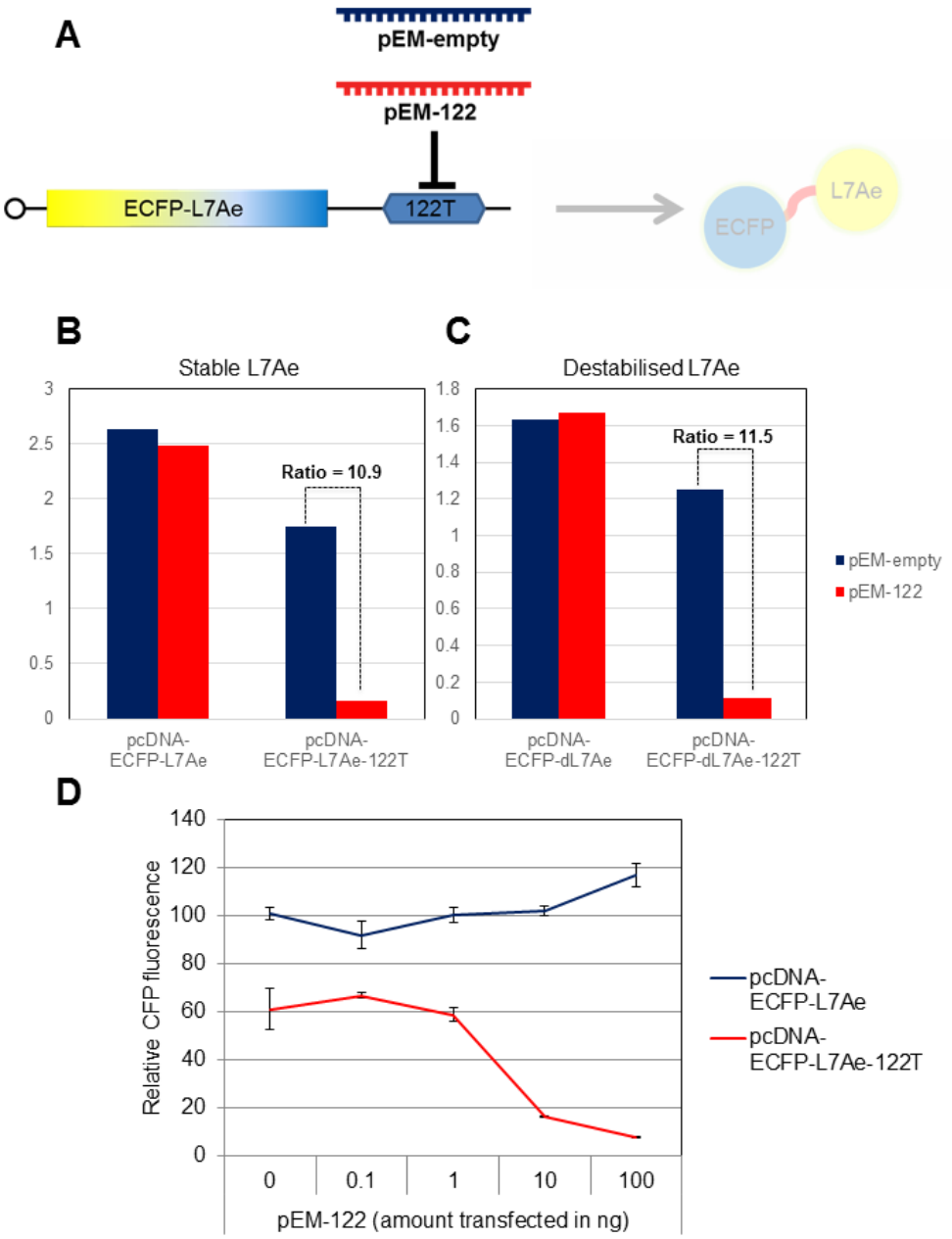



Figure 3.5: Investigation of miR-122 mediated inhibition of a L7Ae regulator

*(A) Schematic diagram illustrating the potential inhibitory interaction between a miRNA and its cognate target site located within the 3' UTR of a transcript coding for an ECFP labelled L7Ae regulator. (B) The levels of L7Ae were indirectly assessed through measurement of relative CFP fluorescence. Approximately 2 × $10^5$ HEK293T cells were transfected with 50 ng of the L7Ae-ECFP ± miR-122 3' UTR target site-bearing construct, 100 ng of either the empty or miR-122 expression vector, and 150 ng of piBlue-N1. These data demonstrated that there was a specific reduction in L7Ae-ECFP fluorescence which was dependent on the presence of both miR-122 and its cognate target site (n = 1). (C) Approximately 2 × $10^5$ HEK293T cells were transfected with 50 ng of the dL7Ae-ECFP ± miR-122 3' UTR target site-bearing construct, 100 ng of either the empty or miR-122 expression vector, and 150 ng of piBlue-N1. Continued...*

CFP fluorescence from HEK293T cells transfected with pcDNA-ECFP-L7Ae-122T was reduced by over 10 fold when co-transfected with pEM-122 as compared to pEM-empty (Fig. 3.5B). This inhibition was miR-122 target site-dependent as demonstrated by the almost unchanged levels of CFP fluorescence following co-transfection of pcDNA-ECFP-L7Ae with either pEM-122 or pEM-empty. Thus these results are consistent with miRNA-mediated inhibition of ECFP-L7Ae expression via an interaction with the two introduced miR-122 3' UTR-located target sites.

The sensitivity of this circuit would be predicted to be dependent on the degree to which the L7Ae regulator could be repressed and therefore assessment of the latter was subject to further investigation. Certain endogenous proteins, such as those associated with the cell cycle and other regulatory processes, are expressed in a highly regulated temporal manner. These proteins often exhibit extremely short half-lives, a feature that is key to the tight coupling of cognate transcript production to steady state protein levels. Mouse ornithine decarboxylase (MODC), an enzyme involved in polyamine biosynthesis, is one of the shortest-lived eukaryotic proteins known, with a half-life estimated at approximately 30 minutes. The trigger for rapid decay of MODC has been attributed to a C-terminal domain rich in Pro, Glu, Ser and Thr residues, termed the PEST domain [165]. The discovery of PEST domains in other short-lived proteins supported its role as a key signal for rapid decay [166]. It has been shown that the PEST-dependent destabilising effect can be conferred upon other stable proteins (EGFP to dEGFP) through generation of fusion constructs [167]. In an attempt to sensitise the L7Ae regulator the modified expression vectors pcDNA-ECFP-dL7Ae and pcDNA-ECFP-dL7Ae-122T were constructed for the expression of the L7Ae variant (dL7Ae) containing the MODC type C-terminal PEST domain.

The CFP fluorescence of cells transfected with pcDNA-ECFP-L7Ae or pcDNA-ECFP-dL7Ae could then be used to indirectly assess the effect of destabilisation. Given that these

*Figure 3.5 contiued: These data demonstrated that there was a specific reduction in dL7Ae-ECFP fluorescence which was dependent on the presence of both miR-122 and its cognate target site (n = 1). Additionally, comparison of these data with those in panel B revealed that use of the ECFP-tagged dL7Ae as compared to L7Ae construct had negligible, if any, impact on the switching ratio but did result in notably weaker relative CFP fluorescence signals (D) Investigation of miRNA-related concentration-dependent effects on the repression of cognate target bearing L7Ae-ECFP-associated transcript. Approximately 2 × $10^5$ HEK293T cells were transfected with 50 ng of a construct coding for the L7Ae-ECFP ± miR-122 target site transcript in the presence of increasing amounts of miRNA expression vector (0 - 100 ng), and 150 ng of iBlue transfection control for normalisation. A dose dependent, inverse relationship is demonstrated between the levels of miRNA and CFP fluorescence in cells harbouring pcDNA-ECFP-L7Ae-122T (n = 3, error bars represent standard deviation).*

constructs lack miR-122 target sites they would be expected to provide information independent of miRNA-dependent regulation regardless of whether investigated cells were co-transfected with pEM-122 or pEM-empty. An approximately 40 % reduction in CFP fluorescence was observed from cells expressing ECFP-dL7Ae (Fig.3.5C) as compared to those expressing ECFP-L7Ae (Fig. 3.5B), based on transfection with either pcDNA-ECFP-dL7Ae or pcDNA-ECFP-L7Ae, respectively. Assuming a comparable rate of translation for both fusion proteins, this observation would support a PEST-mediated increase in protein decay rate and thus lower steady state levels of ECFP-dL7Ae as compared to those of ECFP-L7Ae in the matching transfected cells. Next, the effect of this destabilisation on the performance of the first NOT gate was considered. CFP fluorescence in HEK293T cells transfected with pcDNA-ECFP-dL7Ae-122T was reduced in a miR-122-dependent manner (Fig. 3.5C), in a similar manner to that observed in the equivalent experiment using the stable L7Ae variant (Fig. 3.5B; note different Y-axis scales for Fig. 3.5, B and C). Based on dL7Ae levels as assessed by the ratio of CFP fluorescence in the absence compared to the presence of miR-122 the performance of the dL7Ae-containing NOT gate was very similar to that of the NOT gate using the more stable regulator (11.5 fold vs 10.9 fold, respectively). Given that use of the destabalised variant led to a highly comparable performance to that attained with the stable L7Ae variant all subsequent experiments were performed using the latter as it was reasoned that its greater absolute CFP fluorescence levels would allow for output signals that were better discriminated from those of background or artifactual noise, and potentially allow for more accurate measurements of ON:OFF ratio changes.

Finally, the sensitivity of L7Ae expression to miRNA-mediated repression was determined across a range of increasing miR-122 amounts (Fig. 3.5D), as predicated based on transfection with an increasing dose of the encoding plasmid. Transfection with 1 ng or less of pEM-122 was insufficient to elicit any repression of CFP fluorescence from an ECFP-L7Ae-122T construct in HEK293T cells. By contrast, transfection with either 10 ng or 100 ng of pEM-122 led to a strong dose-dependent decrease in CFP fluorescence. Intriguingly, in the absence of miR-122, cells harboring an ECFP-L7Ae construct lacking the miR-122 binding sites displayed approximately 60 % higher levels of CFP fluorescence as compared to the target containing ECFP-L7Ae-122T (Fig. 3.5B). These two constructs are identical, apart from a 97 bp 3' UTR insertion encoding the two perfectly complementary miR-122 target sites and a short intervening spacer sequence. The differential CFP fluorescence observed may have resulted from unintended regulatory motifs within this insertion. Indeed, it is possible that repression by endogenous miRNAs within HEK293T cells may account for this difference given the degeneracy permitted between an endogenous miRNA and its target sites, in which an interaction may be mediated by as few as 6 complementary bp within the seed region. A putative tar-

get site search using the PITA algorithm [168] identified 16 predicted targets sites (including the two miR-122 target sites that were intentionally inserted) corresponding to 12 miRNA species within the 3' UTR of the ECFP-L7Ae-122T construct as compared to only four target sites corresponding to 3 miRNA species but not miR-122 in the 3' UTR of the ECFP-L7Ae construct, lending weight to this hypothesis (Fig. 3.6).

### 3.2.3 Exploitation of the L7Ae/Kink-turn RNA motif interaction allows for highly regulated repression of the output signal from the second NOT gate

The inhibitory relationship between the regulator and the output modules represents the second NOT gate and a potentially additionally tunable parameter of this circuit (Fig. 3.7A). Published work characterising the positional effect of the Kt motif within the 5' UTR of reporter transcripts revealed that maximal translational repression could be achieved by placing this structure up to 50 nt downstream of the 5' terminus [169]. This positional effect was proposed to be dependent on the L7Ae/Kt motif interaction itself rather than the associated alteration in 5' UTR length due the demonstration that the corresponding control reporter constructs harboring a defective Kt motif that could not bind L7Ae were translated at equally high efficiency independent of 5' UTR length [169].

| 3' UTR of pcDNA-ECFP-L7Ae | | |
|---|---|---|
| microRNA | Sites | Score |
| hsa-miR-602 | 2 | -14.47 |
| hsa-miR-887 | 1 | -10.58 |
| hsa-miR-654-3p | 1 | -10.53 |

| 3' UTR of pcDNA-ECFP-L7Ae-122T | | |
|---|---|---|
| microRNA | Sites | Score |
| hsa-miR-122 | 2 | -35.81 |
| hsa-miR-574-5p | 2 | -20.73 |
| hsa-miR-639 | 1 | -19.26 |
| hsa-miR-432 | 2 | -16.28 |
| hsa-miR-887 | 1 | -13.92 |
| hsa-miR-608 | 1 | -13.45 |
| hsa-miR-328 | 1 | -12.64 |
| hsa-miR-193b | 1 | -12.38 |
| hsa-miR-182 | 2 | -11.38 |
| hsa-miR-1250 | 1 | -11.36 |
| hsa-miR-503 | 1 | -10.51 |
| hsa-miR-885-3p | 1 | -10.21 |

Figure 3.6: Prediction of 3' UTR-located miRNA target sites within the relevant regions of the ECFP-L7Ae ± 122T constructs
*The PITA algorithm [168] was used to predict miRNA target sites within the 3' UTR of the L7Ae-associated transcripts of pcDNA-ECFP-L7Ae (A) and pcDNA-ECFP-L7Ae-122T (B). Only sites corresponding to miRNA species whose provisional putative target site(s) produced composite scores below -10 are shown as these are more likely to be functional. The miR-122 target sites which were cloned into the 3' UTR of pcDNA-ECFP-L7Ae-122T yielded the highest miRNA-associated score of -35.81, further supporting their likely functionality (B). Apart from the miR-122 target sites, pcDNA-ECFP-L7Ae-122T contained many more sites meeting the selected cut-off threshold which typically and frequently (7 of the 16 predicted hits) scored more highly than any of the four sites found within the corresponding region of pcDNA-ECFP-L7Ae, thus potentially explaining the observed differences in CFP florescence in the absence of transgenically-supplied miR-122.*

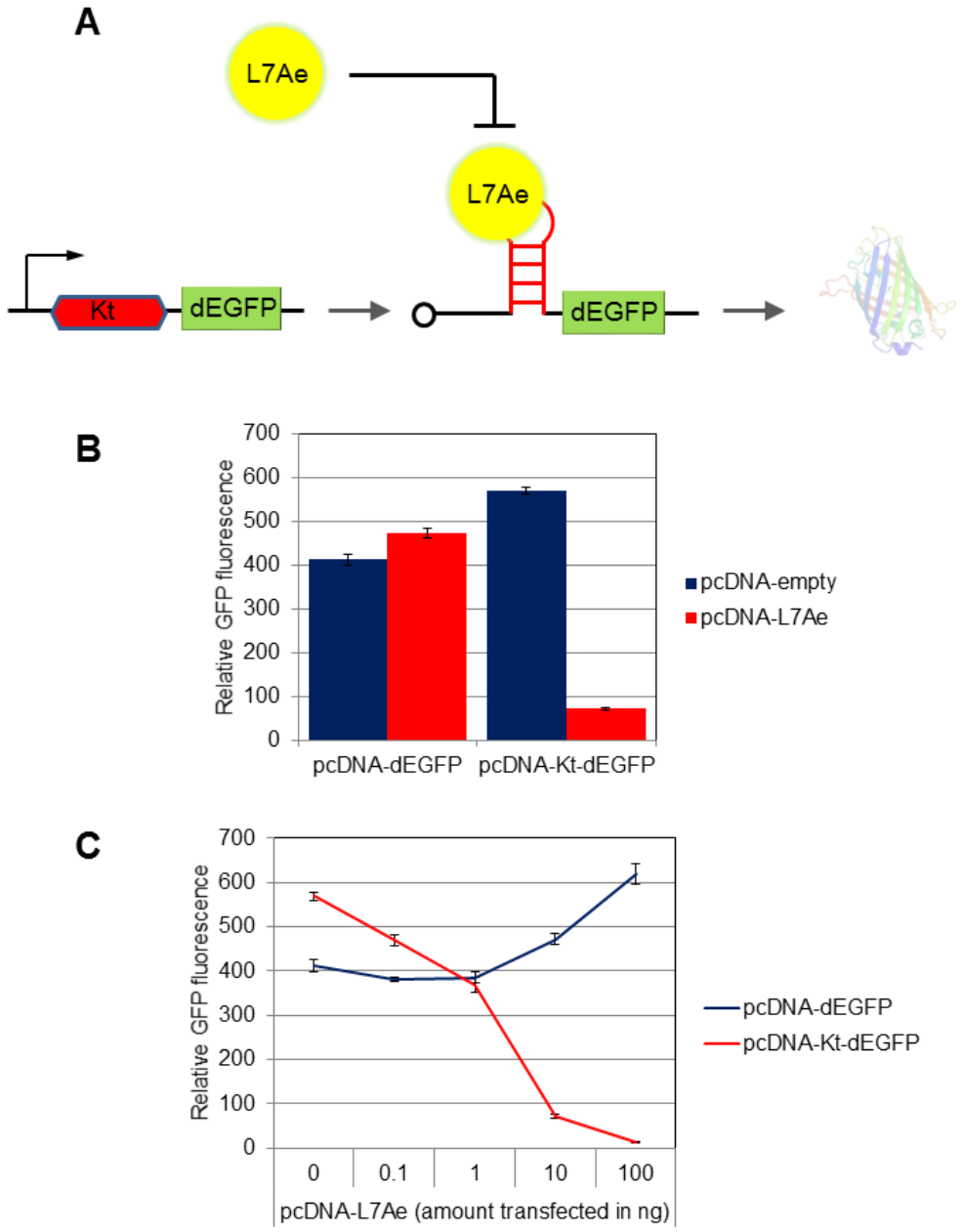



Figure 3.7: Investigation of L7Ae mediated translational inhibition

*(A) Schematic diagram illustrating the dEGFP reporter construct engineered to contain a 5' UTR Kt motif thus rendering it sensitive to translational inhibition upon L7Ae binding. (B) Relative GFP fluorescence was used as a readout of translation from Kt motif-containing reporters in the absence or presence of L7Ae. Approximately 2 × $10^5$ HEK293T cells were transfected with 50 ng of the dEGFP reporter ± Kt motif construct, 100 ng of the empty or L7Ae coding pcDNA expression vector, and 150 ng of the piBlue-N1 transfection control. A specific L7Ae-dependent inhibition of GFP fluorescence from the Kt-containing transcript was observed (n = 3, error bars represent standard deviation). (C) Investigation of L7Ae-associated concentration-dependent effects on the repression of the Kt-bearing dEGFP reporter construct. Continued...*

On this basis, our constitutive dEGFP expression vector (pcDNA-dEGFP) was reengineered to construct pcDNA-Kt-dEGFP to confer responsiveness to L7Ae via the insertion of a Kt motif within the 5' UTR of the dEGFP gene, 29 nt downstream of its 5' terminus. Relative GFP fluorescence in cells transfected with pcDNA-Kt-dEGFP was reduced by approximately 85 % when co-transfected with an L7Ae expression vector (pcDNA-L7Ae) compared to an empty expression vector (pcDNA-empty) (Fig.3.7B). No such L7Ae-mediated inhibition of GFP fluorescence was observed in cells transfected with pcDNA-dEGFP, thus supporting a repressive mechanism dependent on the Kt/L7Ae interaction (Fig. 3.7B). In the absence of L7Ae, GFP fluorescence was slightly increased in cells transfected with pcDNA-Kt-dEGFP compared to pcDNA-dEGFP strongly supporting the hypothesis that the secondary structure of the Kt motif alone is of insufficient stability to exert a negative effect on the translational efficiency of the transcript in which it is contained (Fig.3.7B).

In order to further study the relationship between L7Ae and its Kt-dEGFP associated target transcript, a titration experiment was performed. Co-transfection of HEK293T cells with 50 ng of pcDNA-Kt-dEGFP reporter and a 10-fold dilution series of pcDNA-L7Ae ranging from 0.1 – 100 ng, revealed a strong, inverse, dose-dependent, relationship between GFP fluorescence and L7Ae levels (Fig. 3.7C). Unexpectedly, expression of dEGFP from the transcript lacking the 5' UTR Kt motif appeared to be enhanced in the presence of L7Ae, in a positive dose-dependent manner but only when amounts in excess of 1 ng of pcDNA-L7Ae were transfected (Fig. 3.7C). Orthologues of L7Ae that recognise and bind to similar Kink-turn RNA structural motifs exist in mammalian cells. Despite its archaeal origin, it is possible that the L7Ae protein used in our experiments retains some affinity for mammalian Kt motifs. At very high levels of expression achieved when transfecting the largest quantities of L7Ae expression vector, L7Ae may begin to occupy some of these lower affinity target sites exacting changes within the cell which may globally promote translation, leading to the observed Kt-independent stimulation of dEGFP expression. Alternatively, it is possible that the higher L7Ae levels investigated may have led to impaired expression of the iBlue protein (transfection control), thus resulting in artifactually elevated relative GFP fluorescence at the higher pcDNA-L7Ae transfection doses. If so, this would also suggest that the repressive effect on dEGFP expression for the Kt-bearing construct was even greater than that suggested by the relative GFP fluorescence data shown in Fig. 3.7C.

*Figure 3.7 contiued: Approximately 2 × $10^5$ HEK293T cells were transfected with 50 ng of the dEGFP reporter ± Kt motif plasmid in the presence of increasing amounts of the L7Ae-encoding plasmid (0 - 100 ng), and 150 ng piBlue-N1. These data demonstrated a dose-dependent, inverse relationship between the levels of L7Ae and the effect of this protein on the Kt-dEGFP reporter readout (n = 3, error bars represent standard deviation).*

### 3.2.4 The assembled double-negative gate circuit exhibits a markedly reduced ON:OFF ratio than was expected based on the performance of its component NOT gates

The complete double-negative gate miRNA positive sensor containing the L7Ae based sensor-regulator module and the Kt-regulated dEGFP output, along with its cognate input signal, was encoded on three plasmids (pcDNA-L7Ae-122T, pcDNA-Kt-dEGFP and pEM-122, respectively) and transfected simultaneously into HEK293T cells. Upon assembly of the complete circuit, it was important to explore the limits of output expression under artificially constrained ON and OFF states. The ratio of GFP fluorescence in these states (ON:OFF ratio) can be used as a measure of circuit performance. Given its central position within the circuit, the regulator module was manipulated to force either an ON or OFF state.

A forced ON state was achieved by replacing the L7Ae-encoding expression vector with an empty vector (pcDNA-empty). Without the L7Ae regulator, dEGFP expression from the pcDNA-Kt-dEGFP vector would be expected to exhibit maximal uninhibited levels of fluorescence. Irrespective of the presence or absence of the specific input miRNA, GFP fluorescence was high under this enforced ON state (Fig. 3.8A). An approximately 18 % reduction in GFP fluorescence was observed in the presence of miR-122 (pEM-122) compared to the control (pEM-empty), an effect assumed to be indirect and/or non-specific due to the absence of cognate miR-122 target sites on either the empty regulator transcript (pcDNA-empty) or the dEGFP-encoding output transcript (pcDNA-Kt-dEGFP) (Fig. 3.8A). The OFF state was enforced upon the circuit by transfecting with pcDNA-L7Ae which coded for the L7Ae regulator but lacked miR-122 target sites and was therefore predicted to be insensitive to this entity. Thus, L7Ae is expected to accumulate within cells transfected with pcDNA-L7Ae and consequently exert maximal repression of the Kt-dEGFP transcript in the presence of either a control miRNA or miR-122 itself (Fig. 3.8A). Based on the observed differences in GFP fluorescence between these artificially forced ON and OFF states it would appear that the maximal ON:OFF dynamic range that can be achieved with this circuit would be about 10 - 14-fold.

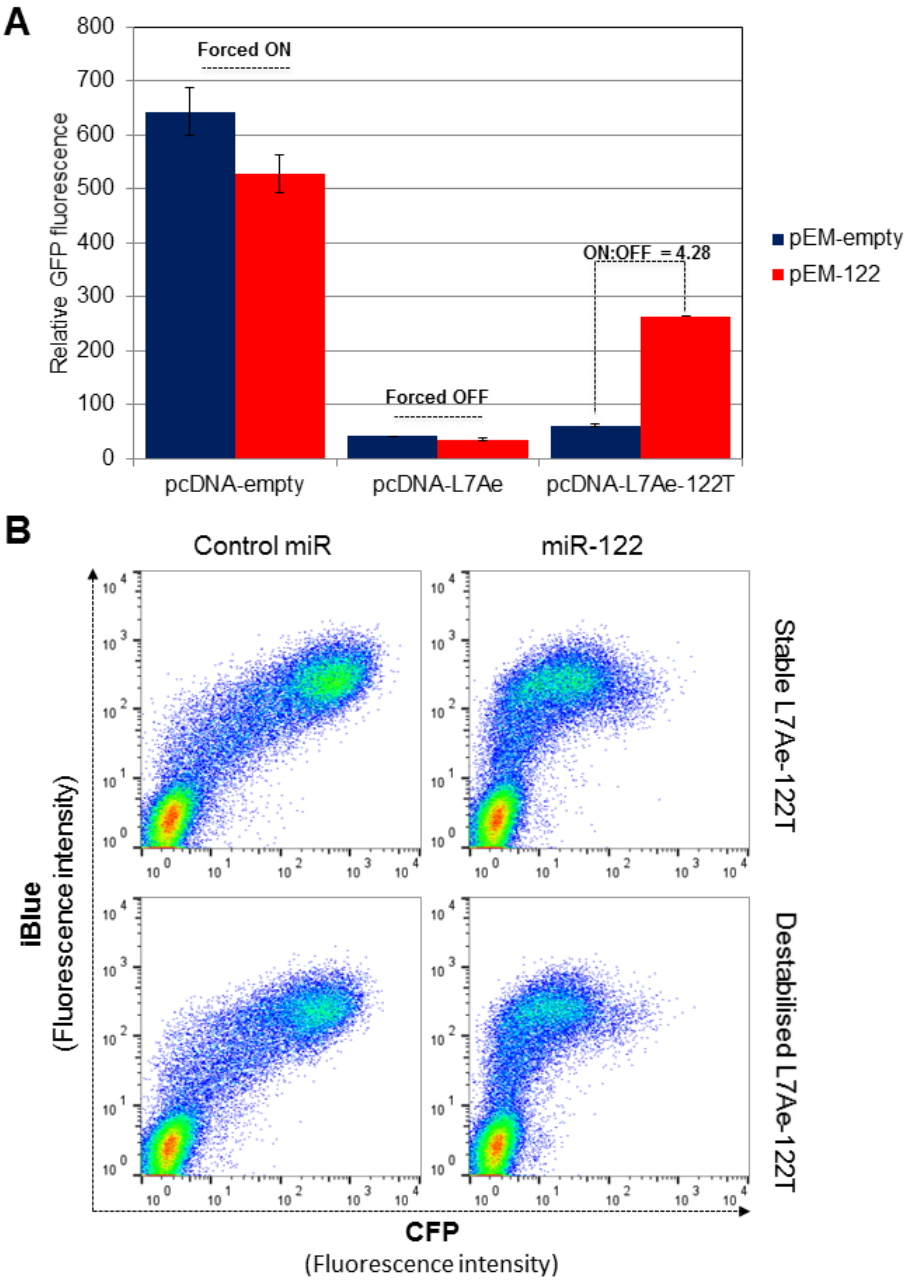



Figure 3.8: Investigation of the function of the fully assembled miRNA positive sensor circuit *(A) Approximately 1 × $10^5$ HEK293T cells were transfected with 450 ng of the empty or miR-122-encoding pEM vector, 50 ng of the corresponding empty or L7Ae ± 122T pcDNA expression vector, 150 ng pcDNA-Kt-dEGFP, and 250 ng the piBlue-N1 transfection control. Independent of the presence or absence exogenous miR-122, GFP fluorescence was maximal in the absence of the L7Ae regulator protein (forced ON state) and minimal when repressed by the miR-122-insensitive L7Ae-encoding construct (pcDNA-L7Ae; forced OFF state). The complete circuit, containing the L7Ae-122T regulator construct (pcDNA-L7Ae-122T), demonstrated a 4.28-fold miR-122-dependent increase in GFP fluorescence (n = 3, error bars represent standard deviation). Continued...*

Finally, a functional sensor circuit was transfected into HEK293T cells by introducing the L7Ae-122T miR-122-sensitive regulator construct and the Kt-dependent dEGFP output plasmid, and interrogated using either a control miRNA or miR-122 expressed by matching co-transcribed plasmids. In the presence of the cognate miR-122, circuit output as measured by GFP fluorescence was approximately 4.28 fold higher than in the presence of a control miRNA, thus demonstrating a functional positive miRNA sensor circuit (Fig. 3.8A). In the absence of the cognate miRNA, conditions under which the circuit should be OFF, the levels of GFP fluorescence approached that of the control circuit which was constitutively forced into the OFF state. However, in the presence of the cognate miR-122 the complete circuit only exhibited an ON-level of GFP fluorescence that was approximately half of that observed for the constitutively ON control circuit. This latter finding strongly suggests an inability of miR-122, at least at the quantities studied, to completely inhibit expression and/or activity of the L7Ae regulator, which in turn leads to a degree of residual repression of the Kt-responsive dEGFP output. In conclusion, the ON:OFF ratio of the fully assembled miRNA sensor circuit fell short of its theoretical optimal behavior as defined by the forced-ON:forced-OFF ratios observed using the constrained circuits (Fig. 3.8A).

Further analysis of the previously utilised construct containing ECFP-tagged L7Ae provided additional evidence for the incomplete repression of this regulator by miR-122. This was revealed through the transfection of a miR-122 targetable ECFP-L7Ae (pcDNA-ECFP-L7Ae-122T) construct with either a control miRNA or miR-122 (Fig. 3.8B; iBlue served as a transfection control). Under all conditions studied, a well-defined non-transfected population of cells existed as shown by data points in the lower left corner of the plots which correspond to cells that do not express either the ECFP or iBlue fluorescent proteins. A second, more dispersed population, corresponding to transfected cells, exhibited broadly similar high levels of iBlue fluorescence (transfection control) across the four scenarios examined but more variable CFP fluorescence, consistent with differing levels of ECFP-L7Ae inhibition. Cells transfected with either miR-122-targetable stable or destabilised ECFP-L7Ae constructs exhibited greater than 10-fold reduction in CFP fluorescence in the presence of the cognate miR-122 compared to the control miRNA (Fig. 3.8B, note leftward shift of data points corresponding to

*Figure 3.8 contiued: (B) Flow cytometry plots charting CFP fluorescence driven by the dL7Ae-ECFP ± 122T construct against iBlue fluorescence (transfection control). Approximately 2 × $10^5$ HEK293T cells were transfected with 50 ng of the dL7Ae-ECFP ± 122T construct, 100 ng of either the empty vector control or the matched miR-122-encoding plasmid, and 150 ng of piBlue-N1. These data clearly demonstrate that CFP readouts from the miR-122-responsive constructs coding for both the stable L7Ae protein or its destabilised variant (dL7Ae) are only partially inhibited in the presence of miR-122.*

transfected cells in panels on the right (miR-122) as compared to those on the left (control miRNA)). Despite the observed miR-122-mediated reduction in CFP fluorescence, the vast majority of the population of transfected cells (high iBlue fluorescence) that were subjected to miR-122-mediated inhibition still exhibited CFP fluorescence levels that were greater than the population of non-transfected cells (low iBlue fluorescence), suggesting moderate residual levels of L7Ae even under these conditions. The circuit containing the destabilised L7Ae is still compromised by incomplete miR-122-mediated repression, although to a significantly lesser extent than that based on the stable L7Ae protein (Fig. 3.8B). Indeed, this latter point strongly supports the above postulate but also emphasises the challenges of building a protein regulator based switch that would allow for cleaner ON:OFF interchanges.

### 3.2.5 Simultaneous positive and negative circuit outputs

By employing the ECFP-L7Ae-122T construct as regulator module for a complete miRNA sensor circuit, flow cytometry can be used to simultaneously measure the output from both constituent NOT gates in the form of CFP and GFP fluorescence signals, respectively. In this instance, the output of the first NOT gate (expression level of the regulator protein) will result in CFP fluorescence that negatively correlates with the level of the target miRNA, whilst the output from the second coupled NOT gate, results in variable GFP fluorescence that positively correlates with the activity of the same miRNA. The circuit OFF state (absence of input miRNA) would be characterised by high CFP and low GFP fluorescence and *vice versa* for the circuit ON state (presence of cognate input miRNA). The caveat in relation to the above is the potential for significant cross-talk in the detection of CFP and GFP fluorescence signals by flow cytometry although partial correction with compensation is possible [162].

The above circuit was examined through the co-transfection of HEK293T cells with constructs coding for a control miRNA or miR-122, the ECFP-L7Ae or ECFP-L7Ae-122T regulator module, and the Kt-dependent dEGFP output expression module. Cells transfected with the miRNA-insensitive ECFP-L7Ae regulator (pcDNA-ECFP-L7Ae) are forced into the OFF state irrespective of which miRNA they express. A flow cytometry scatter plot charting GFP against CFP fluorescence, reveals that under the conditions mentioned above, the bulk of the cell population displays a phenotype associated with relatively high levels of CFP fluorescence and sub-maximal levels of GFP fluorescence, as would be expected in the OFF state (Fig. 3.9A, top two panels). Likewise and as expected cells transfected with both the miRNA-sensitive ECFP-L7Ae-122T regulator and Kt-dependent dEGFP output constructs but an empty miRNA expression vector exhibited the OFF state circuit behavior and the expected fluorescence pattern (Fig. 3.9A, bottom left panel). Only when cells were co-transfected with the fully functional miRNA-sensitive regulator and the Kt-dependent output module, in addi-

tion to the construct expressing the cognate input miR-122, was an ON state induced within cells. This can be observed from a shift of the majority of cells as these become "repositioned" to extend horizontally within a zone adjacent to the X-axis (Fig. 3.9A, lower right panel). This reflects a simultaneous reduction in CFP fluorescence levels and an increase in GFP fluorescence levels that is predicted to accompany the repression of ECFP-L7Ae and consequent de-repression of the dEGFP output protein.

Whilst the CFP output from the first NOT gate undergoes more than 10-fold change in signal intensity in the presence of miR-122 compared to the empty vector control, the dEGFP output from the final NOT gate only demonstrates an approximately 2-fold ON:OFF ratio (Fig. 3.9, B and C). The magnitude of the miRNA-mediated repression of ECFP-L7Ae as measured by the reduction in CFP fluorescence is consistent with our earlier optimisation of the first NOT gate in isolation (Fig. 3.5B). However, the 2-fold observed ON:OFF ratio from the final dEGFP output of the complete double-negative gate circuit was distinctly reduced from the 4.28 fold ON:OFF ratio achieved with the un-tagged L7Ae regulator module (Fig. 3.8A). This was most probably the result of cross-talk between the GFP and CFP fluorescence emission spectra, which despite compensation could not be fully eliminated. Future implementations of such a circuit would ideally make use of different fluorescent proteins characterised by more widely separated emission spectra in an attempt to avoid this interference.

## 3.3 Discussion

### 3.3.1 Persisting Kt-interaction-related inhibitory levels of L7Ae under conditions of miR-122-mediated repression contribute to the observed suboptimal ON:OFF ratio

Although the ON:OFF ratio of the complete double-negative gate circuit was approximately five, the performance of its constituent NOT gates in isolation was far greater. This may be the result of indirect linkage between the two components. The input miRNA signal is targeted against the mRNA transcript encoding L7Ae, yet it is the L7Ae protein that interacts with the output dEGFP-associated transcript. Therefore, the processes governing the steady state levels of L7Ae, comprising principally the rates of protein production and degradation, are a key determinant of the overall circuit ON:OFF ratio. The optimal situation would involve a very tight correlation between L7Ae transcript and protein levels, whilst also allowing for a large dynamic range and a Kt-interaction-related subinhibitory lower limit for L7Ae cellular levels. Among other less amenable factors, maximal rates of both L7Ae production and degradation would be key to achieving these goals. Attempts were made in this study to enhance the rate

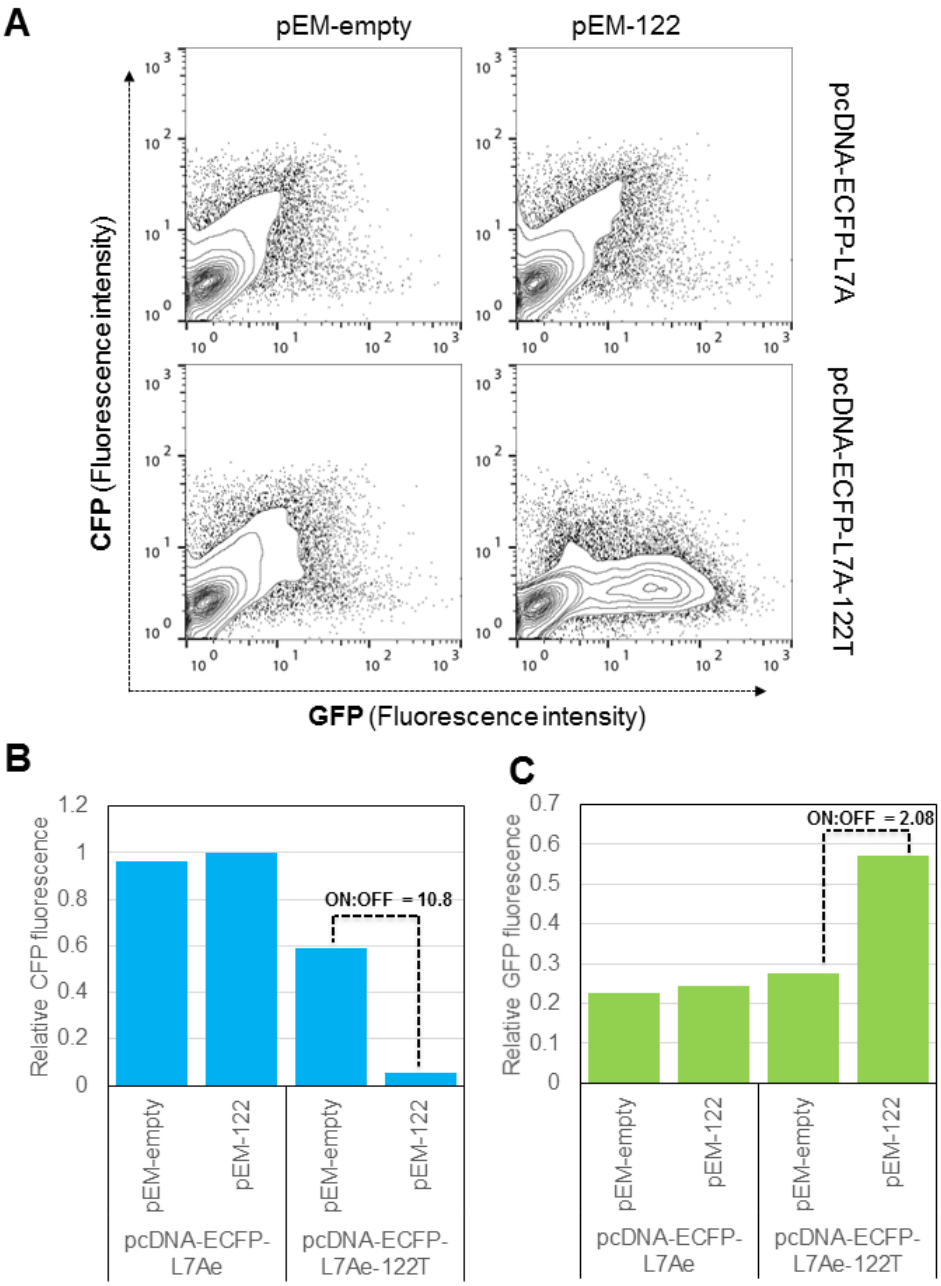



Figure 3.9: Simultaneous measurement of the individual output from each of the two constituent NOT gates in the miRNA positive sensor circuit
*(A) Compensated flow cytometry plots charting CFP fluorescence driven by the L7Ae-ECFP ± 122T regulator construct against GFP fluorescence arising from the Kt-dEGFP output construct. Approximately 2 × $10^5$ HEK293T cells were co-transfected with 150 ng of the empty or miR-122 encoding pEM vector, 250 ng of the L7Ae-ECFP ± 122T construct, 150 ng of pcDNA-Kt-dEGFP, and 250 ng the piBlue-N1 transfection control. The circuit ON state (lower right panel, L7Ae-122T + miR-122) was characterised by a shift to high GFP and low CFP fluorescence. (B) Quantification of relative CFP fluorescence from the above data. (C) Quantification of relative GFP fluorescence from the above data.*

of L7Ae degradation through the fusion of a destabilising PEST domain. However, despite evidence of a lower L7Ae steady state level when input miRNA was present for the destabilised as opposed to stable variant (Fig. 3.8B) this modification only resulted, in at best, a marginal though highly questionable enhancement of the ON:OFF ratio of the sensor-regulator module (first NOT gate) when studied in isolation (Fig. 3.5B; note that n = 1 in this experiment). However, even on the basis of the available data it would appear that significant dL7Ae levels were still present in HEK293T cells carrying the fully assembled double-negative gate sensor 48 h post-transfection of an input miRNA (miR-122) generating plasmid, suggesting that the solitary PEST domain introduced had not led to a sufficient reduction in the half-life of this protein.

Alongside alternative strategies to further destabilise L7Ae (see Section 3.3 for discussion of this point), future efforts should also be directed towards enhancement of the rate of L7Ae production as steady state levels are a function of the balance between production and decay. Additionally, if further reduction of the minimal L7Ae cellular concentration remains problematic, raising the upper cellular concentration so long as it is not deleterious to the cell and or in excess of the circuit-saturating level, would allow for a greater functional dynamic range in L7Ae levels and potentially yet-another opportunity to raise the ON:OFF ratio. Optimisations to the Kozak sequence and other 5' UTR elements of the L7Ae gene that may enhance its translational efficiency could also be investigated. Likewise, tuning of the rate of L7Ae-associated transcription could be explored by way of use of a tightly regulated inducible promoter capable of driving transcription at a set point of choice over a wide dynamic range; ideally spanning complete gene silencing to a transcription rate exceeding even that of the $P_{CMV}$ constitutive promoter used in this study. As with many complex systems it is not possible to predict whether a higher or lower level of L7Ae-associated transcription than that governed by $P_{CMV}$ would lead to an enhanced ON:OFF ratio of the complete circuit. The same is likely to apply to the other components of this circuit. Hence, the need for more comprehensive experimental exploration of transcriptional and/or translational levels (as appropriate) associated not just with L7Ae but also all other components of the complete circuit, including the input miRNA transcript itself. Furthermore, it may be beneficial to explore the time course over which L7Ae undergoes miRNA-mediated repression. For consistency, all the results described above originated from flow cytometry analysis of cells 48 h post-transfection. As noted, within this timeframe residual L7Ae is still observed following miRNA inhibition. By conducting similar experiments and analyzing the data at either 24 h or 72 h it may be possible to determine whether steady state levels of L7Ae have been reached by 48 hours. Once again wider investigations of this nature with regards to the other parts of this multicomponent system are likely to offer valuable insights and undoubtedly further pointers towards

optimisation.

### 3.3.2 Exploration of alternative regulator proteins and/or architectures comprising more complex orthogonal circuits

Besides tuning of the currently designed circuit, the option of investigating other circuit architectures could also be pursued.The identification of alternative regulator protein components in addition to L7Ae will enable the generation of more complex orthogonal circuits. Theoretically any exogenous RNA binding protein with a known short binding motif could be a potential candidate for additional regulators. The endoribonuclease Csy4 is a promising candidate for this purpose, as it is known to bind a short RNA motif with very high affinity [170]. A further advantage of Csy4 is that it is a nuclease, thus possessing the ability to repress target expression through mechanisms of both translational inhibition as well as mRNA cleavage followed by degradation. This activity can be further investigated through use of catalytically dead mutant versions of Csy4, which would function entirely through translational repression, much like L7Ae.

L7Ae, when present at high levels, appeared to have a nonspecific stimulatory effect on the translation of dEGFP from the Kt-motif-lacking control plasmid (pcDNA-dEGFP), though as noted earlier this could have been an artifact of an inhibitory effect on production of the transfection control comparator, iBlue. Nevertheless, given that L7Ae is a ribosomal protein and that the archaeal version used in these studies is a close orthologue of the mammalian protein, it is possible that overexpression of L7Ae is capable of substituting for the endogenous variant and thus globally promoting translation. This would be another justification for the exploration of alternative proteins and cognate DNA or RNA binding motifs that could function as regulator modules.

### 3.3.3 Future strategies for the further destabilisation of intermediate protein modules and associated characterisations

A key factor limiting the performance of the circuit constructed in this study is the half-life of the intermediary module, the regulator. This problem is exacerbated by the fact that the input signal is indirect. Although the input miRNA targets the regulator mRNA for efficient cleavage and degradation, it is the regulator protein which binds to and determines the output expression. Thus, L7Ae-mediated translational repression of the output may continue after miRNA-mediated silencing of the regulator-encoding mRNA transcript. The stability of the L7Ae regulator will determine the length of this delay in signal conduction through the circuit.

As described above, we have taken some preliminary steps towards tackling this problem by fusing the regulator protein to a known destabilising element (PEST domain of MODC) in order to accelerate its rate of degradation, however there is scope for further effort. These could include inserting multiple destabilising domains within L7Ae, investigation of alternative destabilising domains, use of various permutations of destabilising domains and even exploration of the optimum insertion site(s) of such motifs. Critically, such studies should be supplemented by application of assays (such as live cell imaging of ECFP-L7Ae containing cells) that permit precise quantification of cellular levels of L7Ae not only at a fixed time point but over timeframes of relevance.

In order to systematically evaluate the merits of the currently applied and future L7Ae regulator protein destabilisation strategies it would be very useful to utilise a system whereby the stability of L7Ae could be titrated. One such way in which this could be achieved is through ligand-regulated destabilisation domains. Well characterised examples of these have been generated through the screening of engineered mutants of the human FKBP12 protein and the discovery of unstable variants which underwent stabilisation in the presence of a small molecule ligand (Sheild-1[Shld1]) [171]. It was then shown that a representative of these destabilisation domains could be fused to other proteins of interest to confer upon them ligand-regulated stability. It has been reported that this system is capable of very tight regulation, with almost complete degradation of target protein in the absence of its stabilising ligand. This may be a particularly powerful system to incorporate into our double-negative gate sensor as we could theoretically obtain very fine control over the half-life of the regulator component using varying concentrations of the Shld-1 ligand.

# Chapter 4

# RT-PCR based selection of miRNA-responsive allosteric ribozymes

## 4.1 Introduction

Having shown in the previous chapter that it is possible to design a double-negative gate sensor for the detection of miRNA activity, we sought to simplify the system and create a single molecular module capable of performing the function of both NOT gates. This required the construction of a combined regulator and output module, in which the regulator could function in *cis* to repress expression of the output. Self-cleaving ribozymes are ideally suited for this purpose as they can be genetically encoded within the 3' UTR of protein coding genes. Consequently, they can function as regulators whereby their cleavage results in the degradation of the output transcript and therefore repression of output protein expression. To complete the circuit, the activity of this regulator module must be attenuated by a specific input signal. Theoretically, this can be achieved through the use of allosteric self-cleaving ribozymes, in which cleavage activity can be regulated by a specific ligand. Naturally occurring allosteric self-cleaving ribozymes which perform in this way have been previously described [33]. More recently this concept has been exploited by a number of laboratories leading to the generation of a wide range of synthetically engineered allosteric ribozymes, responsive to a diverse set of ligands [42, 43, 63, 64, 66, 151, 172–175].

We therefore sought to generate such ribozymes with a miRNA binding allosteric site for the construction of a miRNA-responsive ON-OFF genetic switch. In the absence of input miRNA, the circuit would be maintained in an OFF state as a result of ribozyme-mediated self-cleavage, thus promoting output transcript degradation and inhibition of output protein expression (Fig. 4.1A). Binding of a specific input miRNA to the allosteric site would result

in structural changes that are transmitted to the ribozyme catalytic core, resulting in inhibition of self-cleavage activity. This woud lead to the stabilisation of the entire output transcript and a transition into the ON state in which the expression of output protein is permitted (Fig. 4.1B). This strategy would enable the creation of a direct intramolecular connection between the regulator and output components of the proposed circuit, thus tightly linking the two NOT gates (Fig. 4.2B). This is conceptually distinct from the circuit described in the previous chapter where the output of the first NOT gate (L7Ae regulator mRNA) is separated from the input to the second NOT gate (L7Ae protein) by the process of translation (Fig.4.2A). The aims of re-engineering the connection between the two NOT gates were to minimise the discrepancy that had been observed between the ON:OFF ratios of the two individual NOT gates compared to that of the the full circuit, and to potentially allow for a more rapid transition from the OFF-to-ON state and *vice versa* as driven by the timing of the initiation or withdrawal of the input signal, respectively (Chapter 3).

To inform our rational design process, we scrutinised the features of the published examples of short oligonucleotide-responsive, allosteric, self-cleaving ribozymes. The first such ribozyme, termed N79, was reported by Yen *et al.* [63] in 2004. N79 was designed to respond to a modified oligonucleotide complementary to segments of stems I and III of the hammerhead ribozyme. Upon oligonucleotide binding conformational changes to the ribozyme secondary structure were predicted to have occurred that in turn almost certainly accounted for the observed inhibition of its catalytic activity. This seminal study highlighted the potential to engineer oligonucleotide-modulated ribozyme switches, but was limited to a single defined cognate input oligonucleotide sequence. Subsequently, a computational method was developed to facilitate rational design of hammerhead ribozymes capable, with a high probability, of responding to ~22 nt oligonucleotide triggers regardless of their sequence [64]. The oligonucleotide binding site (OBS) of such ribozymes is located within stem III, which upon binding to its cognate ligand transmits structural changes through a short linker to the rest of the ribozyme inhibiting its catalytic activity.

More recently, *in vitro* selection techniques have been effectively utilised for *de novo* design of allosteric hammerhead ribozymes [176]. Such methods exploit the direct phenotypic-genotypic link between functional RNA devices. Typically, a large starting ribozyme pool is generated through the fusion of a ligand binding aptamer domain to a ribozyme, via a short randomised linker sequence. Subsequent selection enables the isolation of ribozyme variants that undergo allosteric modulation in response to a specified cognate input ligand. Unlike SELEX, this process consists of repetitive cycles of both positive and negative selection. In order to select for either allosterically activated or inhibited ribozymes the ligand is included during the positive or negative selection cycles (see below for clarification), respectively. This

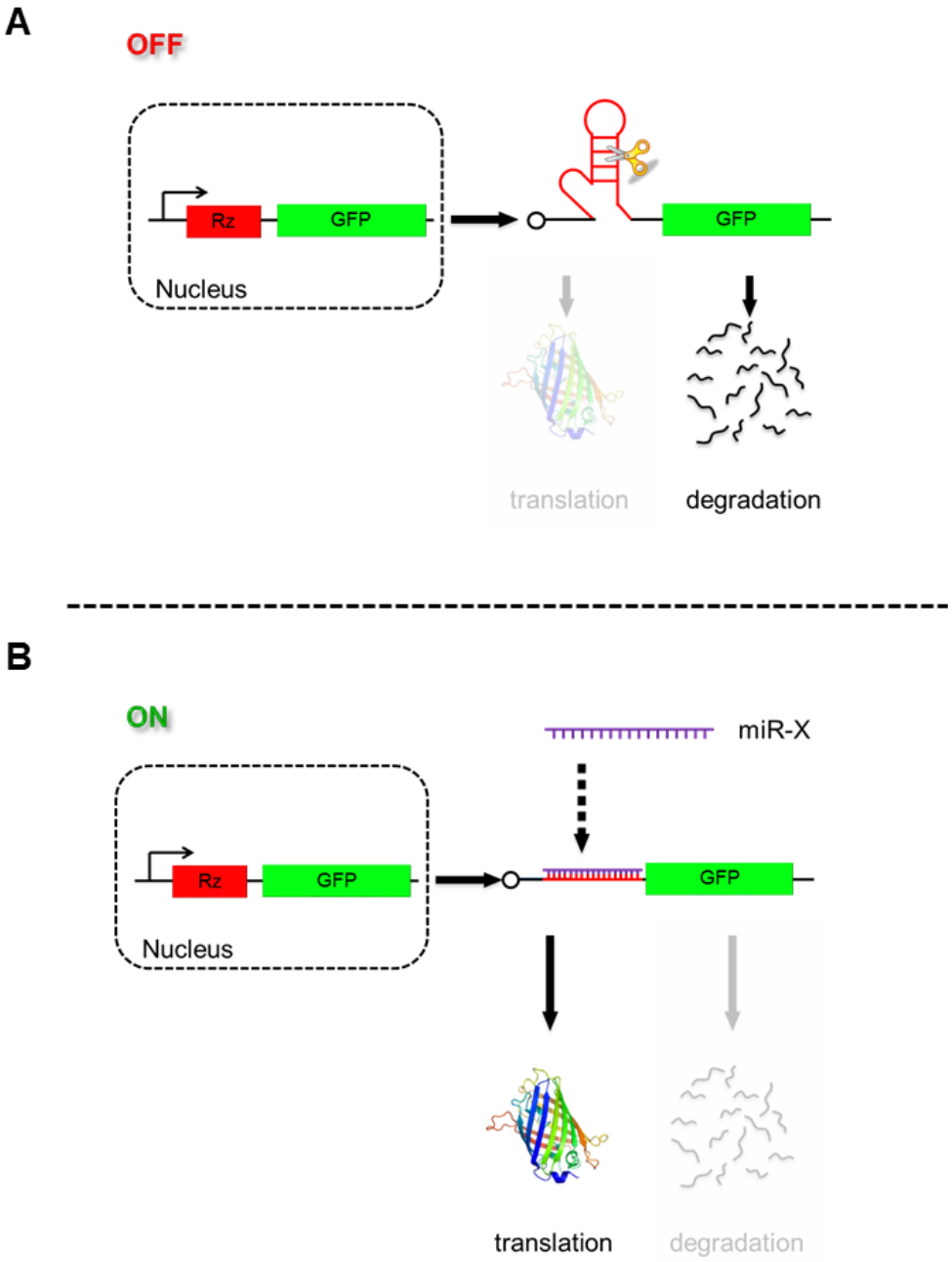



Figure 4.1: Schematic diagram of a ribozyme-regulated sensor construct
*This circuit functions in a similar way to those described in Chapter 3 however here the regulator (ribozyme) has been incorporated into the output transcript so as to reduce the complexity of the circuit and more directly couple the constituent components. (A) A gene encoding the fused regulator/output mRNA is continuously transcribed from a constitutive promoter ($P_{CMV}$). Once transcribed, the ribozyme can fold into an active conformation upon which it will catalyse its own cleavage, leading to degradation of the entire transcript and the inhibition of output protein expression (circuit OFF state). (B) Following binding of a cognate miRNA to the allosteric site within the ribozyme, the propagation of structural changes will lead to inhibition of self-cleavage, thus stabilising the transcript and permitting its translation and the production of output protein (circuit ON state).*

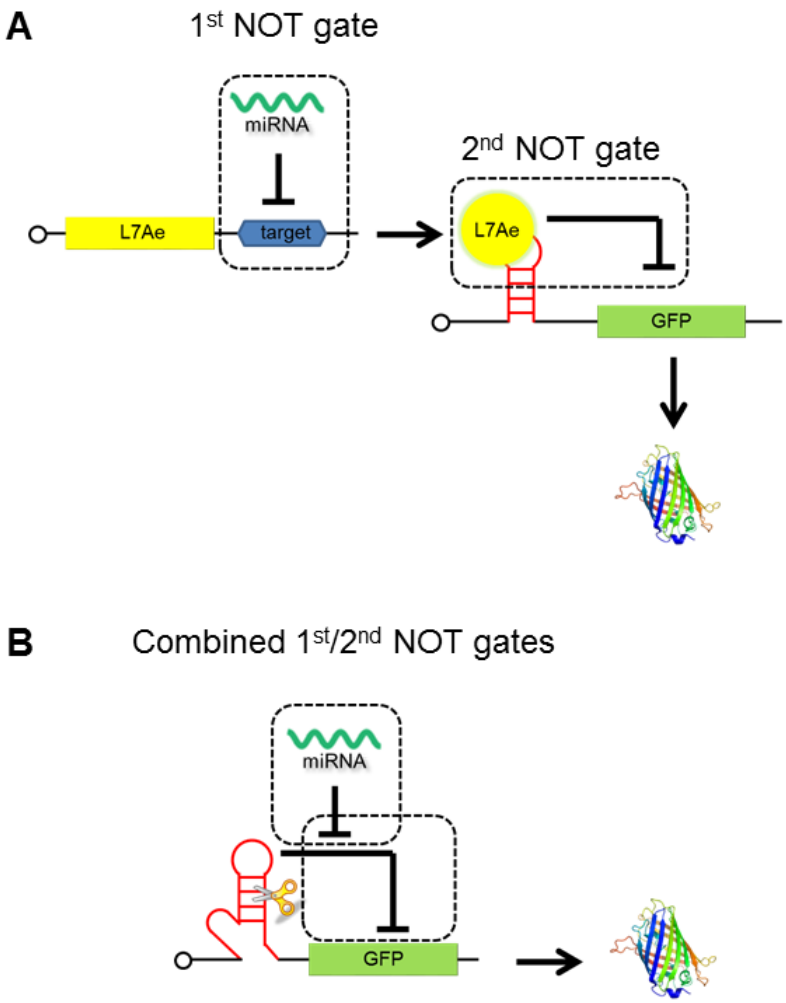



Figure 4.2: An allosteric ribozyme makes for a more direct and condensed circuit
*(A) The double-negative gate circuits constructed in Chapter 3 consisted of two discrete NOT gates; miRNA-mediated inhibition of expression of L7Ae, and L7Ae-mediated inhibition of expression of the output (GFP) gene. Inefficiencies in the linkage between the two NOT gates may have contributed to the reduction in the overall circuit ON:OFF ratio as opposed to the potential optimum based on the ON:OFF ratios of the individual NOT gates. (B) The circuit architecture exploited in this chapter fuses the functionality of both NOT gates into a single RNA molecule; miRNA-mediated inhibition of ribozyme cleavage activity, and ribozyme-mediated inhibition of output (GFP) expression. This should couple their performance in a tighter manner and would be hypothesised to enhance the ON:OFF ratio of the circuit.*

process may be conducted entirely *in vitro* [177], using a combination of *in vitro* and *in vivo* steps [150], or exclusively within living cells [60]. Notably, high throughput sequencing has enabled high-resolution interrogation of this selection process through the direct measurement of each ribozyme variant throughout the entire multicycle procedure.

The following chapter describes the strategies employed as part of this study to select for a miRNA-responsive allosteric ribozyme. First, the activities of NOT gate ribozymes designed by Penchovsky and Breaker were tested and optimised in mammalian cells in culture [64]. Next, these NOT gate ribozymes were used as a framework for building a pool of putative miRNA-responsive allosteric ribozymes. This comprised the modification of stem III of the hammerhead ribozyme to accommodate both a miR-21 complementary OBS and a randomised linker sequence. This ribozyme pool was then subjected to a mixed selection protocol, consisting of *in vitro* and *in vivo* steps. Using this strategy, the critical step during which ribozyme cleavage may occur is carried out within cells, thus maximising the chances that evolved allosteric ribozymes identified by the selection process would function within a cellular environment. High-throughput sequencing was used as a parallel and complementary strategy to monitor the process of selection. Finally, candidate miRNA-responsive allosteric ribozymes were cloned into the UTR of a reporter gene and transfected into cells, in the presence or absence of their cognate miRNA, to assess their utility as miRNA sensors.

## 4.2 Results

### 4.2.1 The wild-type NOT-1 hammerhead ribozyme that is proposed as the basis of NOT gate-type oligonucleotide-responsive allosteric ribozymes functions within the eukaryotic cellular environment

The catalytic activities of ribozymes are highly dependent on their secondary structures, which are influenced by conditions such as the chemical environment and temperature. This presents a significant hurdle when migrating *in vitro* designed ribozymes to an *in vivo* cellular context. The extended hammerhead ribozyme upon which NOT gate ribozymes designed using the Penchovsky algorithm are built, has a higher catalytic rate and a lower $Mg^{2+}$ ion requirement compared to those of the minimal hammerhead ribozyme [64]. These features are a result of tertiary interactions between a bulge in stem I and the loop of stem II [70]. Such NOT gate ribozymes are therefore attractive candidates for *in vivo* use. In order to validate the use of these NOT gate ribozymes as a conceptual framework for the design of our randomised pool it was first necessary to demonstrate their functionality within mammalian cells.

To test this, we have cloned the wild type NOT-1 sequence [64] into the 5' UTR of an EGFP

reporter within the LentiLox 3.7 [156] expression vector to produce pLL-NOT-1-EGFP. This ribozyme insertion location was chosen based on previous positional optimisation by Yen *et al*. [63], who demonstrated that ribozymes located within 5'UTRs elicited the strongest inhibitory effect on protein expression. To control for any cleavage-independent inhibition that may occur as a result of the complex secondary structure present in the 5' UTR, a mutant ribozyme (NOT-1mut) was designed to fold into a dominant innactive conformation that rendered the ribozyme catalytically inactive. NOT-1mut was generated and positioned identically upstream of the reporter EGFP gene to produce the control plasmid pLL-NOT-1mut-EGFP. Transfection of HEK293T cells with the matching EGFP expression vector (pLL-EGFP), containing an unstructured 5' UTR, led to efficient production of EGFP protein as observed based on the strong signal produced under fluorescence microscopy (Fig. 4.3, top left panel). As predicted, the expression of EGFP protein was efficiently inhibited by the 5' UTR-located NOT-1 ribozyme (Fig. 4.3, middle left panel), but not by the structurally equivalent NOT-1mut control ribozyme (Fig. 4.3, bottom left panel), which was associated with a level of fluorescence comparable to that of the ribozyme-lacking construct, pLL-EGFP. As a further control, it is apparent that similar numbers of cells were analysed under each transfection condition, as evidenced by DAPI staining (Fig. 4.3, right panels).

### 4.2.2 Introduction of two ribozyme-independent modifications into the sensor construct led to a marked enhancement of the ON:OFF ratio

Having demonstrated that the NOT-1 ribozyme can be used to inhibit transcript-associated protein expression within mammalian cells we attempted to further engineer the regulator/output transcript to enhance its potential ON:OFF ratio. In addition to ribozyme cleavage, a number of additional factors may act to influence the degree to which output protein expression is inhibited. By altering these ribozyme-independent factors, we aimed to identify the optimal genetic environment within which to embed the randomised sequence-bearing ribozyme pool for the subsequent selection experiments.

One contributory factor is the stability of the output EGFP protein, which can decouple the ON:OFF ratio at the level of ribozyme self-cleavage to its impact on GFP fluorescence. The wild type EGFP protein has been shown to have a half-life of approximately 26 h [178]. This introduces a lag period between changes in the level of EGFP transcript and consequent downstream alterations in the level of accumulated protein. Amplification emerging from the ability of a single mRNA transcript to direct the translation of many protein copies can further diminish the apparent ON:OFF ratio as protein expression reaches levels of saturation. To

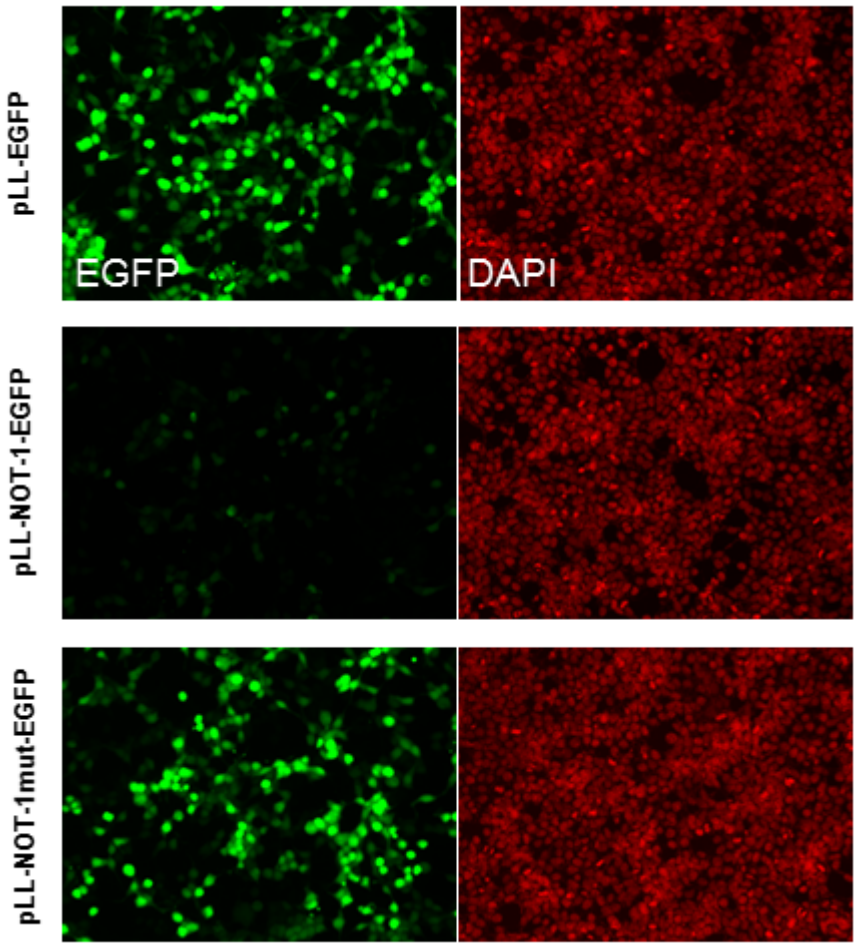



Figure 4.3: The NOT-1 hammerhead ribozyme undergoes self-cleavage within a cellular environment

*1.75 × $10^6$ HEK293T cells were transfected with 4 µg of pLL-EGFP, pLL-NOT-1-EGFP or pLL-NOT-1mut-EGFP. The incorporation of the NOT-1 ribozyme into the 5' UTR of the EGFP gene carried by the expression vector results in potent inhibition of GFP fluorescence following transfection into HEK293T cells as compared to transfection of cells with the EGFP expression vector lacking the 5' UTR-inserted ribozyme (middle left panel compared to top left panel). No qualitative difference can be observed between cells transfected with vectors lacking the ribozyme or those carrying the matched construct containing the catalytically-inactive mutant ribozyme (NOT-1mut). This would suggest that the observed NOT-1-dependent inhibition of GFP fluorescence results from the catalytic activity of this ribozyme.*

overcome these limitations, destabilised EGFP variants have been developed through fusion of the destabilisation domain of the mouse ornithine decarboxylase to the C-terminal domain of EGFP, leading to half-lives of the new fusion protein as low as 2 h (a strategy discussed and pursued in Chapter 3 to reduce the half-life of the L7Ae regulator protein) [167, 178]. The levels of a destabilised EGFP output protein should provide a closer reflection of transcript abundance. The surrounding sequence environment is another factor that may influence the ON:OFF ratio of an embedded ribozyme. The manner in which a ribozyme folds in isolation is very different to its behaviour and potentially its secondary structure when introduced into the context of a much larger mRNA transcript. In general it has been proposed that flanking sequences can influence ribozyme folding and can therefore directly affect catalytic activity. To minimise these effects, ribozymes have been previously flanked by repetitive poly(A) spacer motifs [172]. These motifs are predicted to have an unstructured conformation and are therefore much less likely to interfere with ribozyme folding.

Three pairs of constructs were generated in order to quantitatively assess the impact of EGFP destabilisation and the introduction of ribozyme-surrounding spacer motifs on the ON:OFF ratio. The first pair was generated through modification of the previously described pLL-NOT-1(mut)-EGFP vectors (the bracketed "(mut)" notation in this label is used to indicate the two related vectors, one carrying NOT-1 and the second NOT-1mut) to introduce ribozyme-flanking poly(A) spacer sequences (pLL-spaced-NOT-1-EGFP and pLL-spaced-NOT-1mut-EGFP). The pLL-NOT-1(mut)-EGFP vectors were also separately modified to replace the gene coding for EGFP with that of the destabilised EGFP (dEGFP) to generate the second pair of constructs (pLL-NOT-1-dEGFP and pLL-NOT-1mut-dEGFP). The third pair of constructs encompassed both of these modifications (pLL-spaced-NOT-1-dEGFP and pLL-spaced-NOT-1mut-dEGFP). These constructs were then individually transfected into HEK293T cells and the levels of GFP fluorescence measured after 48 h by flow cytometry to determine their respective ON:OFF ratios (Fig. 4.4A). The consequence of substituting EGFP with its destablised derivative (dEGFP) could be inferred by comparing the ON:OFF ratios corresponding to the first (EGFP-encoding) and third (dEGFP-encoding) pairs of the above mentioned constructs (pLL-spaced-NOT-1(mut)-EGFP versus pLL-spaced-NOT-1(mut)-dEGFP), which revealed an increase from 5.69 to 24.2. Likewise, comparison between the second and third pairs of constructs (pLL-NOT-1(mut)-dEGFP versus pLL-spaced-NOT-1(mut)-dEGFP) revealed a spacer sequence motif-dependent increase in the ON:OFF ratio from 17.0 to 24.2. Given the positive impact on ON:OFF ratios conferred by these two modifications, both were incorporated into all subsequent ribozyme-containing constructs.

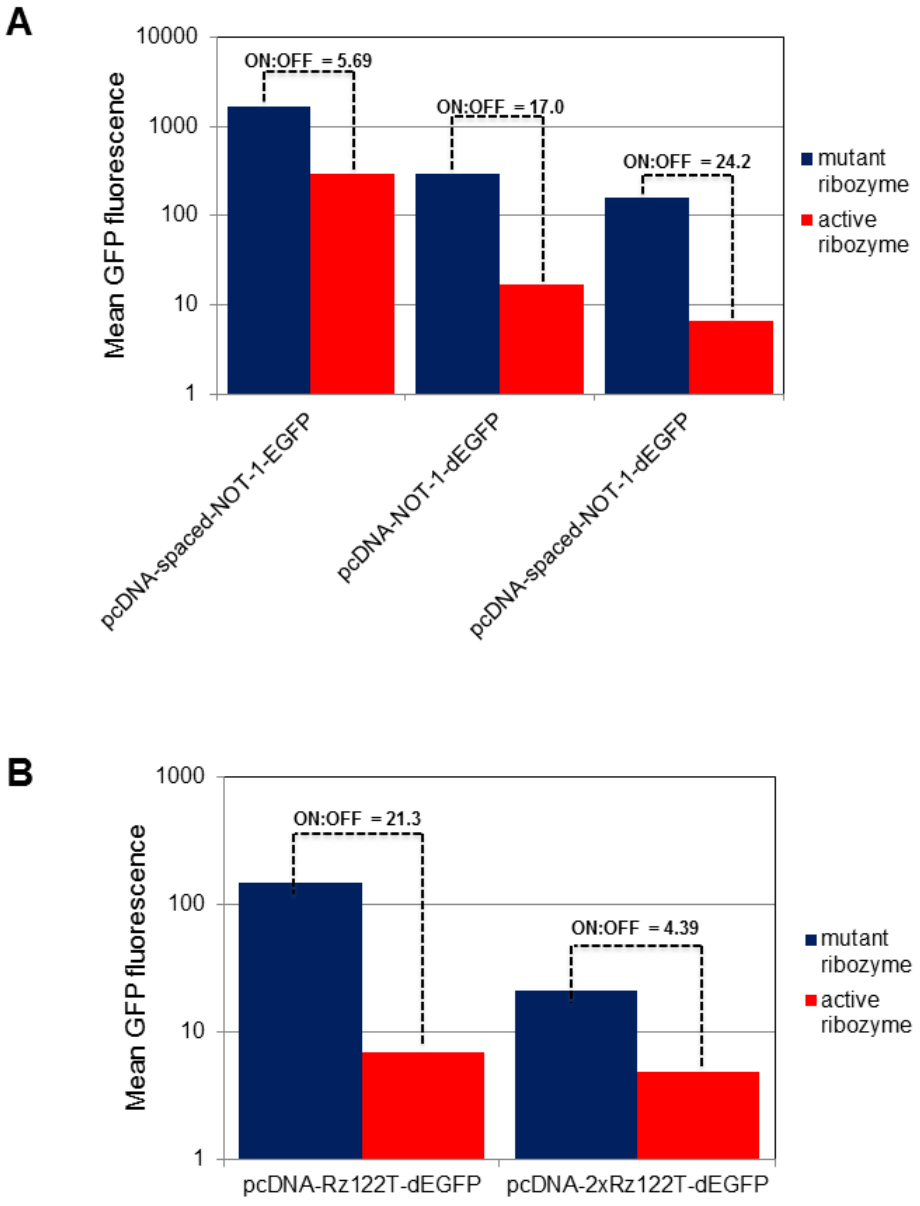



Figure 4.4: Exploration of ribozyme-independent measures to optimise the NOT-1-mediated ON:OFF ratio

*(A) 1.75 × $10^6$ HEK293T cells were transfected with 2 µg of pcDNA-spaced-NOT-1(mut)-EGFP, pcDNA-NOT-1(mut)-dEGFP or pcDNA-spaced-NOT-1(mut)-dEGFP and mean GFP fluorescence was measured 48 h later by flow cytometry. The notation "(mut)" within the plasmids named indicates that the two matched constructs encoding either NOT-1 or NOT-1mut were examined in parallel. The ON:OFF ratio for these experiments was defined as the ratio of the mean GFP fluorescence of the NOT-1mut (ON) bearing construct to that of the NOT-1 (OFF) matched construct. Both the addition of a destabilising tag to EGFP and the flanking of NOT-1 with spacer sequences independently resulted in an improved ON:OFF ratio. The observed effect was additive such that the greatest ON:OFF ratio was achieved by the construct incorporating both of these modifications. Continued...*

### 4.2.3 Introduction of a pair of tandemly located NOT-1 type ribozymes into the sensor construct led to an impairment of the ON:OFF ratio

Next, we tested the effect on the output ON:OFF ratio of introducing a pair of identical ribozymes in series within the 5' UTR of the transcript. The probability with which a mRNA transcript is cleaved by an internal ribozyme and subsequently degraded is determined by a rate constant associated with a particular ribozyme sequence. Thus, this probability should theoretically scale linearly with the addition of one or more identical ribozymes in the same transcript. To experimentally test this hypothesis a new pair of constructs was generated containing tandem 5' UTR-located Rz122T(mut) ribozymes which are structurally equivalent to NOT-1 and were designed using the Penchovsky and Breaker algorithm [64](pcDNA-2x-Rz122T(mut)-dEGFP). These constructs, as well as the co-constructed matched single ribozyme-containing comparators (pcDNA-Rz122T(mut)-dEGFP), were then transfected into HEK293T cells for analysis by flow cytometry. The presence of two active ribozymes appeared to be capable of inhibiting EGFP expression by an additional 30 % compared to that associated with the single ribozyme-bearing construct, although further replicates would be required to determine the significance of this result (Fig. 4.4B). This would be consistent with a ribozyme copy number-dependent partial additive effect on transcript cleavage, and the consequent influence on reporter protein expression levels. Intriguingly though, the matched construct harbouring the pair as compared to the single mutant ribozyme, also revealed a decrease in GFP fluorescence levels (Fig. 4.4B). Given that these mutant ribozymes are unable to undergo self-cleavage, a conclusion supported by data shown in Fig. 4.3, the observed repression must almost certainly result from a cleavage-independent inhibitory mechanism impacting preferentially on the construct with the tandem Rz122T mutant ribozymes. Stable RNA secondary structures within the 5' UTR of a transcript can substantially attenuate and even potentially prohibit translation, a phenomenon exploited in Chapter 3. This could explain some or all of the inhibitory effect conferred by a mutant ribozyme and if so the effect would likely be exacerbated upon increasing the number of ribozyme copies. Furthermore, the presence of ATG sequences within 5' UTR-located ribozymes has been reported to serve as premature start codons for translation,

*Figure 4.4 continued: (B) $5 \times 10^5$ HEK293T cells were transfected with 1 µg of pcDNA-Rz122T(mut)-dEGFP or pcDNA-2xRz122T(mut)-dEGFP and mean GFP fluorescence was measured 48 h later by flow cytometry. The addition of a second ribozyme to the 5' UTR of the dEGFP transcript had the unexpected effect of reducing the ON:OFF ratio. This was mainly a consequence of the reduced GFP fluorescence associated with the construct encoding 2xRz122Tmut (duplicate tandem copies of Rz122mut) compared to that encoding 1x Rz122Tmut (single copy of Rz122mut).*

thus inhibiting initiation from the downstream dEGFP start codon [173]. This second confounding effect may also be enhanced by the presence of two mutant ribozymes as compared to one. Due to this unexpected behaviour, the ON:OFF ratio of the single ribozyme-bearing constructs exceeded by a very substantial margin that of the double ribozyme-bearing constructs (Fig. 4.4B, 21.3 vs. 4.39). Hence, given these results, all subsequent miRNA sensors were constructed using a single NOT-1 based ribozyme.

### 4.2.4 The location of the ribozyme within the mRNA transcript influences the formers inhibitory effect

Although the rate of ribozyme-mediated catalytic self-cleavage should be largely independent of its location within an mRNA transcript (assuming minimal-to-no impact on the secondary conformation of the ribozyme), this location parameter may still influence the expression of the mRNA encoded gene as a direct consequence of the outcome of ribozyme cleavage. Two major distinctions in the inhibitory mechanisms apply depending on whether the ribozyme is located in the 5' or 3' UTR of a reporter gene. Based on the established model of translation, cleavage mediated by a 5' UTR-located ribozyme will result in an immediate block to translation as the 5' cap would be removed and therefore scanning ribosomes initiating from this structure would be prevented from reaching the open reading frame of the transcript. In addition the removal of the 5' cap would result in a destabilised mRNA transcript susceptible to degradation by the Xrn1p exonuclease. By contrast, cleavage by a 3' UTR-located ribozyme will have a similar ultimate effect but be mediated by a different pathway. In this case, separation of the poly(A) tail from the transcript coding sequence and the resulting loss of interaction between PABP and eIF4G will lead to an inhibition of translational initiation. Additionally, the 3' UTR-cleaved mRNA will be targeted for degradation in a 3' to 5' direction by the exosome complex.

Yen *et al.* [63] performed a systematic analysis in mammalian cells of ribozyme-mediated inhibitory effect as a function of position within a reporter transcript. This initial study revealed that 5' UTR-located ribozymes could efficiently repress output reporter expression between 62 and 225-fold compared to a ribozyme mutant control, far in excess of the inhibition achieved from 3' UTR-located ribozymes, which did not exceed 4-fold. To avoid the risk of cleavage-independent translational inhibition that may result from the presence of stable RNA secondary structure in the 5' UTR, several subsequent studies have argued that a 3' UTR-located ribozyme may preferable, and Win *et al.* have shown that such an approach was capable of achieving a 50-fold inhibition of reporter expression in yeast cells [172]. Strong inhibition (> 100-fold) from a 3' UTR-located ribozyme has since also been demonstrated in

mammalian cells [175]. Furthermore, in this latter study the observed level of repression was superior to that achieved by way of an equivalent 5' UTR-located ribozyme [175]. Part of the explanation for these discrepancies may relate to minor differences between the classes of ribozyme used in the individual studies. Although all three studies were conducted using *Schistosoma mansoni*-derived hammerhead ribozymes, Yen *et al.* [63] used a type I configuration in which the ribozyme is joined to the transcript through stem I. In contrast, Win *et al.* [172] and Beilstein *et al.* [175] used a type III configuration based on stem III linkage to the mRNA which has been shown to have a far greater repressive ability [60].

Considering the contradictory results generated by these previous studies, we sought to investigate the positional effects that may occur in our system. All constructs designed and used in this chapter fall under the type I class of *S. mansoni* hammerhead ribozymes. The published NOT-1 ribozyme was subjected to minor sequence editing and analysis based on published methods for the generation of small oligonucleotide-responsive allosteric ribozymes to produce the nominated variants and randomised sequence-bearing ribozyme pools investigated in this study [64, 66]. To study positional effects, active or mutant ribozymes were cloned into the 3' UTR of the dEGFP gene to produce the pcDNA-dEGFP-spacer-Rz122T(mut) plasmid-pair and their ON:OFF ratio was compared to that of equivalent constructs harbouring 5' UTR-located ribozymes (pcDNA-spacer-Rz122T(mut)-dEGFP). These constructs were then transfected into HEK293T cells and analysed 48 h later by flow cytometry. Potent reductions (~10 - 20-fold) in GFP fluorescence were observed when comparing constructs containing an active ribozyme to those with a mutant ribozyme regardless of whether the ribozymes was 5' or 3' UTR-located (Fig. 4.5). The absolute mean GFP fluorescence associated with the construct possessing the 5' UTR-located active ribozyme was approximately 50 % that of the matched 3' UTR-located active ribozyme-bearing construct. This was however more than offset by an even greater reduction in GFP fluorescence observed following transfection of constructs bearing the 5' UTR compared to the 3' UTR-located mutant ribozyme. These results suggests that when located within the 5' UTR of a reporter transcript, both the catalytically inactive mutant ribozyme and its wild-type parent are capable of inhibiting EGFP expression through a cleavage-independent mechanism; this is consistent with earlier published studies. Together, these data revealed an ON:OFF ratio of 10.6 and 20.3 for 5' UTR-located and 3' UTR-located ribozymes, respectively. Based on the superior ON:OFF ratio conferred by 3' UTR-located ribozymes, this ribozyme insertion location was chosen for subsequent experiments.

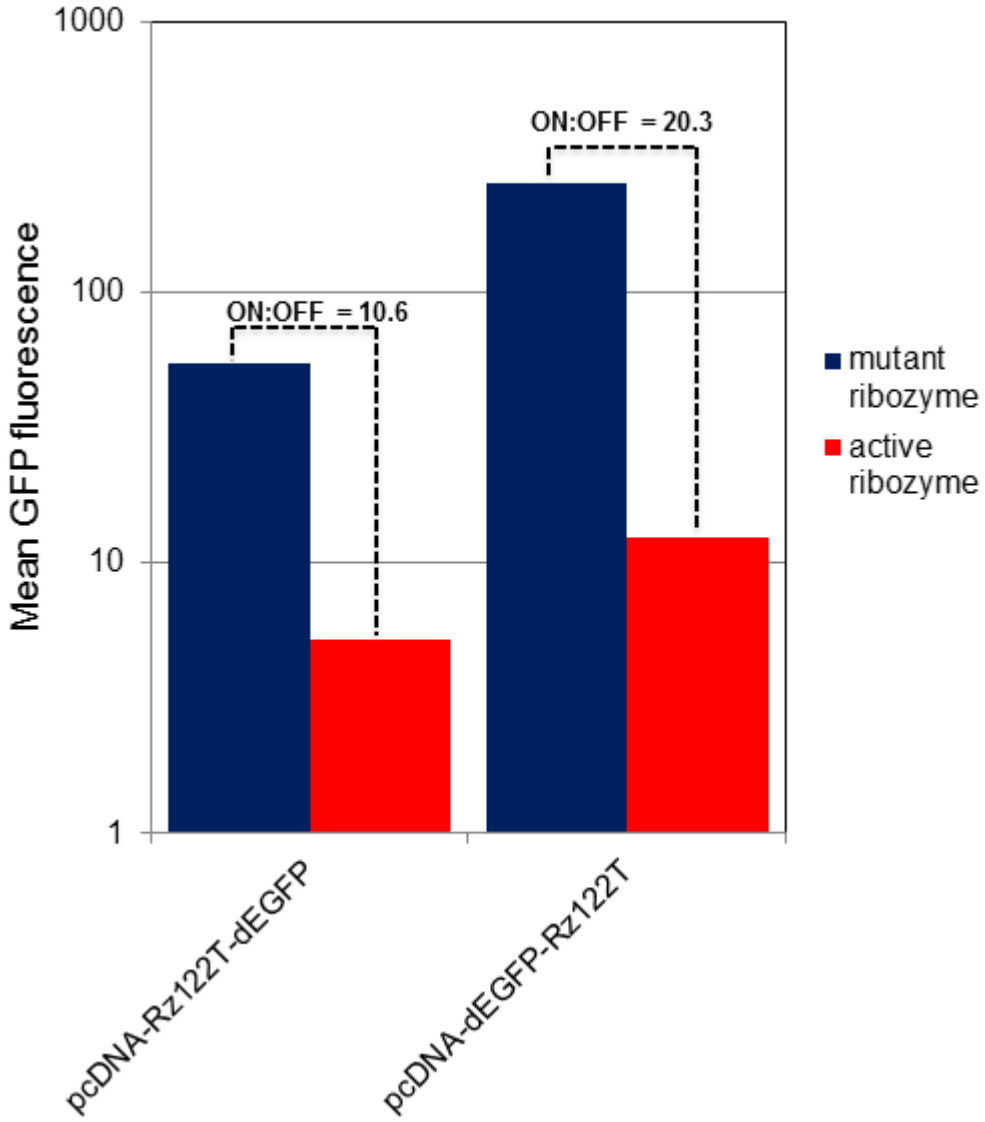



Figure 4.5: Investigating the ribozyme ON:OFF ratio as a function of the location of the ribozyme within the output mRNA transcript
*$5 \times 10^5$ HEK293T cells were transfected with 1 µg of pcDNA-Rz122T(mut)-dEGFP (construct encoding the 5' UTR-located active or mutant ribozyme) or 1 µg of pcDNA-dEGFP-Rz122T(mut) (construct encoding the 3' UTR-located active or mutant ribozyme). Flow cytometry was performed 48 h later to measure mean GFP fluorescence. Based on these data it can be seen that integration of Rz122mut within the 3' UTR as compared the 5' UTR of the transcript has led to a partial reduction of the cleavage-independent inhibition obseved in association with this mutant ribozyme-carrying construct. However, this has been achieved at the expense of a smaller reduction in the potency of inhibition mediated by the active ribozyme when this enzyme was located within the 3' UTR. Hence, overall, location of the Rz122 ribozyme within the 3' UTR of the transcript under study demonstrated the superior ON:OFF ratio.*

### 4.2.5 Rationale underlying the design and construction of the diverse RzN8 ribozyme pool intended for *in vitro* selection

Based on our optimisation studies, a random pool of constructs was generated containing a ribozyme with a randomised region, flanked by poly(A) spacer sequences within the 3' UTR of a dEGFP reporter gene. Additionally, the template ribozyme required engineering in order to confer upon it potential allosteric modulation in response to target miRNA binding. The process and rationale behind this engineering is described below.

Biochemical analysis of the hammerhead ribozyme has revealed that its minimal configuration consists of a catalytic core containing 15 highly conserved nucleotides, from which three base-paired stems arise [179]. A more extended hammerhead ribozyme conformation has been subsequently described which contains tertiary interactions between the loops on stems I and II, thus conferring the ability to autocleave under conditions of low $Mg^{2+}$ concentrations found in the cellular environment [70]. The sequence of the loop of stem III is not involved in catalysis and hence amenable to modifications, as has been demonstrated by the incorporation of ligand or oligonucleotide binding domains at this locus [64, 66]. Ligand binding to these allosteric sites may result in either the stabilisation or destabilisation of stem III, a process that will activate or inhibit ribozyme self-cleavage activity, respectively.

A random pool of ribozymes, termed the RzN8 pool, was designed by taking into account these considerations. In essence, the stem III sequence of ribozyme NOT-1 was replaced with a miR-21 oligonucleotide binding site (OBS) and a random linker consisting of nine degenerate nucleotides (Fig. 4.6). The miR-21 OBS was rationally designed to maximise binding affinity whilst minimising the potential for miRNA-mediated cleavage via the RNAi pathway. A previous systematic assessment of miRNA binding affinity and Ago2-catalysed cleavage based on examination of a comprehensive set of miRNA-target mismatches revealed that target sites containing non-complementary bases at positions 8 and 9 reduced the $k_{cat}$ of Ago2-mediated cleavage by as much as 93-fold [95]. At the same time, this 8-9 dinucleotide mismatch also demonstrated a higher affinity for the cognate miRNA than its perfectly matched counterpart, thus making a sequence that fits this criterion an ideal OBS candidate for our studies. Based on this information, the miR-21 OBS was designed to be a perfect reverse complement of miR-21 except for mismatches at positions 8-9. Additionally, it was decided to place the nine base randomised linker within a critical region at the 5' base of stem III, as it was hypothesised that *in vitro* selection should yield an evolved region that possesses sufficient energy to form a stable stem III when the OBS is unoccupied, yet of low enough energy to ensure disruption of stem III upon cognate miRNA binding. Notably, the 5'-most of these random nucleotides is at an invariant position in which an adenine (A) is required for the correct folding of a catalytically active ribozyme. This was chosen as a design feature to serve as a positive control

for the selection of active ribozymes, all of which would be expected to contain an A in this position.

### 4.2.6 Selection of allosteric ribozymes via an RNA-based RT-PCR strategy

To accomplish exponential enrichment of hammerhead ribozymes exhibiting miRNA-dependent allosteric modulation of cleavage activity, we devised a two-step strategy with similarities to SELEX and *in vitro* ribozyme selection (Fig. 4.7). First, a multicycle positive selection step was applied to select active ribozyme variants from within RzN8 pool. This was termed OFF state selection as it sought to enrich for more active ribozymes which would be expected to be associated with higher levels of silencing of EGFP expression, and therefore an output-defined circuit OFF state. This was followed by a multicycle negative selection step performed in the presence of the cognate miR-21 for the selection of ribozymes capable of miRNA-dependent allosteric inhibition. This was termed ON state selection due to the preferential ability of the sought ribozymes to turn EGFP expression ON in the presence of the cognate miRNA. Five iterative cycles of positive selection were performed followed by five iterative cycles of negative selection. To enable this experimental workflow, we devised a selection strategy that could discriminate between active and inactive ribozyme variants by RT-PCR (Fig. 4.8 and 4.11). The pooled ribozyme-containing constructs were transfected into HEK293T cells and total RNA was subsequently harvested at a defined timepoint. Ribozymes, either cleaved or uncleaved depending on the selection being applied, were recovered and reconstituted to form a new enriched pool (Fig. 4.7). The starting point for the outlined overall selection process was the RzN8 ribozyme pool containing a region of 9 randomised bases and therefore up to 262,144 unique ribozyme variants.

#### 4.2.6.1 OFF state selection for highly active self-cleaving ribozymes yields strong evidence of the desired selection but candidate selected ribozymes exhibit lower than parental levels of cleavage activity

Starting with the NOT-1 hammerhead ribozyme and upon randomisation of the sequence of one of its two predicted stem III arms and insertion of a miR-21 binding site to create the RzN8 ribozyme pool, the vast majority of constituent variants would be predicted to be incapable of folding into catalytically active structures. Consequently, it was important to first isolate from this pool only those sequences capable of sufficient self-cleavage activity to in-turn inhibit expression of a coupled EGFP transcript. This is a prerequisite characteristic of a miRNA-inhibitable allosteric ribozyme.

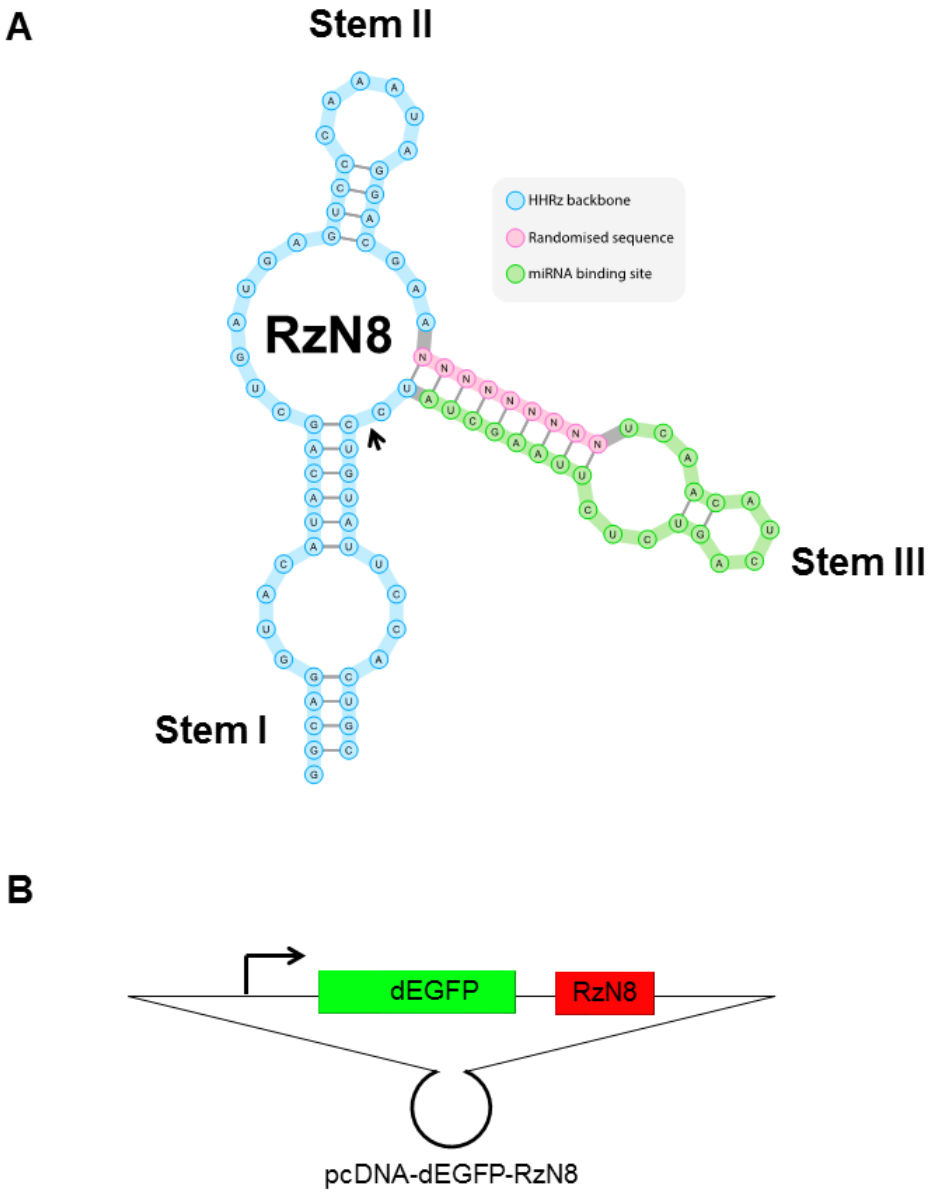



Figure 4.6: Putative secondary structure of the parent ribozyme used to generate the RzN8 ribozyme pool comprising variants carrying a nine nucleotide region of randomised sequence *(A) Shown here is the secondary structure of the parent ribozyme used to generate the RzN8 ribozyme pool folded into an active conformation in which stems I, II and III are correctly formed with relation to its catalytic core. Cleavage is predicted to occur at the scissile phosphodiester bond indicated by the arrow. Stem III has been replaced by a miR-21 OBS (yellow) and a nine base random linker (red). An active conformation requires an adenine at position N1 in order to base pair with the opposing uracil. (B) Schematic diagram indicative of the pcDNA-dEGFP-RzN8 plasmid pool. Each plasmid within this pool encodes one RzN8 variant, with the ultimate and ideal aim being for the plasmid pool to represent all RzN8 ribozyme variants. The transcription of the mRNA encoding dEGFP and the 3' UTR-located ribozyme is driven by the constitutive CMV promoter.*

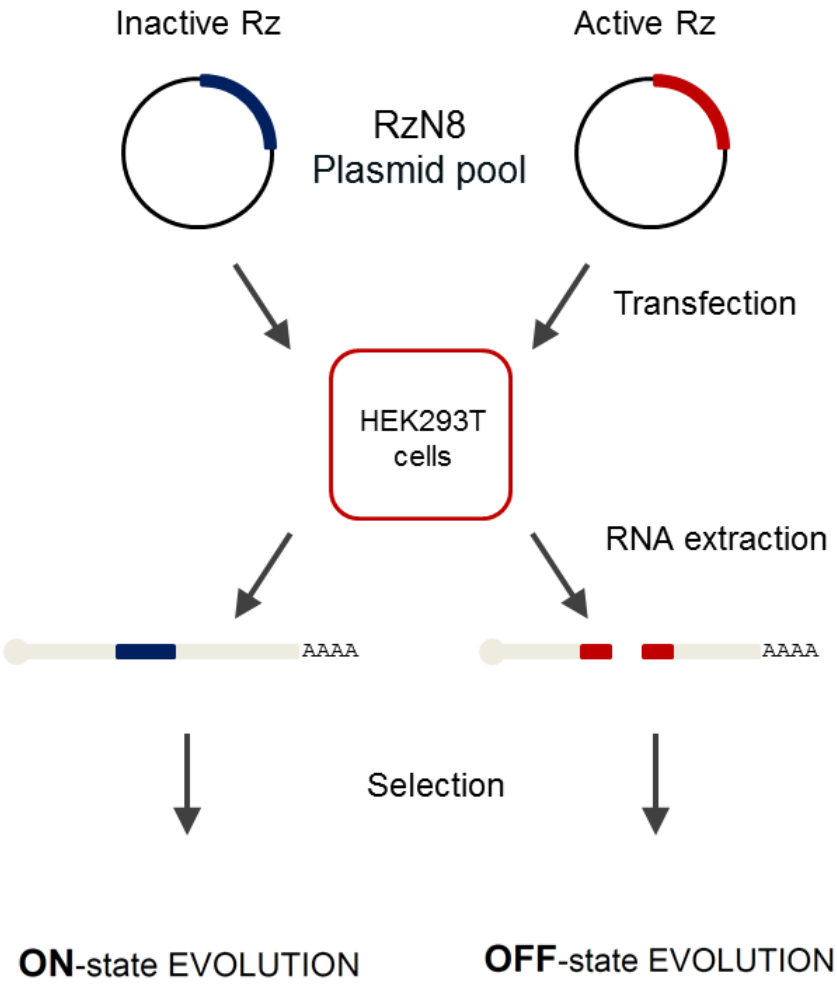



Figure 4.7: Overview of the ribozyme autocleavage activity-focused selection scheme that was performed as part of this study
*The RzN8 plasmid pool, comprising a random mixture of individual near-identical plasmids carrying either an active or inactive ribozyme variant, was transfected into HEK293T cells. Following RNA extraction, specific ON or OFF state selection pressures were applied to enrich for inactive (shown above as an uncleaved ribozyme) or active ribozymes (shown above as a cleaved ribozyme), respectively. Key to this strategy was the selective amplification, quantification and/or identification of cleaved versus uncleaved ribozyme sequences.*

As shown in Figure 4.8, each cycle of the OFF state selection implemented in this study is itself a multi-step process. It begins with transfection of HEK293T cells with the pcDNA-dEGFP-RzN8 plasmid pool, which drives intracellular expression of the diverse ribozyme-containing transcripts. These ribozymes subsequently undergo self-cleavage at varying rates depending on the identity of their random sequence and consequent secondary structure. After a varying periods of time (eg. 1.5 - 48 h), total cellular RNA containing full length ribozyme and ribozyme cleavage products is isolated. The time between transfection and RNA extraction is used as a determinant of selection stringency on the basis that the more active ribozymes would expected to be more likely to be found in a cleaved state at earlier timepoints. Extending this logic, at the earliest timepoint investigated, the majority of ribozyme cleavage products present will have originated from the subset of particularly highly active ribozyme variants for which the probability of cleavage within this brief window of time is high. As time progresses, cleavage by-products from less active ribozymes will accumulate alongside those originating from their more active counterparts. Consequently, low stringency selection may be performed long after transfection to capture all ribozymes capable of self-cleavage without undue rate-based discrimination. High stringency selection on the other hand must occur shortly after transfection to enable the preferential selection of of ribozyme variants mapping towards the top end of the activity spectrum. Based on these general principles, the stringency of selection is expected to correspond in an inverse manner to the time interval between transfection and RNA extraction.

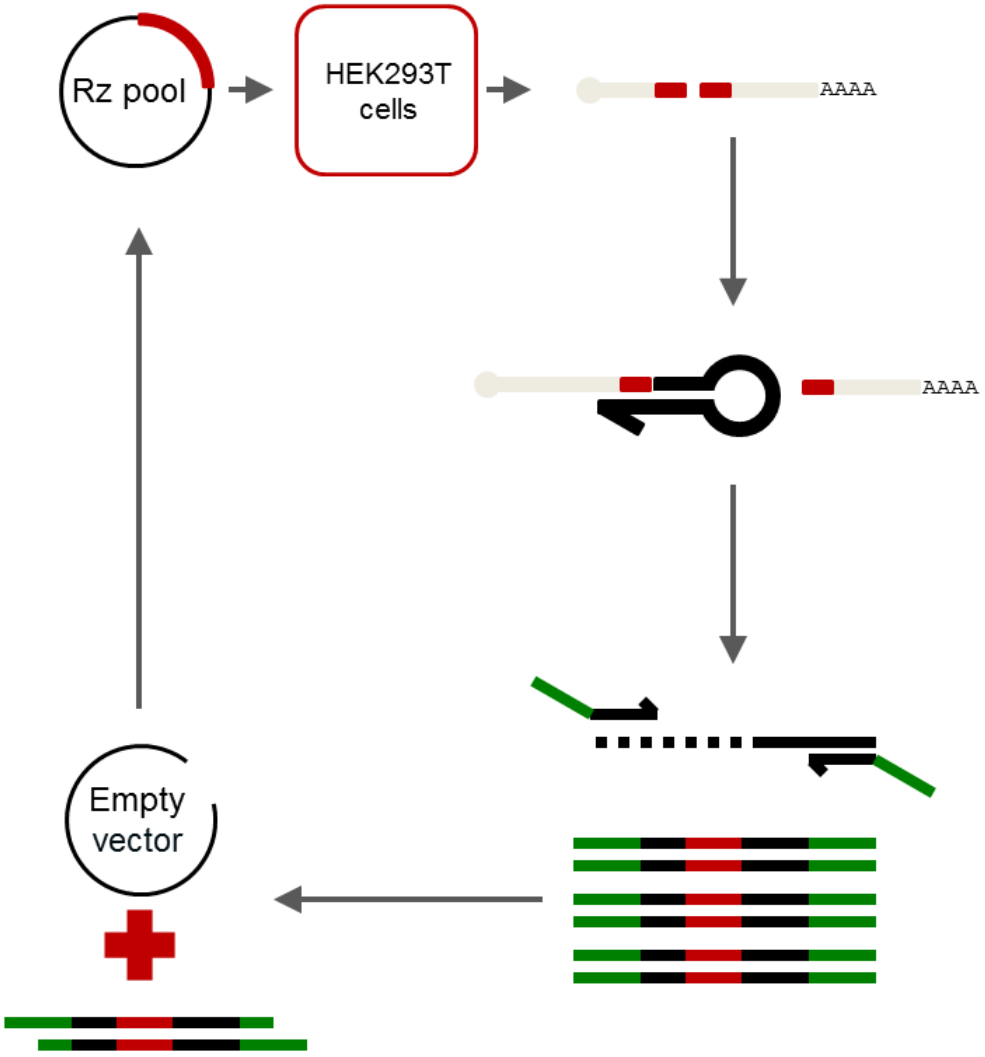



Figure 4.8: Detailed OFF state selection scheme for the selection of active ribozymes
*OFF state selection proceeded via the following experimental steps: (1) Transfection of the plasmid pool encoding the Rz variants into HEK293T cells, (2) extraction of total RNA including both cleaved and uncleaved ribozyme containing transcripts, (3) use of a structured stem loop primer ensures the specific reverse transcription of the targeted cleaved ribozyme fragment into cDNA, (4) PCR amplification of this cDNA and reconstitution of full-length ribozymes through use of PCR primers carrying the necessary 5' extension sequences to reconstitute the ribozyme and append near-terminal restriction sites, (5) restriction enzyme digest of the resulting ribozyme-encoding amplicons and the pcDNA-dEGFP empty vector to enable cohesive-end cloning and recovery of an enriched plasmid pool. Continued...*

Once RNA was extracted, active ribozyme variants were isolated using a structured stem-loop RT-PCR method which specifically and highly selectively captured and reconstituted their associated cleavage by-products. This primer was designed using the same principles as those used by previous investigators to generate tools for the differential amplification of mature miRNAs over their precursors [158, 180]. The structured stem-loop primer used in this study was designed to fold into an intramolecular hairpin conformation in which the 5' end is recessed into a double stranded stem region and the six last nucleotides at the 3' end are exposed as a single strand overhang (Fig. 4.8). The 3' overhang was designed to be complementry to the last six bases of the 5' ribozyme cleavage by-product, to which it was intended to bind so as to initiate reverse transcriptase-mediated cDNA synthesis. Such a primer is not specific for the cleavage product at a sequence level because its complementary binding site is present in both the full-length ribozyme and its 5' cleavage by-product. Instead, specificity and selectivity are derived from the steric hindrance that occurs between the double-stranded stem region of the structured primer and the RNA sequence immediately downstream of the putative cleavage site within the context of the full-length ribozyme. Following cleavage, the 3' end of the 5' cleavage by-product is now permitted to interact with the complementary sequence of the structured stem-loop primer. Following specific synthesis of cDNA from the 5' by-product of ribozyme self-cleavage, PCR amplification was used to reconstitute the entire ribozyme sequence by the addition of 5' extended tails to the forward and reverse primers. These double-stranded ribozyme-encoding DNA amplicons were then digested with a restriction endonuclease directed at newly incorporated near-terminal sites, and the now sticky-ended amplicons cloned into a destination vector to generate a new enriched ribozyme plasmid pool, at which point the process of selection was initiated again (Fig. 4.8).

In order to track the progress of selection, a sample of the ribozyme pool generated from each cycle of the selection procedure was analysed by Sanger sequencing. Despite the heterogeneity of the plasmid pool, the only readily observable sequence differences were confined to a small nine base window, as was anticipated, lying approximately midway along the region targeted for sequencing. As the sequencing reaction passed through the invariant 5' and 3' ribozyme regions, it was expected to "encounter" the same base across the huge multitude of independent DNA strands undergoing "pooled" sequencing, at each and every position in turn (note: each position of the invariant regions possesses its own specific base); thus, leading

*Figure 4.8 continued: Specifically, this strategy would be predicted to lead to an enriched plasmid pool carrying a greater proportion of more active ribozymes than the starting plasmid pool used at step (1). Hence, when the process is applied reiteratively and each plasmid pool in turn is subjected to a new round of selection the process would over a number of cycles be expected to promote the selection plasmids encoding more active ribozymes.*

to the generation of a clean signal and a unique high confidence base call for each of the invariant regions positions in turn. Conversely, when the randomised region was sequenced, an assortment of bases was expected to be encountered at each of the nine randomised positions resulting in a mixed sequencing trace over this short window. The relative signal intensity corresponding to each of the four bases found within a single mixed peak would correlate with the relative frequencies with which each these bases occur at that particular position in the pool of post-selection, reconstituted ribozymes from which they are derived. Assuming that the RzN8-containing plasmid pool was created in an unbiased manner and that all potential sequence combinations across the randomised 9 bp window were equally represented, this sequence distribution pattern should have been observed in a pre-selection Sanger sequencing trace. Consistent with this expectation, pre-selection sequencing of the RzN8 plasmid pool revealed the presence of mixed base peaks, tightly confined to the randomised nine nucleotide region (Fig. 4.9A). However, based on a qualitative interpretation of the sequencing chromatogram, although the A, T and C bases were found with similar frequencies at all nine positions, a moderate G bias (G peaks were approximately twice the hight of those of the other three bases) was observed for all randomised positions except the first. This may be a result of the chemical synthesis protocol used to generate the random oligonucleotide library from which the ribozyme pool was generated. For the generation of this randomised pool, the machine mixing method was chosen, in which an equal mix of all four base precursors are injected at each extension step during oligonucleotide synthesis. Theoretically, this should enable the incorporation of each base with equal frequencies across the pool of oligonucleotides. However, it is known that base incorporation is not entirely random and that such a machine mixing method is prone to context dependant biases, possibly accounting for the observed G bias at 8 of the 9 randomised positions within the RzN8 plasmid pool. The more costly method of hand mixed random bases can be used to precisely define 25% frequency of occurrence of each base at every position. Alternatively and far less likely, it is possible that low-level cloning biases impacting on the construction of the RzN8 plasmid pool may have favoured the primary cloning and/or *E. coli*-based amplification of plasmid variants carrying one or more G bases at these positions.

Having finalised a selection scheme (Fig. 4.8) and confirmed the generation of a highly diverse RzN8 plasmid pool (Fig. 4.9A), we initiated the process of selection. The RzN8 plasmid pool was subject to five iterative cycles of OFF state selection for active ribozymes. Subtle differences in the appearance of the mixed sequencing traces mapping to the region of randomisation started to emerge from the very first cycle of positive selection, and became more pronounced by the end of the fifth cycle (Fig. 4.9B). The most striking change was the increasing signal intensity corresponding to an A base at the first position, which was ex-

pected given its requirement for catalytic activity of the ribozyme. Following the completion of five cycles of positive selection conducted at increasing levels of stringency (decreasing times between transfection and RNA isolation), nine individual plasmid clones were isolated at random and sequenced individually. An A nucleotide was found at the first position in all nine clones, as was a T at the second position, as per expectations arising from the observed chromatogram obtained following five OFF state selection cycles (Fig. 4.9B). Indeed, this result was strongly indicative of successful selection of active ribozyme variants. The AT dinucleotide which emerged from the random sequence would be expected to base pair with the TA nucleotides at the base of stem III in the fixed ribozyme sequence, a feature that is known to be essential for the catalytic activity of hammerhead ribozymes. Alignment of sequences obtained from the nine randomly selected clones was used to generate a Weblogo representation of the consensus sequence and base composition within the region of randomisation (Fig. 4.9B). It is interesting to note that even after five cycles of OFF state selection that there were only relatively low levels of base biases at positions 3 - 9, and in particular at position 6 which exhibited no clearly apparaent bias either by the sequencing chromatogram data corresponding to the entire captured amplicon pool or the Weblogo information derived from sequences of the nine arbitraly selected clones. This was an important finding as it demonstrated that huge diversity remained within the ribozyme population even following the OFF state selection process, and hence maximised the chance of the existence within the newly enriched pool of allosteric variants that may have been subsequently identified by ON state selection.

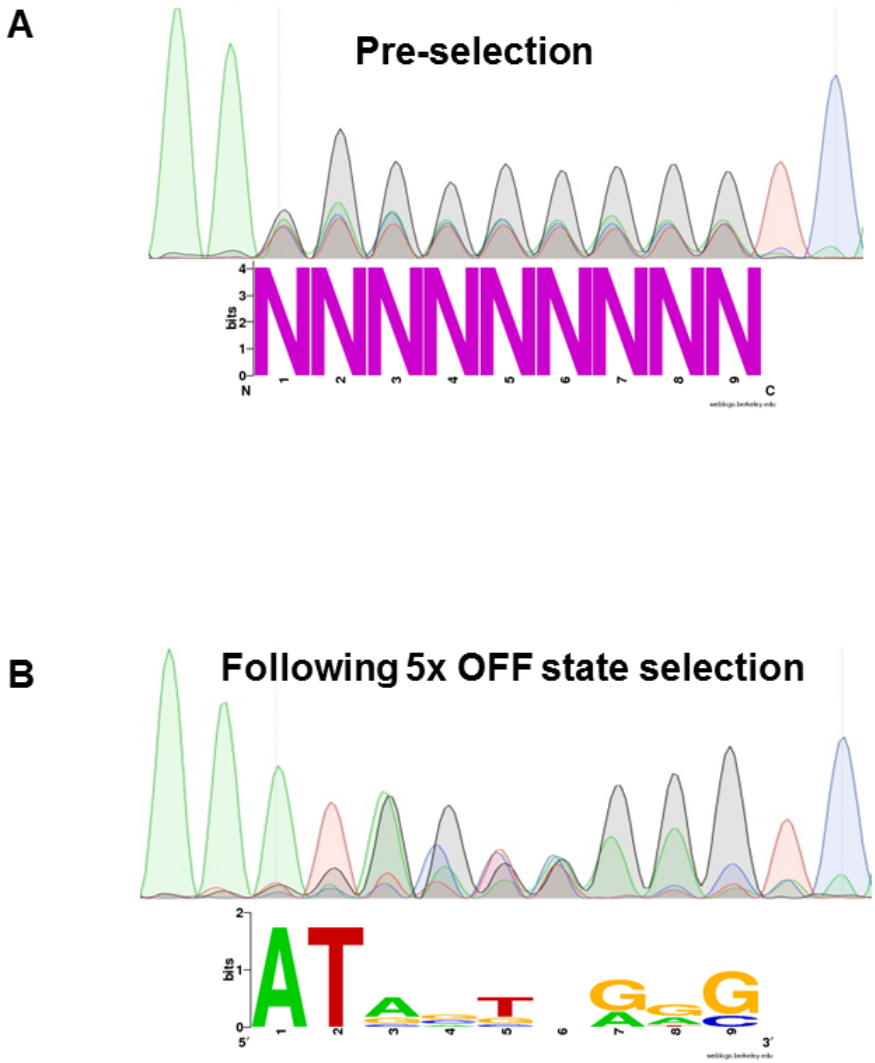



Figure 4.9: Tracking the selection journey via Sanger sequencing
*(A) Pre-selection sequencing chromatogram of the relevant region of the randomised RzN8 ribozyme pool and corresponding Weblogo. Mixed sequencing traces were confined to the nine base region of randomisation and indicated a high degree of heterogeneity amongst the plasmid pool carrying the RzN8 ribozyme variants. With the exception of a substantial G bias that was observed at all but the first position within the randomised region, the remaining three bases appeared to be equivalently represented at each of the nine positions. (B) Sequencing chromatogram of the plasmid pool reconstituted following five cycles of OFF state selection for active ribozymes. By comparison to the pre-selection sequencing chromatogram, it was clearly apparent that strong biases were observed to have emerged within the region of randomisation, in particular at positions 1 - 4 and 7 - 9, as may occur due to selection of bases that are required for catalytic activity, or even those that are favoured as these are associated with higher-levels of catalytic activity. Continued...*

Finally, a more direct approach to investigate self-cleavage activity of the nine ribozymes isolated following the positive selection strategy was pursued. Given that the ribozymes were contained in the 3' UTR of a reporter dEGFP gene, their activity could be measured through the efficiency with which they could reduce GFP fluorescence following transfection into cells. pcDNA-dEGFP was transfected into HEK293T cells to establish a baseline level of GFP fluorescence by flow cytometry. In parallel, each of the nine pcDNA-dEGFP based ribozyme-encoding plasmids (see Fig. 4.10 for individual plasmid designations) was transfected into HEK293T cells and GFP fluorescence measured using the same parameters. All nine tested ribozymes were capable of reducing the level of GFP fluorescence by approximately 50 % as compared to that associated with the ribozyme-lacking pcDNA-dEGFP (Fig. 4.10). This result confirmed the specificity of the selection strategy for active self-cleaving hammerhead ribozyme variants. However, despite their ability to destabilise the EGFP reporter mRNA (based on reduced GFP fluorescence readouts), all chosen variants at this stage of selection demonstrated activity well below (< 10%) that of their parental ribozyme (Rz122T; see Fig. 4.5 for data on pcDNA-dEGFP-Rz122T), suggesting a potential major limitation in the selection approach that was applied as described above.

*Figure 4.9 continued: Nine randomly-selected plasmids from this pool were sequenced individually and aligned to create the displayed Weblogo. This is consistent with the chromatogram in showing the emergence of preferences for particular bases at specific positions, though clearly as only nine plasmids were sampled the Weblogo signature obtained was unlikely to capture the full picture of sequence diversity represented within this pool of plasmids. Nevertheless, of particular note was the occurrence of an A at the first position (positive control) in all nine plasmids sequenced, a finding that was very strongly concordent with the relative heights of the four chromatogram peaks at this position.*

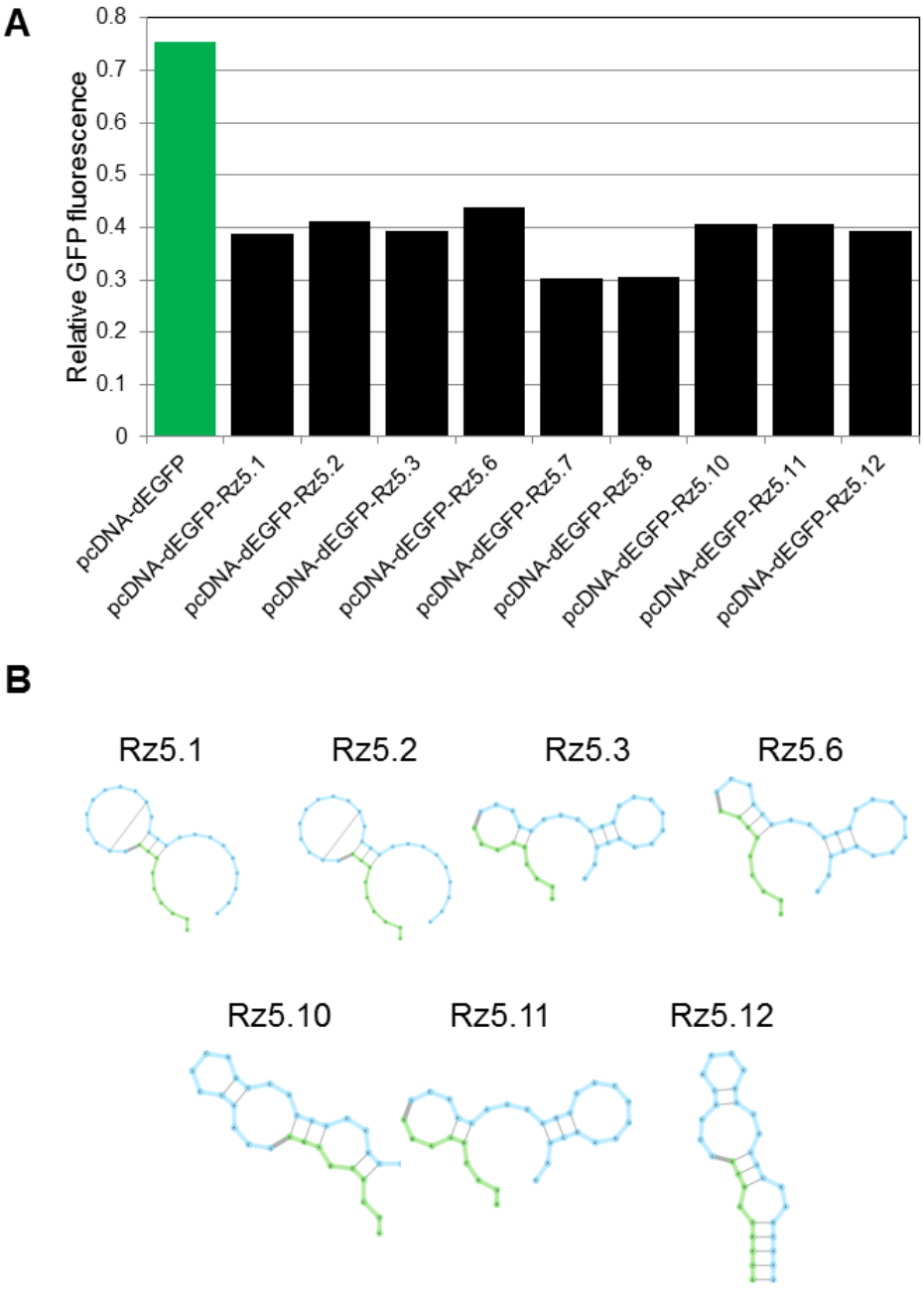



Figure 4.10: Nine randomly chosen RzN8 ribozymes present following the 5th selection cycle are capable of inhibiting dEGFP expression

*(A) 5 × $10^5$ HEK293T cells were transfected individually with 500 ng of pcDNA-dEGFP or one of the nine pcDNA-dEGFP-Rz5.X plasmids (Note: Rz5.X, 5 denotes a ribozyme drawn from the reconstituted pool following the $5^{th}$ cycle of selection and X denotes the specific clone number), in addition to 500 ng of piBlue-N1 as the transfection control. Fourty-eight hours post transfection GFP and iBlue fluorescence were measured by flow cytometry to calculate the relative GFP fluorescence. All nine RzN8 clones chosen from this $5^{th}$ cycle pool demonstrated repressive activity towards dEGFP expression. This result supported the specificity of the OFF state selection protocol for the selection of catalytically active ribozymes as none of the randomly selected clones failed to exhibit at least 40 % inhibition of relative GFP fluorescence as compared to that associated with the control ribozyme-free plasmid. Continued...*

#### 4.2.6.2 ON state selection for ribozymes capable of miRNA-mediated allosteric inhibition of cleavage activity was confounded by the large-scale generation and enrichment of indel-containing "cheater" ribozymes

Having set out to enrich for ribozymes that exhibit high levels of self-cleavage over the course of five cycles of OFF state selection, an ON state selection was carried out to select ribozyme variants capable of allosteric inhibition by the cognate miRNA. The process of ON state selection closely resembled that of OFF state selection, with one important distinction at the stage of RT-PCR recovery and reconstitution. For ON state selection reverse transcription primers were designed to bind downstream of the cleavage site such that only uncleaved ribozyme transcripts could be converted to cDNA and subsequently amplified by PCR with cleavage site-spanning primers (Figure 4.11). As before, the ON state selection stringency could be modified by changing the time between transfection and RNA extraction. Given that the input plasmid pool for the ON state selection encoded ribozymes that had been already selected for cleavage activity, the levels to which these accumulated in cells in the uncleaved form were predicted to display an inverse relationship to corresponding cleavage activities. Theoretically, a subset of these ribozymes may undergo allosteric inhibition in the presence of miR-21, thus prolonging the time these persist in the uncleaved form. RNA extraction performed shortly after transfection should enrich for ribozymes that are unable to autocleave, have not yet had time to undergo self-cleavage, as well as for ribozymes that have been allosterically inhibited by miR-21. Assuming that ribozymes lacking catalytic activity have largely been depleted following the initial OFF state selection, the longer the time between transfection and RNA extraction, the greater the proportion of ribozymes enriched by ON state selection that comprise those that have undergone allosteric modulation. Therefore, the longer the interval of time the higher the stringency of selection for allosterically inhibited ribozymes.

ON state selection was initiated by co-transfecting into HEK293T cells the $RzN8_{5xOFF}$-encoding plasmid pool obtained from the fifth cycle of OFF state selection and an expression vector governing synthesis of the cognate input miR-21. Five iterative ON state selection cycles were performed according to the scheme outlined in Fig. 4.11. As with prior OFF state selection, Sanger sequencing of the plasmid pools obtained following each ON state selection cycle was used as a means of tracking changes in the sequence composition within the region of randomisation. Unexpectedly, it was observed that the region of mixed sequence

*Figure 4.10 continued: (B) Predicted secondary structures of stem III for the 7/9 RzN8 clones tested produced using NUPACK [40] (blue = stem III constant sequence, yellow = post-selection randomised sequence; structure prediction was unsuccesful for stem III of Rz5.7 and Rz5.8).*

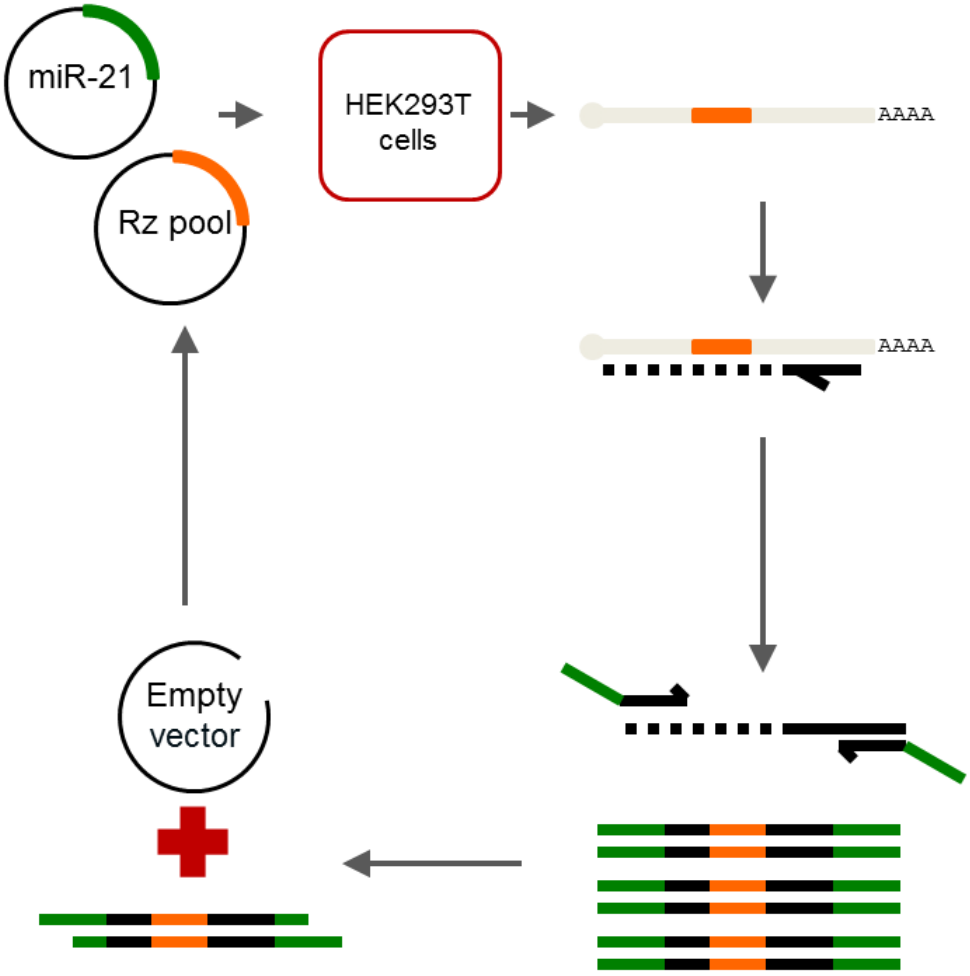



Figure 4.11: Detailed ON state selection scheme for the enrichment of allosterically inhibited ribozymes

*ON state selection proceeded via the following experimental steps:, (1) co-transfection into HEK293T cells of the plasmid pool carrying the set of ribozyme (Rz) variants and the miR-21 expression vector, (2) extraction of total RNA including both cleaved and uncleaved ribozyme containing transcripts, (3) specific conversion of uncleaved ribozyme transcripts into cDNA through use of a reverse transcription primer that binds downstream of the cleavage site, (4) PCR amplification of cDNA and reconstitution of full length ribozymes with 5' extended primers as described previously for OFF state selection, (5) restriction enzyme digest of ribozyme amplicons and the pcDNA-dEGFP empty vector to enable the cloning and recovery of an enriched plasmid pool carrying a greater proportion of allosterically inhibited ribozymes, and potentially variants that are subject to higher levels of allosteric inhibition.*

peaks generated by Sanger sequencing began to extend well beyond the nine nucleotide window that had been randomised in the starting plasmid pool and to eventually encompass the entire amplicon length (Fig. 4.12A). Theoretically, this outcome could have been explained if the plasmid pool contained multiple out-of-frame variants with respect to the binding site of the sequencing primer that arose through short and/or long unintended insertion and/or deletion (indel) mutations. Alternatively, large-scale accumulation of high densities of point mutations throughout the ribozyme sequence may also have explained the expansion of the mixed trace sequence region. A third scenario potentially accounting for this observation-may have been the occurrence of primer slippage or mis-priming during the recovery (reverse transcription step) and/or reconstitution (PCR step) phase of selection, or during the Sanger sequencing reaction itself. However, the clean umambiguous chromatogram calls from the final ON state selection amplicon pool at all ribozyme-invariant positions, argues against a sequencing primer associated cause. In order to explore this matter further, eight single plasmid clones were isolated from the plasmid pool that had been obtained following the five cycles of ON state selection to individually sequence their constituent ribozymes. In addition to the expected variations in sequence observed within the region of randomisation, each of the clones examined contained its own unique set of indel mutations mapping to sites within the surrounding ribozyme sequence (Fig. 4.12B). Thus, given that all eight randomly selected clones examined were characterised by unintended indel mutations and that the precise nature of these mutations differed from clone-to-clone, we can conclude that this "procedural bug" alone accounted for the vast majority if not all of the observed occurrence of mixed sequencing tracks resulting from Sanger sequencing of the intermediate and final plasmid pools obtained following OFF state selection.

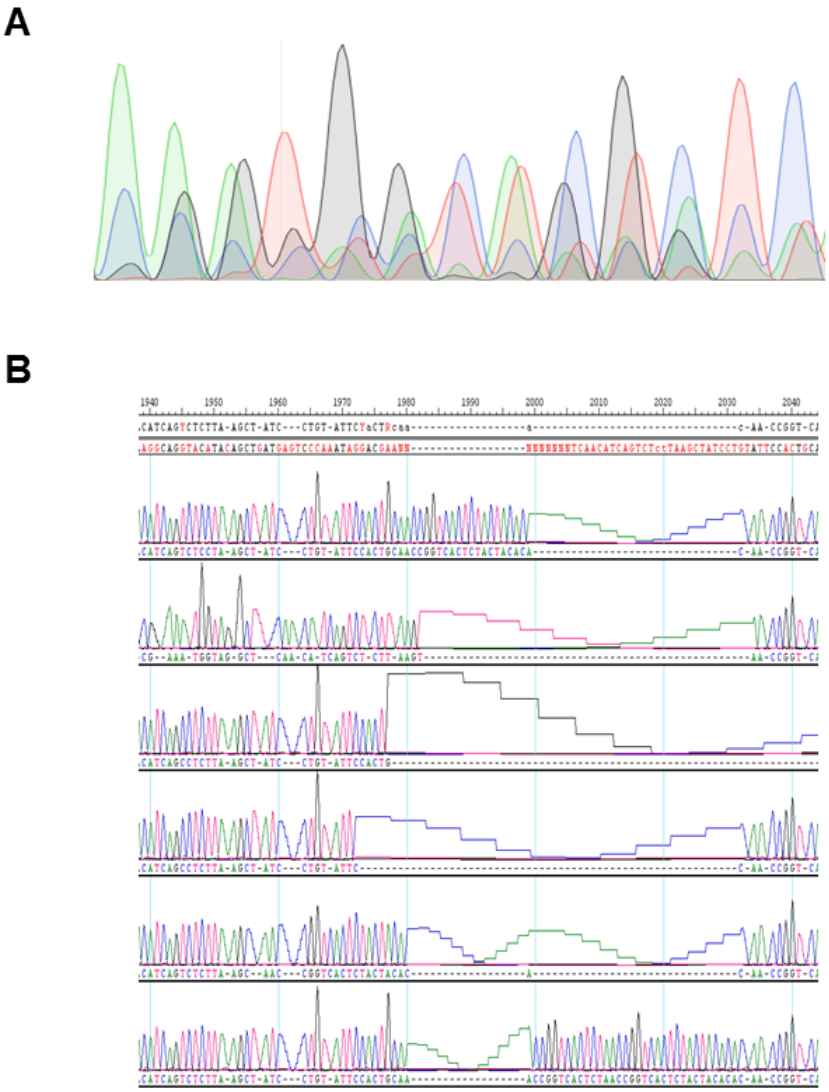



Figure 4.12: The unintended accumulation of ribozyme-inactivating indels over the ON state selection cycles confounded discovery of allosterically-inhibited ribozyme variants
*(A) Sanger sequencing chromatogram corresponding to 13 positions, centred around the nine base randomised region, from the ribozyme pool generated at the end of the $10^{th}$ selection cycle (5 cycles of OFF state selection, followed by 5 cycles of ON state selection). In contrast to Fig. 4.8, the mixed sequencing peaks extend beyond the defined region of randomisation (central 9 bases). (B) Six individual plasmid clones were isolated following the end of the $5^{th}$ cycle of ON state selection (10 completed cycles in total) and Sanger sequenced separately. Alignment of the sequences of these ribozyme variants revealed the occurrence of a different indel mutation in each of the sequences corresponding to the six randomly chosen clones. Continued...*

Although the ON state selection protocol enabled the selection of uncleaved ribozyme variants, it could not discriminate between those that had undergone allosteric inhibition and those that had been inhibited via other mechanisms. The occurrence of indels within conserved regions of the ribozyme that are required for catalytic activity was an alternate means by which catalytic activity was blocked. Therefore, so long as the mutations did not occur within the binding sites of the primers used for ribozyme recovery or reconstitution, such indel containing "cheater" ribozymes would have undergone exponential enrichment as efficiently, and potentially even more efficiently (eg. large cummulative deletions giving rise to shorter amplicons and/or ones possessing less PCR-inhibitory secondary structure) than the desired allosteric ribozymes.

Given the dominance of indel-containing "cheater" ribozyme sequences at the end of the negative selection step, it was decided that testing for miRNA-induced allosteric modulation among randomly selected representative obtained at this stage would have been hugely problematic (identification of indel-free candidates or those likely to have retained catalytic activity despite short indels would have been a substantial task) and likely of limited yield. Instead, further steps were taken in an attempt to filter out the "cheater sequences" and thus enable the identification of *bona fide* allosteric candidates.

#### 4.2.6.3 Additional OFF state selection as a means to filter out indel-containing ribozymes

Having completed sequentially five cycles of positive selection followed by five cycles of negative selection, the vast majority of ribozymes encoded within the newly constituted plasmid pool contained indels and were therefore unlikely candidates for miRNA-responsive allosteric ribozymes. It was however not possible to rule out the presence of indel-free ribozyme variants at low copy numbers amongst the huge masking population of "cheater" ribozymes. Indeed, if indel-free ribozymes were present, due to their presumed high activity throughout positive selection and low activity in the presence of miR-21 (negative selection), these would constitute very promising candidates for miRNA-responsive allosteric ribozymes.

In an attempt to identify such allosteric ribozyme candidates, the final plasmid pool obtained after five cycles of OFF state selection, followed by five cycles of ON state selection ($RzN8_{5xOFF5xON}$ encoding plasmid pool) was used as the starting material for additional cycles of OFF state selection, according to the scheme outlined in Fig. 4.8. Sanger sequencing of the

*Figure 4.12 continued: The undoubtedly frequent presence of ribozymes containing such indels within the pool and the observation that these mutations appeared to comprise a diverse set of variably sized indels markedly increased the probability that the sequences of a substantial pool of variants would be out of frame with respect to the sequencing primer and thus would explain the above observation.*

amplicon/plasmid pool obtained following each subsequent round of selection was conducted in order to track the process. Indeed, as predicted and in accordance with the application of strong selective pressure for active ribozymes, the presence of mixed peaks outside of the predetermined region of randomisation decreased with each cycle of OFF state selection. Even more interesting, was the concurrent reduction in mixed peaks within the region of randomisation. Strikingly, following the third of these additional OFF state selection cycles, Sanger sequencing suggested that a single ribozyme variant was overwhelmingly dominant within the population, consequent upon the near-total or total extinction of all other ribozyme species (Fig. 4.13A)! Consistent with this observation, the isolation and sequencing of several arbitrarily chosen single plasmid clones harvested from this pool, all of which had an identical sequence, supported the inferrence that the population was very substantially made up of a single ribozyme variant. Since this construct was the product of a combined 13 cycles of selection, it was termed Rz13.1 (Fig. 4.13B). At this point the selection process pursued in this study was deemed to be completed.

### 4.2.7 Functional analysis of the evolutionarily-selected candidate Rz13.1 allosteric ribozyme reveals a paradoxical apparent miR-21-associated activation of its cleavage activity

Having passed through successive OFF, ON, and OFF state selection steps, the "final" ribozyme remaining within the pool at the end of the 13th round (Rz13.1) was expected to demonstrate the desired allosteric modulation in response to the presence of the miR-21 input signal.

Since Rz13.1 was encoded within the 3' UTR of a dEGFP reporter gene (pcDNA-dEGFP-Rz13.1), measurement of GFP fluorescence could be used as an indirect readout of ribozyme cleavage activity. HEK293T cells were transfected with a control plasmid (pcDNA-dEGFP) or the Rz13.1-regulated construct (pcDNA-dEGFP-Rz13.1), and in parallel co-transfected with a control (pEM-empty) or miR-21 (pEM-21) expression vector as input. piBlue-N1 was also introduced into the cells under all conditions as a transfection control. High levels of GFP fluorescence were observed in cells transfected with pcDNA-dEGFP irrespective of the identity of the co-transfected miRNA expression vector (Fig. 4.14A). In contrast, GFP fluorescence in cells transfected with the construct coding for the Rz13.1-regulated output (pcDNA-dEGFP-Rz-13.1) and the control (pEM-empty) miRNA expression vector was reduced by approximately 50 % compared to that associated with pcDNA-dEGFP (control output) in the presence of either miRNA expression vector (pEM-empty or pEM-21; Fig. 4.14A). This effect could have been attributed to Rz13.1 mediated transcript cleavage and subsequent degrada-

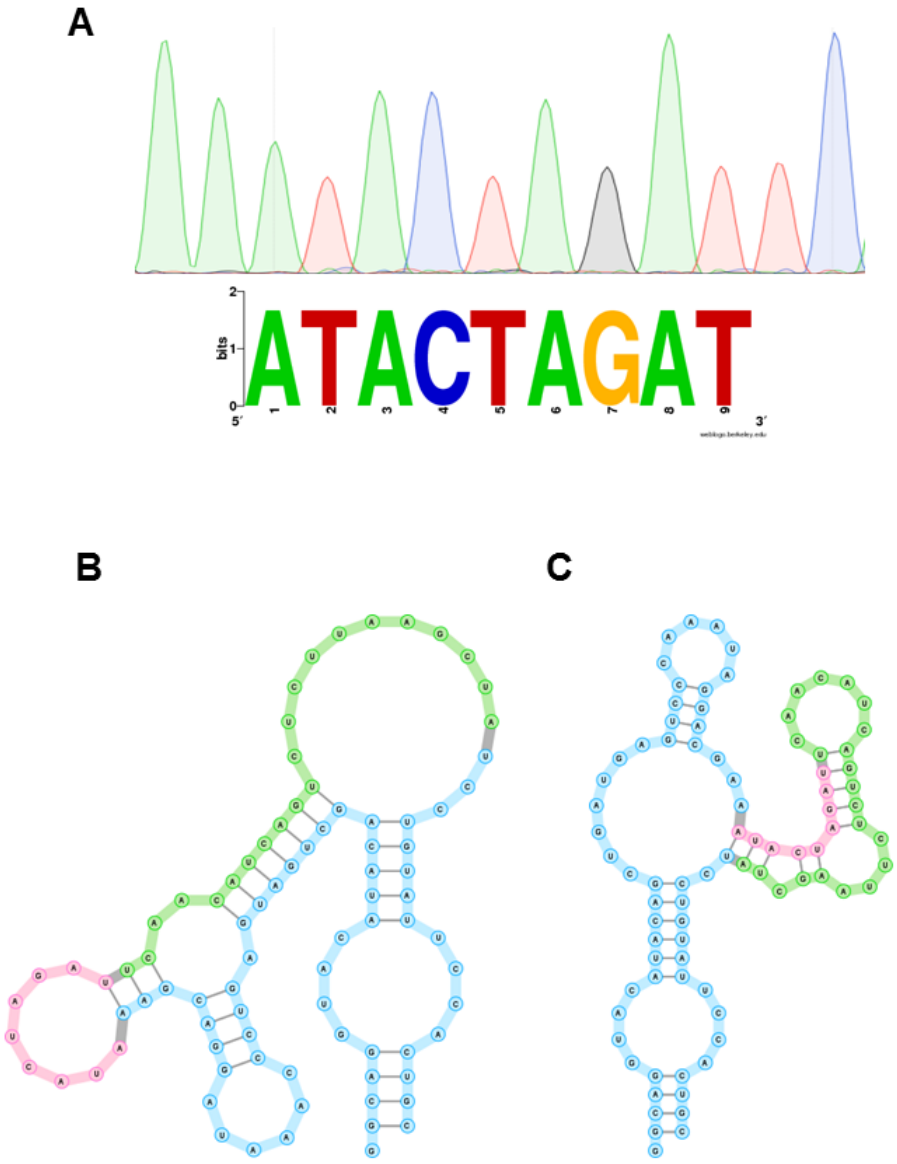



Figure 4.13: Further cycles of OFF state selection eliminate the indel-carrying ribozymes and converge on a single ribozyme

*(A) Sanger sequencing chromatogram of the salient region obtained by pooled sequencing of amplicons produced following the 13$^{th}$ selection cycle (5 cycles of OFF state selection, 5 cycles of ON state selection, and a further 3 cycles of OFF state selection). The matching Weblogo for the initial nine nucleotide randomised region alone is shown below. Mixed peaks were now absent throughout the ribozyme sequence (including the regions of the ribozyme not shown in the above chromatogram), and most remarkably the same applied to the starting randomised region, suggesting the dominant-to-overwhelming presence of only a single ribozyme variant within the pool. (B-C) The structure of the single unique ribozyme (Rz13.1) isolated at the end of the selection process (blue = hammerhead ribozyme backbone, yellow = miR-21 OBS, red = post-selection randomised region). The minimum free energy secondary structure of Rz13.1 (B), and the structure of Rz13.1 when constrained to fold into an active conformation (C).*

tion, and the consequent inhibition of output expression. If so, a comparison between the levels of GFP fluorescence in cells transfected with the Rz13.1-regulated output coding construct (pcDNA-dEGFP-Rz-13.1) and either the empty or miR-21 expression vector should have revealed evidence of allosteric modulation. However, intriguingly, instead of an increase in GFP fluorescence in the presence of miR-21 that may have been expected from an allosterically-inhibited ribozyme, a 4.28-fold reduction in fluorescence was observed (Fig. 4.14A). This was entirely unexpected as Rz13.1 had been identified following a seemingly rigorous and rational mutli-cycle, three stage (OFF, ON and OFF) selection procedure that was aimed at identifying an optimised miR-21-regulated allosteric ribozyme. Additionally, and most remarkably this selection procedure had led to dramatic evolutionary purification and the consequent fluorishing of Rz13.1, with this species alone potentially comprising the sole ribozyme survivor, thus further strengthening the impression that the selection process had been successful.

This paradoxical outcome may have arisen due to the indirect nature of the GFP fluorescence readout used to assess the function of Rz13.1 in contrast to the more direct RT-PCR methods used during the selection process for allosteric ribozymes. Additionally, any miRNA-mediated allosteric modulation of Rz13.1 may have been confounded by miRNA-mediated repression of dEGFP expression. Although the miR-21 target site was designed with mismatches at positions 8-9 to eliminate interference from Ago2-mediated cleavage, repression by alternative mechanisms of translational repression and/or mRNA destabilisation may still have occured.

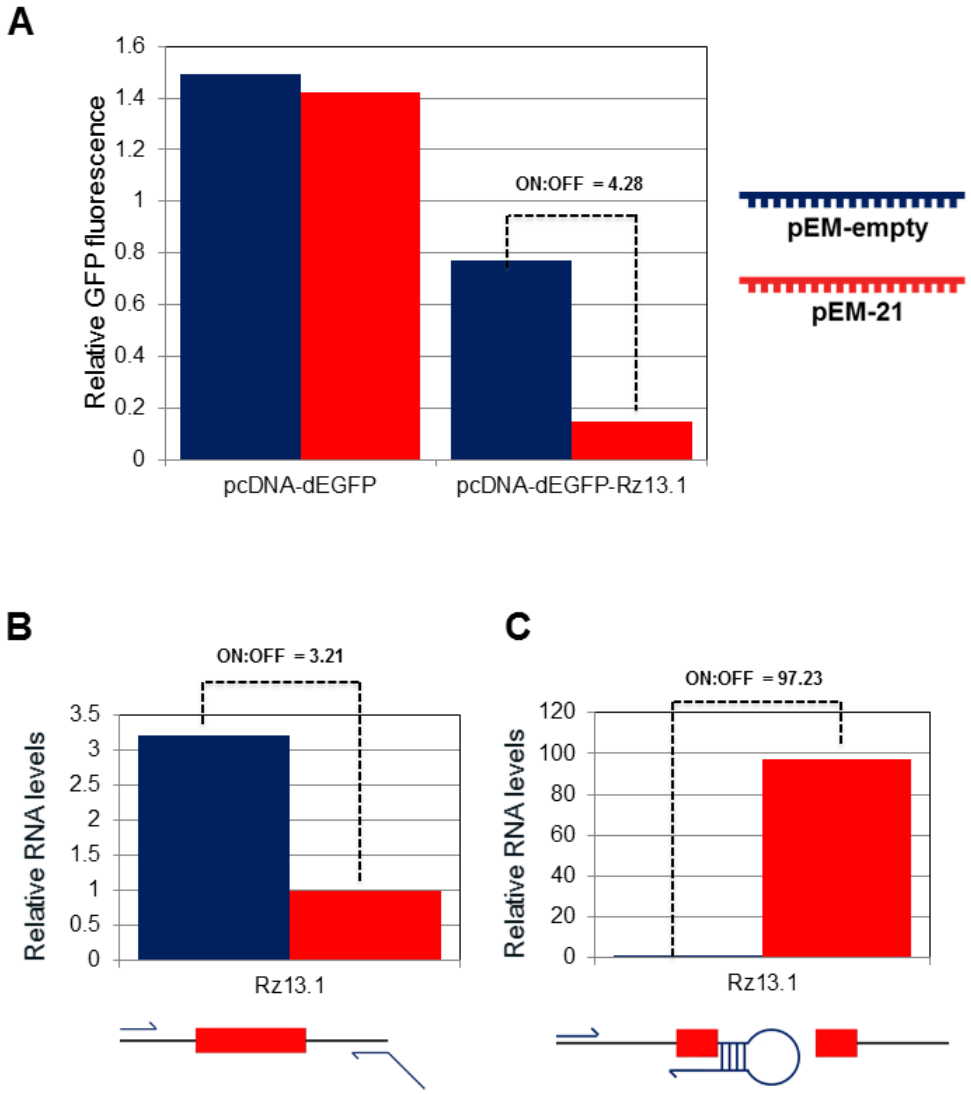



Figure 4.14: Characterisation of the effect of a 3' UTR-located Rz13.1 sequence on transcript abundance and corresponding reporter protein-associated phenotype
*(A) 5 × $10^5$ HEK293T cells were transfected with 500 ng of pcDNA-dEGFP or pcDNA-dEGFP-Rz13.1, in combination with 500 ng of pEM-empty or pEM-21, and 500 ng of piBlue-N1 as the transfection control (n = 1). Flow cytometry was performed 48 h post-transfection to measure GFP and iBlue fluorescence and calculate the relative GFP fluorescence. A Rz13.1-mediated reduction in GFP fluorescence of about 50 % was observed in the presence of the control miRNA. This may have been a function of ribozyme cleavage-triggered degradation of the transcript. Surprisingly though, and contrary to our prediction, GFP fluorescence associated with pcDNA-dEGFP-Rz13.1 was further inhibited in the presence of miR-21 by a factor of 4.28 fold. Continued...*

To further dissect the behaviour in response to miR-21 of Rz13.1 it was necessary to probe its function more directly using RT-qPCR. Exploiting primers used during the selection process it was possible to perform RT-qPCR to specifically measure the levels of either full length uncleaved Rz13.1 or specific ribozyme-mediated cleavage by-products. RNA was extracted from HEK293T cells transfected as described above (pcDNA-dEGFP-Rz13.1 + pEM-empty or pcDNA-dEGFP-Rz13.1 + pEM-21). Since the RT-qPCR primers were specific for sequences contained within the ribozyme itself, it was not possible to perform similar measurements on RNA extracted from the control transfection (pcDNA-dEGFP only). First, the levels of uncleaved Rz13.1 were measured in the absence or presence of miR-21. This was achieved using the ON state selection RT-PCR primers that spanned the cleavage site and were therefore only capable of amplification from an uncleaved ribozyme transcript template. In parallel, RT-qPCR quantification of the housekeeping gene coding for GAPDH was performed for normalisation. This assay revealed a miR-21-dependent reduction in the levels of full length ribozyme by a factor of approximately 3.21 (Fig. 4.14B). This reduction in full length mRNA abundance closely mirrored, and may therefore explain, the reduction in GFP fluorescence levels observed earlier. This result would argue against miRNA-mediated translational repression, since mechanistically this would lead to a reduction in output protein without an associated fall in the level of output mRNA. These data however do not enable differentiation between a reduction in GFP fluorescence caused by miRNA-mediated mRNA destabilisation or that due to an increased level of ribozyme self-cleavage activity. To address this possibility, RT-qPCR was conducted on the same samples using the structured stem-loop RT primers that enabled specific reverse transcription of the Rz13.1 5'-cleavage by-product. This represented the most direct quantitative readout of Rz13.1 self-cleavage activity available in our experimental system. Strikingly, an almost 100-fold increase in the level of the Rz13.1 cleavage by-product was observed by this assay in the presence of miR-21 compared to that

*Figure 4.14 continued: (B - C) RNA was extracted from the cells described above that had been transfected with pcDNA-dEGFP-Rz13.1 and either pEM-empty or pEM-21. RT-qPCR was performed to measure the abundance of either uncleaved ribozyme (B) or the specific 5' cleavage product (C), and the data were expressed as the relative abundance level to that of the transcript of the housekeeping GAPDH gene (average of technical triplicates). A miR-21-dependent reduction in full length ribozyme transcript abundance is observed (B), likely accounting for some or all of the reduction in GFP fluorescence noted. Surprisingly, the abundance of the 5' ribozyme cleavage product underwent an almost 100-fold increase in abundance in the presence of miR-21 as compared to its absence (C), once again suggestive of an increased rate of self-cleavage in presence of miR-21. However, the possibility that an assay or model related artifact has contributed to a false excess measure of cleavage products in presence of miR-21 cannot be ruled out. See text for a detailed discussion of this point.*

found in the presence of the control miRNA (Fig. 4.14C). Hence, these results would strongly support the conclusion that the observed miR-21-triggered reduction in GFP fluorescence associated with the pcDNA-dEGFP-Rz13.1 construct when present in HEK293T cells was due to activation of Rz13.1 cleavage activity in the presence of miR-21. Resolution of the question as to whether or not this activation is allosteric in nature would require the demonstration of direct interaction between Rz13.1 and miR-21. Since mRNA destabilisation operates predominantly through a mechanism of deadenylation, decapping and exonucleolytic degradation in a 5' to 3' direction, it is unlikely to account for these findings, as the exonucleolytic degradation is catalysed by non-site-specific processive enzymes which would therefore be extremely unlikely to result in the accumulation of any one single product of decay possessing a specific cleavage terminus as was being measured by the structured stem-loop primer RT-qPCR assay used in this study.

Collectively, although this would be at direct odds to the selection pressure used to enrich for and identify Rz13.1, the data generated and presented here strongly suggests that Rz13.1 undergoes activation of its catalytic activity in the specific presence of miR-21 . Additionally, the data presented highlights the enormous power of the applied evolutionary approach to distil through a huge number of ribozyme variants to yield a single hit, though clearly our findings are an important and salient reminder of the mysteries of selection and the considerable potential for unintended outcomes!

### 4.2.8 Tracking of the selection landscape by high-throughput sequencing provides useful insights into the nature and outcomes associated with the applied selection pressures

As noted above, the accumulation of ribozymes carrying indels throughout the process of ON state selection may have obscured the evolution and detection of miRNA-regulated allosteric ribozymes. This possibility demanded a strategy to filter out these mutant sequences, allowing the identification of any underlying allosteric ribozymes, even if present at only very low levels. For this purpose, we decided to track the entire process of selection by high-throughput sequencing (HTS) of the ribozyme plasmid pool following each sequential selection cycle. This would enable both identification and exclusion of indel-containing ribozymes, as well as quantitative measurement of the enrichment of individual sequenced ribozyme variants throughout the selection process, thus allowing for inferences to be drawn as to their potential allosteric nature.

A targeted sequencing library was constructed to capture information regarding the full-length sequence of all sampled ribozymes within each of the resulting ribozyme pools follow-

ing each cycle of selection. PCR primers were designed to amplify a ribozyme region centred on the randomised segment of stem III, to generate an amplicon library corresponding to each and every cycle of selection. Each primer contained an extended 5' tail incorporating both a sequencing index to uniquely label amplicons as derived from a particular selection cycle, as well as an Illumina adapter sequence for direct sequencing of amplicons on the MiSeq platform. Following one run of pooled HTS on this platform, the approximately 100,000 resulting reads were processed according to the pipeline shown in Fig. 5.5. Briefly, reads were demultiplexed based on their index, such that these could be assigned to a particular selection cycle. Next, the 5' and 3' invariant ribozyme sequences were trimmed from each read to leave only a sequence corresponding to the randomised region. This trimming was performed with zero error tolerance such that any ribozymes containing indels, SNPs or any other type of mutation within the invariant regions were discarded. Finally, a count of occurrences at each cycle of individual unique randomised sequences was tabulated. These were the data upon which the process of selection was tracked.

During the data analysis process, the extent of the previously observed ON selection-associated "depletion" of indel-free ribozymes and the concurrent accumulation of indel-carrying mutant variants was revealed by the number of reads discarded upon trimming. During the first five cycles of OFF state selection, the invariant ribozyme sequences were highly stable and occurred without error in approximately 99 % of the corresponding reads (Fig. 4.15). These invariant sequences underwent a dramatic "destabilisation" upon application of ON state selection. Thus, after five cycles of such selection pressure, mutation-free variants were found in less than 10 % of sequencing reads (Fig. 4.15). To understand the cause of this phenomenon, it was necessary to identify the types of mutations that had accumulated and whether these were found in a random or systematic distribution. Aligning all sequencing data from each cycle to a reference RzN8 sequence, revealed that the mutations that began to accumulate at cycles 6 - 10 were concentrated very substantially over a contiguous region spanning the 3' terminus of the ribozyme and approximately 80 % of the flanking 3' poly(A) spacer sequence (Fig. 4.16). Three other indel-rich regions were observed to span almost the entire 5' ribozyme segment and an immediate short fragment of the flanking 5' poly(A) spacer sequence, with only two short relatively indel-free intervening stretches. However, none of these three regions exhibited the same mutation density as that spanning the 3' ribozyme-spacer sequence junction. Additionally and most notably, approximately 60 % of sequencing reads arising from the amplicon library corresponding to cycle 10 failed to align with the RzN8 template sequence over a large variable region which at its maximum spanned the entire 3' ribozyme segment and the full-length and beyond of the flanking 3' poly(A) spacer sequence. Closer scrutiny of a random selection of sequence reads that failed to align over this region

revealed the presence of a variable set of large indels impacting on this region and the consequent consistent loss of sufficient stretches of conserved sequence to allow for alignment within this region. The observed non-random distribution of indel locations may suggest that indels mapping to these loci are of biological significance, or alternatively that they have arisen through a technical artefact of selection. Since earlier experiments showed that ribozyme-mediated inhibition of gene expression can be enhanced by spacer sequences (Fig. 4.4A), it is possible that ribozymes containing poly(A) spacer-bearing indels were unintentionally favoured during ON state selection for allosterically inhibited ribozymes. Alternatively, DNA polymerase slippage may have occurred during any and potentially many of the multiple PCR reactions implemented during the recovery and regeneration of the ribozyme pool after each cycle of selection creating a technical artefact. Polymerase slippage is a problem that is particularly likely to occur over homopolymeric regions, such as the poly(A) spacer sequences that were integrated into the constructs. Given that the resulting loss of a poly(A) spacer sequence(s) arising from either or both of these modalities would be predicted to on average result in the reduction in cleavage activity of the transcript-borne ribozyme, it is conceivable that both mechanisms may have contributed co-operatively to the observed accumulation of indels. However, given that the indels observed in this study markedly favoured one poly(A) spacer sequence over the other and that the regions impacted more often than not exhibited relatively sharp boundaries (Fig. 4.16), often external to either poly(A) sequence, it must be postulated that an alternative and as yet unexplained mechanism(s), potentially related to a diverse and variable set of indel-flanking signature sequences, or RNA and/or single-stranded DNA secondary structure, accounted for the non-random distribution of PCR and/or reverse transcription-generated indels produced and selected for during the five cycles of ON state selection.

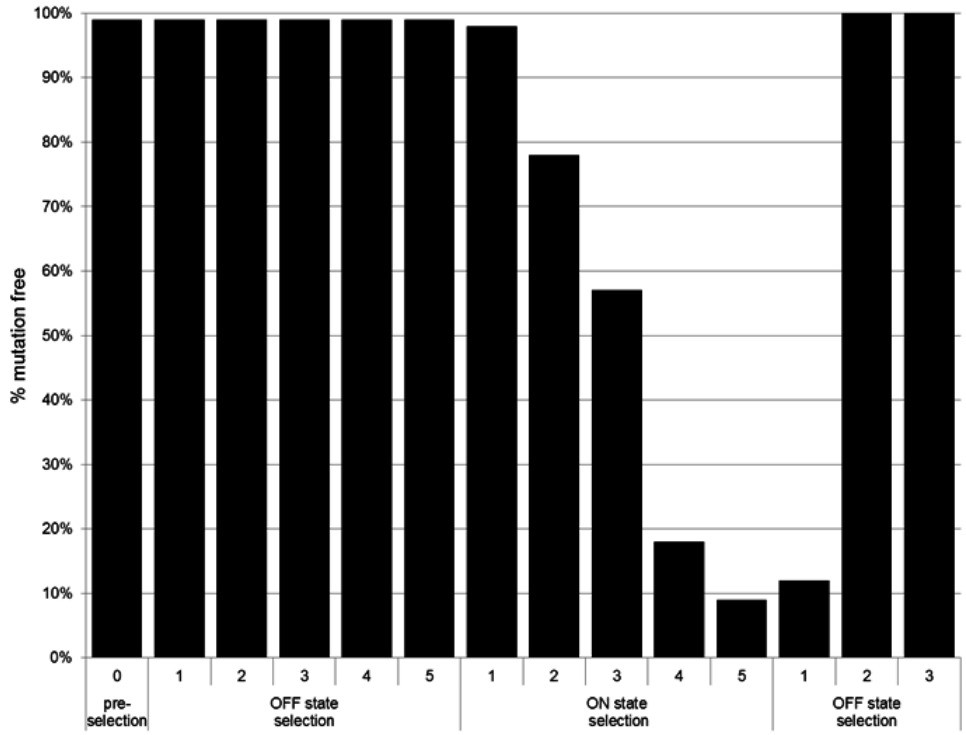



Figure 4.15: The sequences of the ribozyme variants within the pool undergo dramatic and progressive degradation as consequence of the emergence of indels throughout the five cycles of ON state selection

*High-throughput sequencing data obtained following each of the selection cycles was analysed and the percentage of indel-free ribozyme sequences within the pool at the end of each cycle of selection was charted. Up to cycle 5, approximately 98 % of sequences were indel-free. By contrast, indel-bearing ribozyme sequences begin to accumulate substantially and progressively from the very first cycle of the ON state selection until these made up more than 90 % of ribozyme sequences within the pool by the end of the five ON state selection cycles. Even more dramatically, the re-application of the OFF state selection pressure successfully eliminated the vast majority, if not all, of the indel-free ribozyme sequences. Indeed, by the end of only the second of these follow-up OFF state selection cycles, almost 100 % of the sequences obtained were mutation-free.*

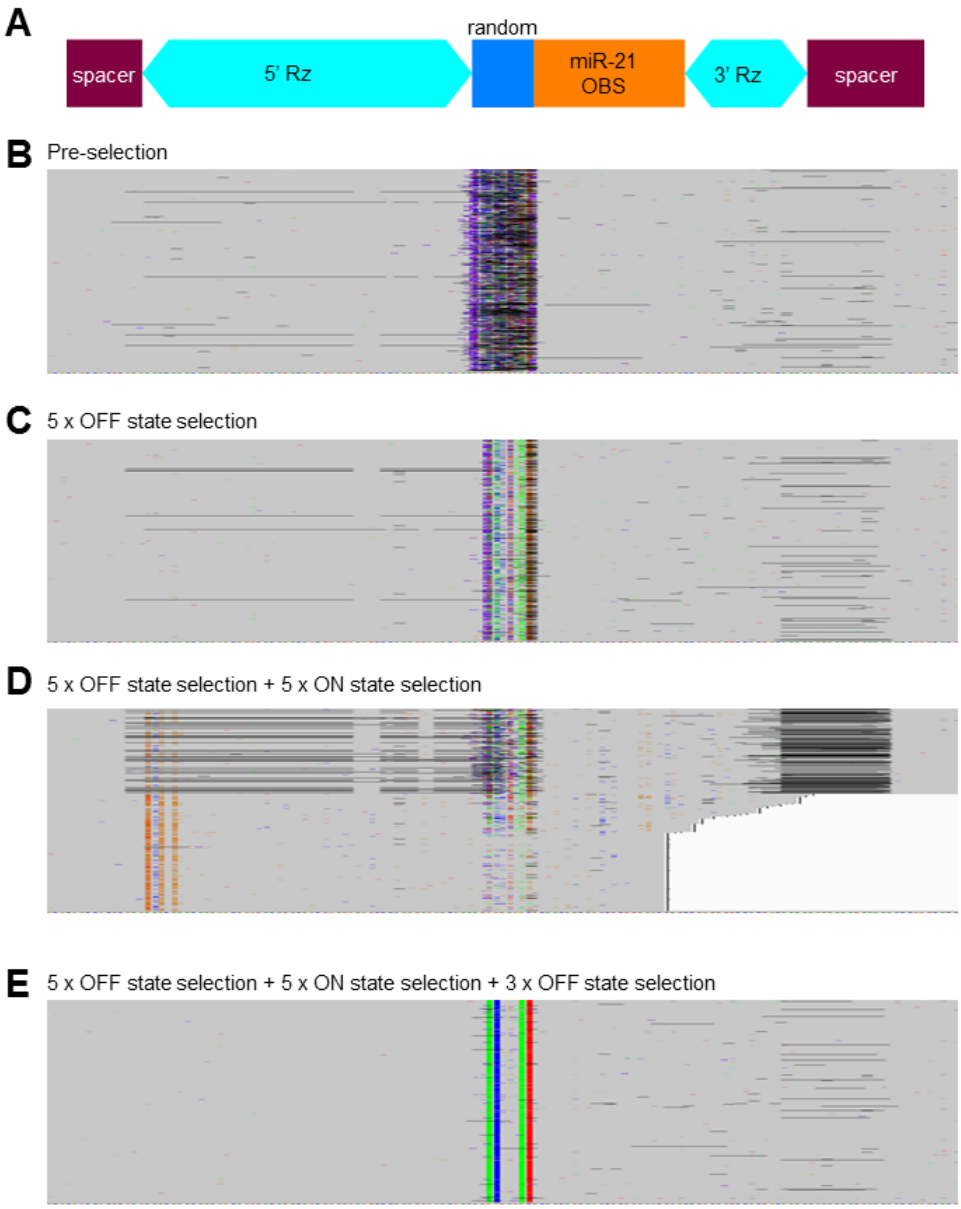



Figure 4.16: Alignment of high-throughput sequencing data broadly identifies the locations and sizes of indels that have emerged during the ON state selection cycles
*(A) Schematic map of the generic ribozyme encoding region present within each of the members of the pcDNA-dEGFP-RzN8 plasmid pool. The schematic has been drawn to scale to correspond directly to the mass sequence alignments shown below. (B - E) High-throughput sequencing reads were mapped to a reference genome which was defined as that of pcDNA-dEGFP-Rz5.7 and visualised using the Integrative Genomics Viewer (Broad Institute) tool. Positions that map to the reference are coloured in grey, missmatches are coloured according to base identified in the individual read (A - yellow, T - red, C - blue, and G - green), insertions are coloured purple, and deletions are coloured black. (B) Variation within the pre-selection pool is almost entirely confined to the randomised region. Continued...*

*Figure 4.16 continued: (C) Following 5 cycles of OFF state selection, a considerable reduction in diversity is observed within the randomised region. Additionally, by comparison to the overwhelmingly grey panels seen in (B), there is a hint of the emergence of a very small proportion of sequences carrying mutations (indels and/or SNPs) outside the region of randomisation, though the detection rates of these mutant sequences may well fall below the threshold of reliability given the inherent error rate associated with HTS. (D) After the $10^{th}$ selection cycle (5 OFF state, followed by 5 ON state selection cylces) diversity within the randomised region would appear to be even further reduced. Much more notably and obvious though is that the frequency of deletions has increased considerably, particularly over a region spanning the majority of the 3' spacer region, but also affecting a short segment of the immediately adjacent 3' portion of the ribozyme. A marked increase in frequency of base substitutions and/or indels was also observed to map to three broadly demarcated regions spanning the entirety, with the exception of short intervening stretches, of the 5' ribozyme segment and a short span of the adjacent 5' spacer region. See further discussion in text for clarification as to possible basis for the "missing alignment data" observed towards the lower right corner of this sequence alignment panel.(E) By the end of the $13^{th}$ cycle (5 OFF state, 5 ON state, and 3 OFF state cycles) deletions have been almost universally eliminated. The vast majority of reads obtained at this point now corresponded to the sequence of Rz13.1 which only differed from the reference sequence at positions 3, 4, 8 and 9 as observed by the corresponding coloured rows within randomised region against the broader and once again overwhelmingly grey background.*

In an attempt to isolate allosteric ribozyme variants from amongst the pool now dominated by indel-containing sequences, further OFF state selection was performed (Fig. 4.8). The result of the selection process over these three final cycles is clearly evident as demonstrated by the dramatic increase in the frequency of ribozymes containing mutation-free invariant regions from cylce 11 to cycle 12 (Fig. 4.15). Although this final OFF state selection step was performed to filter out indel-containing sequences, it came at the potential expense of a reduction in diversity of the sequences being subject to selection, and indeed eventually left only the single Rz13.1 sequence. Given that it is possible to bioinformatically filter out indel-containing ribozymes from the sequencing data, the end of the $10^{th}$ cycle was chosen as the end point for subsequent analysis in an attempt to regress the process and once again maximise the starting diversity so as to increase the chance of identifying potential allosteric ribozyme sequences that may have been selectively enriched over cycles 1 - 10.

Having tabulated the number of occurrences of each ribozyme variant at the end of each cycle of selection, it was possible to identify those sequence variants that underwent a progressive process of enrichment. Such ribozymes presumably satisfied our criteria of selection: efficient self-cleavage activity in the absence of its cognate miRNA, and the reduction of this catalytic activity through allosteric modulation in the presence of the cognate miRNA ligand. Ribozyme variants of interest were defined as those that represented greater than or equal to 2 % of all variants found within the final pool, following a sequential process of five cycles of OFF state selection and five cycles of ON state selection. This would represent approximately 1300-fold enrichment over the predicted frequency with which a single ribozyme variant from the RzN8 pool should occur by chance alone. Ten unique ribozyme variants were found to meet these selection criteria. Their frequencies within the original starting pool and each of the subsequent pools following five cycles of OFF state selection and five cycles of ON state selection was plotted as shown in Fig. 4.17. Cumulatively, these 10 ribozymes constitute approximately 4.4 % of the original starting pool. However, over the course of positive selection these 10 ribozymes came to account for approximately 88.1 % of all remaining ribozyme variant sequences within the pool. The incremental and near linear manner in which these sequences were enriched over the course of five cycles of OFF state selection is supportive of the notion that the increasing stringency that was intended to be applied appeared to be effective leading to the selection of ribozymes exhibiting even greater catalytic activity. While these 10 ribozyme variants remained highly overrepresented during the course of ON state selection, the fraction of the pool that these occupied decreased slightly, though this collective picture masked significant variations in changes in frequency between individual members Three of the 10 ribozymes exhibited greater than 2-fold increases in frequencies at cycle 10 as compared to cycle 5, four displayed the reverse trend and three demonstrated differences

that were below this empirically nominated threshold. These seemingly random changes in the frequencies of the 10 ribozyme species over the course of five cycles of negative selection suggests that the latter was either entirely ineffective or provided insufficient stringency to allow for progressive enrichment over the ON state selection cylces (cycles 6 -10) of the species of interest.

Interestingly, four of the most highly represented ribozymes identified by RNAseq had previously been isolated as single clones following the $5^{th}$ cycle of OFF state selection (Fig. 4.10, Rz5.1, Rz5.3, Rz5.6 and Rz5.7). Rz5.7, the most enriched ribozyme at the end of cycle 10, accounted for approximately 27 % of the final pool and was therefore chosen as the prime miR-21-responsive allosteric candidate for further functional analysis.

### 4.2.9 Functional analysis of the most enriched ribozyme sequence present at the end of the ON state selection fails to reveal detectable evidence of allosteric modulation

Having identified Rz5.7 as a potential allosteric ribozyme candidate from HTS selection, the next step entailed its functional characterisation. Additional control constructs were generated to facilitate an in depth analysis. Rz5.7MM was generated through the introduction of a mismatch between miR-21 and the ribozyme OBS at positions 4-5. Such a mismatch has been shown by Wee *et al*., to nearly abolish the interaction between a miRNA and its target and therefore would serve as a control for the effects of miR-21 binding [95]. Additionally, Rz5.7Mut was generated through targeted mutagenesis at position 14 in the ribozyme to introduce an A14G base substitution, which has been shown to prevent catalytic activity of hammerhead ribozymes [63, 73]. Constructs containing Rz5.7, Rz5.7MM and Rz5.7Mut surrounded by poly(A) spacers in the 3' UTR of the dEGFP gene were then generated (pcDNA-dEGFP-Rz5.7, pcDNA-dEGFP-Rz5.7MM and pcDNA-dEGFP-Rz5.7Mut). pcDNA-dEGFP, as the empty 3' UTR control, and each of these three plasmids were then transfected into HEK293T cells along with a miRNA expression vector (pEM-empty or pEM-21) and piBlue-N1 as a transfection control.

Firstly, the catalytic activity of Rz5.7 was considered. In the absence of miR-21, a robust inhibition of GFP fluorescence of approximately 50 % was observed for constructs containing either Rz5.7 or Rz5.7MM in comparison to the pcDNA-dEGFP control (Fig. 4.18A). This inhibition was substantially but not entirely attenuated for the construct bearing Rz5.7Mut, which would be consistent with the predicted lack of ribozyme-mediated cleavage activity associated with this mutant (Fig. 4.18A). Additionally, the observation that GFP fluorescence from the Rz5.7Mut-containing construct does not reach the level attained by the con-

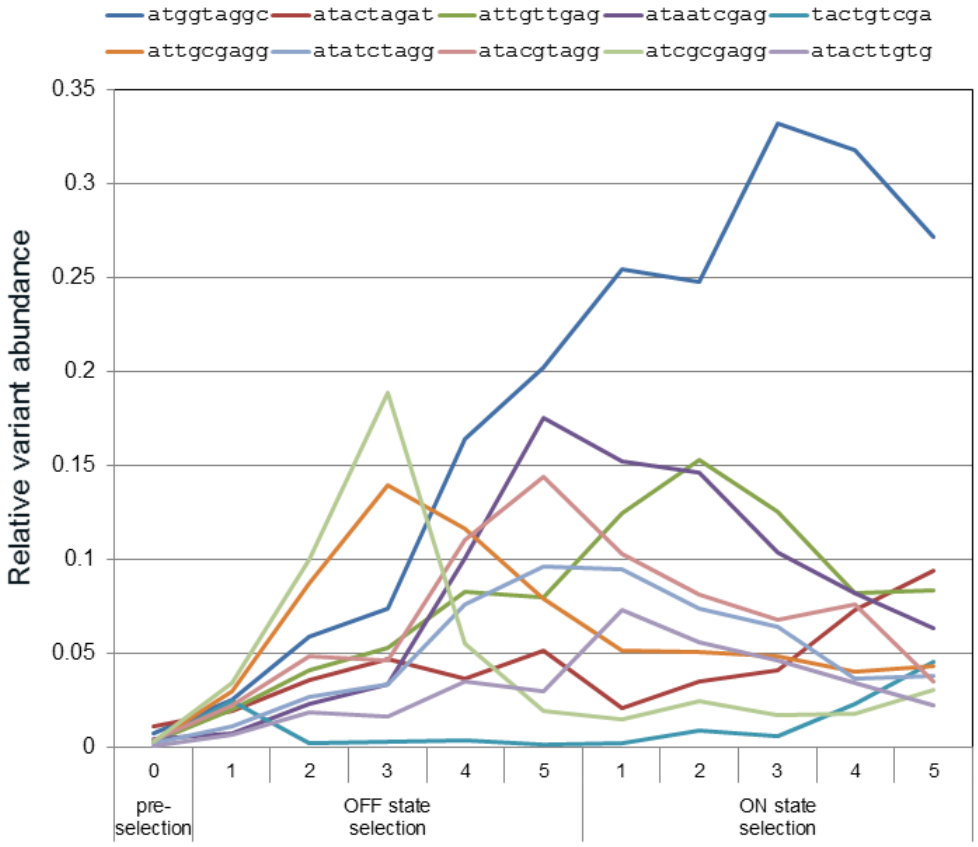



Figure 4.17: Bioinformatic based identification of the most highly enriched indel-free ribozyme sequences present at the end of the $10^{th}$ selection cycle
*The relative abundance of each of the 10 indel-free ribozyme sequence variants found to comprise $\geq$ 2 % of the final pool were traced over the entire selection process. The sequence of the cognate randomised region is displayed in the key. Rz5.7 made up greater than 25 % of the final pool and was therefore chosen for further analysis.*

trol plasmid (pcDNA-dEGFP), is consistent with published findings relating to the presence of a ribozyme sequence within the 3' end of a transcript indicating that a certain degree of cleavage-independent inhibition occurs. The reductions in GFP fluorescence observed between constructs harbouring the mutant ribozyme (Rz5.7Mut) compared to those carrying active ribozymes (Rz5.7 or Rz5.7MM) are therefore likely to represent the downstream consequence of ribozyme-mediated self-cleavage (Fig. 4.18A). Together these data suggests that Rz5.7 possess self-cleavage activity and hence supports its observed enrichment through the five cycles of OFF state selection.

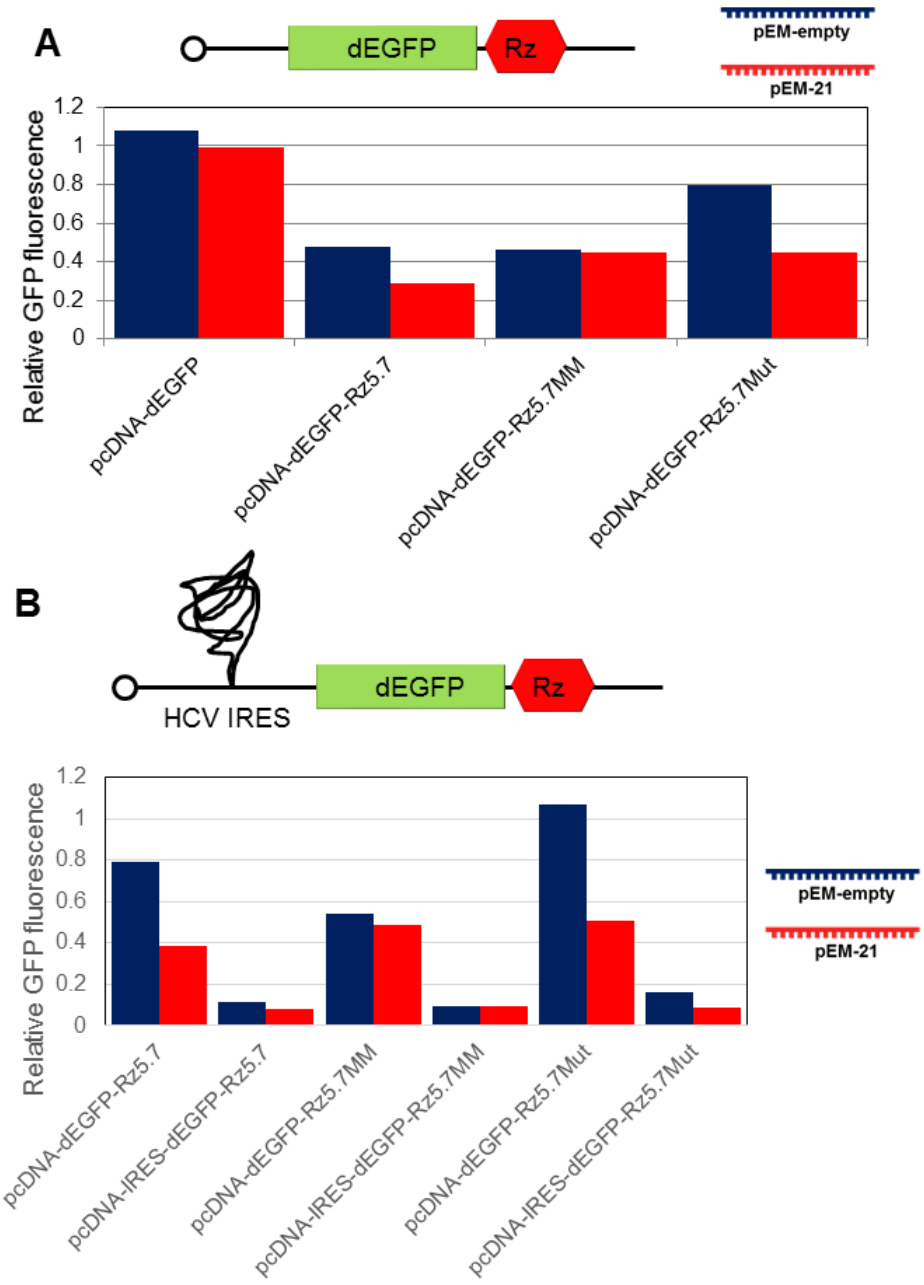



Figure 4.18: Characterisation of the effect of a 3' UTR-located Rz5.7 sequence on reporter protein-associated phenotype

*(A) 5 × $10^5$ HEK293T cells were transfected with 500 ng of pcDNA-dEGFP, pcDNA-dEGFP-Rz5.7, pcDNA-dEGFP-Rz5.7MM, or pcDNA-dEGFP-Rz5.7Mut in combination with 500 ng of pEM-empty or pEM-21, and 500 ng of piBlue-N1 as the transfection control (n = 1). Flow cytometry was performed 48 h post-transfection to measure GFP and iBlue fluorescence and calculate the relative GFP fluorescence. Given that Rz5.7Mut was expected to lack catalytic activity, the modest inhibition of GFP fluorescence associated with pcDNA-dEGFP-Rz5.7Mut in the absence of miR-21 was most probably attributable to a cleavage-independent inhibitory effect. Following on this logic, the even greater inhibition of GFP fluorescence associated with constructs encoding Rz5.7 or Rz5.7MM in the absence of miR-21 was presumed to result from ribozyme cleavage and consequent accelerated transcript degradation. Continued...*

*Figure 4.18 continued: Moderate inhibition of GFP fluorescence in the presence of miR-21 relative to its absence was observed in relation to the two Rz5.7-encoding constructs bearing cognate binding sites (pcDNA-dEGFP-Rz5.7 and pcDNA-dEGFP-Rz5.7Mut), but was barely apparent regards to the Rz5.7MM construct that carried an interaction-minimising mismatched miR-21 binding site. This effect was presumed to result from miR-21-mediated repression of expression of GFP and thus could conceivably have obscured any potential allosteric modulation of ribozyme activity. The latter activity were it to occur on its own would be expected to result in increased GFP fluoresecence in the presence of miR-21 versus its absence. (B) $5 \times 10^5$ HEK293T cells were transfected with 500 ng of pcDNA-dEGFP-Rz5.7, pcDNA-dEGFP-Rz5.7MM, or pcDNA-dEGFP-Rz5.7Mut, or the equivalent IRES-containing constructs (pcDNA-IRES-dEGFP-Rz5.7, pcDNA-IRES-dEGFP-Rz5.7MM and pcDNA-IRES-dEGFP-Rz5.7Mut), in combination with 500 ng of pEM-empty or pEM-21, and 500 ng of piBlue-N1 as the transfection control (n = 1). Flow cytometry was performed 48 h post-transfection to measure GFP and iBlue fluorescence and calculate the relative GFP fluorescence. A major IRES-dependant reduction in GFP fluorescence was observed for all constructs, regardless of whether miR-21 was present or not, consistent with the known reduction in translational efficiency associated with initiation from this element. Otherwise the same trends as above were observed suggesting that miRNA-mediated repression continued to dominate and potentially obscure any contribution that may have arisen from miR-21-mediated allosteric modulation of the activity of the Rz5.7 ribozyme.*

Next, the potential allosteric modulation of Rz5.7 was assessed by analysis of its capacity to increase GFP fluorescence in the presence of miR-21. GFP fluorescence from constructs lacking the ability to interact with miR-21 (pcDNA-dEGFP (no target site), and pcDNA-dEGFP-Rz5.7MM (mismatched target site)) showed negligible differences in the absence or presence of miR-21 (Fig. 4.18A). In contrast, both the Rz5.7 and Rz5.7Mut containing constructs demonstrated a miR-21-dependent inhibition in GFP fluorescence levels of approximately 40 % (Fig. 4.18A). Considering that Rz5.7Mut lacks any catalytic activity and is therefore incapable of allosteric modulation, this reduction is most likely the result of miRNA-dependent silencing, which may also be assumed to have equally affected the Rz5.7-containing transcript.

Given the magnitude of the observed likely miR-21-dependent silencing, it is plausible that this miRNA-mediated inhibition of the Rz5.7-bearing transcript masked potential underlying allosteric modulation of this ribozyme. In an attempt to overcome this problem new sensor constructs were generated in which translation of dEGFP was driven from a Hepatitis C virus internal ribosome entry site (HCV IRES) element. Translational initiation from an HCV IRES proceeds via the direct recruitment of the 40S ribosomal subunit and the involvement of eIF3, eIF2, eIF5, eIF5B and the initiator tRNA, and has been experimentally demonstrated to be insensitive to miRNA mediated repression [113]. The GFP fluorescence level associated with all HCV IRES driven transcripts was markedly reduced compared to that observed from equivalent transcripts capable of translation via the canonical 5' cap-dependent initiation mechanism. However, the overall trend in behaviour of HCV driven transcripts containing Rz5.7, Rz5.7MM or Rz5.7Mut and their effect on GFP fluorescence levels in the absence or presence of miR-21, mirrored that observed earlier from equivalent 5' capped constructs (Fig. 4.18B). Thus, despite HCV IRES mediated translation, both Rz5.7 and Rz5.7Mut containing transcripts appeared susceptible to miR-21-mediated inhibition and therefore it was not possible to discern any potential added contribution made by allosteric ribozyme modulation in relation to the Rz5.7-bearing construct.

## 4.3 Discussion

### 4.3.1 The two selected ribozymes display prerequisite autocleavage activity but do not exhibit detectable miRNA-mediated allosteric modulation

The selection strategy pursued in this chapter identified two candidate allosteric ribozymes that were subjected to further characterisation. Rz 13.1 was identified at the end of 13 selec-

tion cycles (5 OFF state, followed by 5 ON state, followed by 3 OFF state cycles). By contrast, Rz 5.7 was bioinformatically identified as the most abundant ribozyme species following 10 cycles of selection (5 OFF state, followed by 5 ON state cycles). However, rather disappointingly, both Rz13.1 and Rz5.7 failed to demonstrate the desired miR-21-mediated allosteric inhibition.

Particularly puzzling was the paradoxical and seemingly robust miR-21-mediated activation of Rz13.1 cleavage activity (Fig. 4.14C). Oligonucleotide-regulated allosteric ribozymes that undergo both activation and inhibition of cleavage activity have been previously described [64] and therefore it is possible that the observed result had occurred via miR-21 mediated activation of Rz13.1 cleavage. Alternatively, the result may represent a technical artefact of the RT-qPCR process. Total RNA prepared from cells transfected with pcDNA-dEGFP-Rz13.1 and pEM-21 would contain, in addition to all other components shared in common with cells transfected with pcDNA-dEGFP-Rz13.1 and pEM-empty, abundant amounts of miR-21. It is possible that miR-21 itself may have served as a supplementary primer, alongside the intended structured stem loop RT primer, during reverse transcription of these ribozyme transcripts. This however should not have affected final qPCR quantification which used a reverse primer specific to a unique sequence in the structured stem loop primer which would have been appended to the 3' end of cDNA originating from reverse transcription driven solely by this structured stem loop primer. Alternatively, and entirely speculatively, perhaps the presence of miR-21 stablised, potentially by virtue of altered RNA secondary structure, the duplex formed between the structured stem loop primer and the ribozyme cleavage product or even the intact ribozyme itself, thus greatly enhancing the efficiency of the reverse transcription step and potentially reading-off non-cleaved ribozyme sequences as well. Given the dramatic and entirely counter-intuitive result obtained further exploration of this finding would be essential prior to being able to be definitive as to RT-qPCR-alluded miR-21-mediated activation of Rz13.1 cleavage activity. Indeed, one key control would be to repeat the experiment using a construct bearing Rz13.1Mut which carried the ribozyme-inactivating SNP in the presence of pEM-21 or pEM-empty. Theoretically, as Rz13.1Mut would not exhibit any self-cleavage activity, cleavage activity measurements by this same RT-qPCR assay under both conditions would result in similar and neglible cleavage product counts. If however, miR-21 was able to stablise the annealing of the structured stem loop primer to the uncleaved ribozyme sequence, this would have resulted in loss of specificity of the assay and greatly elevated "cleavage product" counts but only in the presence of miR-21. Indeed, potentially such a result may mirror that obtained in this study with the Rz13.1-bearing construct. Considering the matter from an entirely different perspective, it is possible, though very unlikely, that the elevated level of miR-21 led to altered intracellular biochemistry and an entirely coincidental increase in

ribozyme cleavage activity. There are no previous published reports to support either of these speculative models, further emphasising the importance of further experimentation as a means of exploring this phenomenon. Besides the experiment suggest earlier with Rz13.1Mut and potentially Rz13.1MM (OBS mismatched at positions 3-4), further studies should also include investigation of the amounts of uncleaved and cleaved Rz13.1 using alternative techniques such as Northern blotting or other appropriately designed *in vitro* assays.

Intriguingly, Rz5.7 was also identified following random characterisation of a small number of clones as one of the candidates present following the first five cycles of selection (5 OFF state cycles). This was consistent with the demonstrated reduction in GFP fluorescence in cells transfected with pcDNA-dEGFP-Rz5.7 compared to pcDNA-dEGFP, and was likely attributed to the cleavage activity or Rz5.7 in the absence of miR-21 (Fig. 4.10). Rz5.7 was again identified following 10 cycles of selection. Bioinformatic analysis revealed an enrichment of ~7.8-fold over the first 5 selection cycles (OFF state), and a further enrichment of only ~1.3-fold over the following five selection cycles (ON state) (Fig. 4.17). The observed weaker enrichment through the ON state selection process would suggest that Rz5.7 only weakly fulfilled the selection criterion (miR-21-mediated allosteric inhibition of ribozyme activity) intended to be implemented during this stage of selection. This may explain the observed action of Rz5.7 in the absence of miR-21 in relation to inhibition of construct-driven GFP fluorescence, a trait intended to be selected over the first five cycles, and yet no demonstrable miR-21-mediated allosteric regulation, the key feature that was meant to have been selected over the last five cycles. Furthermore, characterisation of Rz5.7 and the mutant controls Rz5.7Mut and Rz5.7MM suggested that constructs containing these ribozymes were subject to miR-21-mediated repression (Fig. 4.18). This posed a major problem in that any miRNA-mediated allosteric modulation of Rz5.7 activity, leading to an expected increase in GFP fluorescence, may have been masked by the opposing effect of miRNA-mediated inhibition of the transcript output (GFP fluorescence), exerted by reduced translational efficiency and/or decreased mRNA stability. A more direct method to quantify Rz5.7 cleavage activity in the presence or absence of miR-21, such as an RT-qPCR or alternative ribozyme-targeted *in vitro* assay could be used to avoid this pitfall in the future.

### 4.3.2 The catalytic activities of ribozyme variants seemingly selected for enhanced cleavage rates by RT-PCR were lower than that of the parent wild-type ribozyme

After five cycles of selection using the RT-PCR approach, nine different candidate ribozymes were isolated, all of which inhibited output expression, relative to that associated with the

pcDNA-dEGFP control, when coupled to the matching dEGFP transcript, (Fig. 4.10). Although these candidate ribozymes were competent to catalyse self-cleavage, their inhibitory effects on output expression were much lower than that observed using the wild type hammerhead ribozyme. The lower rate of catalysis displayed by these chosen candidates may be due to a number of reasons. The level of stringency applied during selection was determined by the duration of time from transfection to RNA purification. It was postulated that the shorter this time interval, the more active a ribozyme must be to generate sufficient cleavage products, which could then be recovered by RT-PCR. In practice, the maximal stringency achievable was determined by experimental limitations related to the delay between transfecting cells and the subsequent transcription of ribozyme-containing transcripts by the host cell machinery. As reported in similar previous studies, the most stringent selection utilised in this study corresponded to RNA extracted at 1.5 h after transfection [150]. Nonetheless, application of this brief time interval may still not have imposed sufficient stringency and thus the experimental conditions utilised may have precluded selection of the most active ribozyme variants.

A potential strategy that could be employed to overcome this experimental limitation is using the Tet-On inducible system to drive ribozyme expression. In this case, cells would be transfected with a plasmid by which expression of the ribozyme-containing transcript was under the transcriptional control of a tetracycline response element (TRE) promoter. Simultaneous transfection of these cells with a constitutive expression vector encoding the reverse tetracycline-controlled transactivator (rtTA) gene would provide an inducible transcription activator capable of stimulating expression from the TRE promoter. In the absence of doxycycline, expression of the ribozyme would be blocked. Upon addition of doxycycline, rapid induction of ribozyme expression would occur. This is due to the fact that the system is maintained in a dormant but "poised" state in the absence of doxycycline. In this OFF state, the rtTA protein is expressed at high levels but cannot bind to the TRE and consequently cannot drive expression of the downstream target gene. Doxycycline rapidly binds rtTA promoting its association with any TRE promoter and the induction of expression of downstream target genes. This system could be used for RT-PCR selection of ribozymes in which selection stringency for the most active ribozymes can be tightly controlled as a function of time between addition of doxycycline and RNA extraction. Previous studies have shown that TRE-initiated transcripts can be detected shortly before 20 minutes post doxycycline induction using this system, hence allowing for far more stringent selection of highly activity ribozymes than was the case in this study [181].

### 4.3.3 Theoretical additional mechanisms and associated modules by which miRNA-mediated repression could potentially be minimised

In the preceding work we have attempted to overcome miRNA-mediated repression that occurs as a result of translational repression. In order to achieve this, we built matching constructs carrying the HCV IRES structure, which can initiate translation in a non-canonical manner and critically which had also previously been reported to confer "immunity" to miRNA-mediated repression [113]. Unfortunately, besides leading to markedly lower levels of GFP fluorescence this measure was ineffective at dampening the confounding miRNA-mediated repression effects observed and therefore the search for additional mechanisms and/or strategies that may lead to transcripts which are insensitive or less prone to miRNA-mediated repression remains warranted.

mRNA destabilisation has been suggested as the dominant mechanism by which miRNAs elicit target repression within cells [101]. This pathway has been proposed to proceed via deadenylation and decapping followed by exonuclease degradation from the 3' end. This has raised the question as to whether a 3' poly(A) transcript tail is required for miRNA-mediated mRNA destabilisation. Wilusz *et al.* have described an RNA triplex structure found at the 3' terminus of the non-coding RNA produced from the metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) locus where it appears to functionally substitute for a poly(A) tail [182]. This unusual structure arises as a result of Hoogsteen hydrogen bonds that form between a U rich motif and the major groove of an A-U rich helix at the 3' end of MALAT1. The function of the MALAT1 triplex was investigated by cloning its encoding sequence downstream of a reporter gene, to be transcribed instead of a poly(A) tail. The resultant transcript was both stable and efficiently translated supporting the role of the triplex in substituting for functionality much more typically associated with a poly(A) tail [182]. Alternatively, an RNA motif found at the 3' terminus of histone-encoding transcripts, the so called histone stem-loop, has been shown to function as a poly(A) tail substitute and is thus yet another potential means by which to avoid poly(A) tail associated deadenylation-dependent miRNA-mediated repression [183]. One or potentially even both of these modification may be incorporated into future ribozyme-containing sensor constructs in an attempt to overcome miRNA-mediated repression.

It may be that combinations of several such modules are required to fully overcome miRNA-mediated inhibition. Overcoming either translational inhibition using an IRES element or some other "translation enhancer module", or mRNA destabilisation using poly(A) substitutes, may alone be insufficient to adequately suppress miRNA-mediated inhibition effects. However, when combined these modules or variants thereof could potentially enable the generation of a transcript that is insensitive to repressive effects consequent upon binding to its

cognate miRNA.

### 4.3.4 Lentiviral transduction and cell viability selection pressure

A major limitation that affected the RT-PCR selection strategy resulted from the accumulation of ribozyme-inactivating indels during the OFF state selection cycles. To minimise this problem, it would be greatly beneficial to use a polymerase enzyme with a much higher replicative fidelity and many fewer cycles of amplification. Both these benefits could in principle be achieved by stable transduction and integration of ribozyme-encoding transcripts into the cellular genome, and their consequent amplification by the endogenous DNA polymerase.

The cell-based nature of this approach would however necessitate a new strategy to select for ribozyme catalytic activity, such as cell viability selective pressure. By coupling the ribozyme sequence to the expression of either an essential or a toxic gene, ON state or OFF state selection could theoretically be carried out by targeted enrichment based on cellular fitness. Candidate genes for ON state selection could include an antibiotic resistance gene, essential for viability in the presence of its cognate antibiotic. OFF state selection on the other hand could conceivably be achieved through the use of toxic gene, such as that encoding for the herpes simplex virus thymidine kinase (HSV-tk) which converts the prodrug ganciclovir to a toxic metabolite. An additional advantage of such an approach would be the ability to finely tune the level of selection stringency through modulation of the concentration of either the antibiotic or the pro-drug. Besides overcoming the huge challenges associated with sub-optimal *in vitro* based nucleic acid replication fidelity, these tools would enable the imposition of even more highly stringent selection pressures than utilised in the present study, thus also overcoming the barrier that led to poor catalytic activity amongst RT-PCR selected ribozymes. However, as with the presently applied approach, key to the success of this latter strategy would be the ability to start with a very large and diverse set of lentiviral transductants, each carrying a construct bearing a distinct ribozyme variant. Indeed, if the aim was to screen and subject to selection ribozyme variants bearing a nine nucleotide randomised region, the number of lentiviral transductants in the preselection pool would need to equal 262,144 ($4^9$), assuming each and every transductant bears a distinct ribozyme variant. As the probability of the latter is vanishingly small, it will be necessary to screen from a starting pool of approximately half a million tranductants or more so as to be confident of capturing the vast majority of randomised ribozyme variants within the actual experimental starting pool.

# Chapter 5

# RNAseq-based *in vitro* selection for the construction of miRNA-responsive allosteric ribozymes

## 5.1 Introduction

High-throughput methodologies, such as those applied in the previous chapter, have great utility in the design of functional RNA devices. When run to completion, exponential enrichment screens typically provide a binary output of performance as either a winner or loser sequence emerges. However, this results in the loss of quantitative information and the dismissal of all sequences that fall below an arbitrary threshold. In addition, such endpoint screens are susceptible to "cheater sequences", which may end up dominating the pool despite not displaying the desired characteristics. To overcome these limitations, RNAseq has recently been employed to provide a quantitative snapshot of the cleavage activity of all members within a diverse pool of random ribozyme sequences [184]. The ability to quantify the ON:OFF ratio (cleaved:uncleaved ratio, or *vice versa* as appropriate) of every individual ribozyme within a rich and diverse population at an RNA level enables identification of active variants without the need to biochemically isolate these over multiple rounds of exponential enrichment. A similar approach has been pioneered by the Smolke laboratory in which random ribozyme sequences were placed into the 3' UTR of the EGFP gene to enable their stratification based on the ability to repress the expression of the reporter protein [152]. Subsequently, RNAseq was used to establish the sequence identity of candidate ribozymes meeting the desired characteristics. When applied to pools of candidate allosteric ribozymes constructed by the fusion of a ligand binding aptamer to a self-cleaving ribozyme via short stretches of randomised

sequences, comparative analysis of cleavage rates in the presence or absence of the cognate ligands enabled the identification of functional allosteric variants.

Here we sought to apply an RNAseq selection technique for the identification of miR-21-responsive allosteric ribozymes. We have applied this technique for the first time in living mammalian cells and shown the ability to identify candidate ribozymes displaying differential activity in the presence of a cognate input miRNA. Furthermore, we have incorporated several technical adaptations which have successfully minimised intrinsic system noise. However, despite the apparent ability of several chosen candidate allosteric ribozymes to regulate their autocleavage rate in response to miR-21 exposure, this was not reflected in their ability to enhance the expression of the dEGFP output from their respective ribozyme coupled reporter constructs.

## 5.2 Results

### 5.2.1 Rationale underlying the design of the three ribozyme pools intended for RNAseq selection

Three different ribozyme pools were generated and subjected to selection by RNAseq. All pools were based on the hammerhead ribozyme design and modified to contain a miR-21 binding site incorporated at different loci, in addition to a short random linker region to generate diversity.

Rather than selection through a process by which the starting pool was whittled down to a specific population (Chapter 4), the RNAseq selection performed as part of this chapter considered all species within the starting pool in the presence or absence of the relevant ligand (miR-21). Theoretically, with this strategy direct comparison between the abundance of each species under conditions with and without the cognate ligand enables subsequent bioinformatics based identification of candidate allosteric ribozymes. To achieve this, all ribozyme variants, regardless of activity, must still be adequately represented within the final RNAseq dataset. Therefore, the size (number of distinct ribozyme variants) of the initial randomised ribozyme pool must be considered by taking account of the number of intended replicate samples, experimental conditions to be studied and expected sequencing platform read count capacity. Given the stochastic nature of the PCR process involved in preparation of RNAseq libraries and the resultant technical noise, sufficient read depth is required across each individual species to draw statistically meaningful conclusions. Hence, we reasoned that a pool size of ~5000 variants should result in an average read depth of ~1000 counts across all ribozyme variants, when explored under four experimental conditions, following a single MiSeq exper-

iment capable of generating up to 20 million unique sequence reads per run. To fall within this limit, all three randomised ribozyme pools generated and characterised in this chapter contained a short contiguous region of six random nucleotides and therefore each had an individual total pool size of 4096 variants. Functionally optimised allosteric ribozymes have been identified using a similar RNAseq method from a pool of 256 variants of the HDV ribozyme [184], as well as by FACS-based pre-selection in bacteria followed by screening in mammalian cells from a pool of 1024 variants of the hammerhead ribozyme [151]. Based on these previously successful approaches, we anticipated that our intial pool size of 4096 variants should provide sufficient diversity, at least with regards to one or other of the three distinct pools to be investigated.

The first two pools, RzSeq1 and RzSeq3, were designed to undergo allosteric modulation through the disruption of stem III, in a manner similar to RzN8 pool ribozymes (Chapter 4; Fig. 5.1, A and C). A full-length miR-21 binding site, mismatched at positions 8 and 9 to prevent Ago2 cleavage, was incorporated into the 5' or 3' arm of stem III in RzSeq1 or RzSeq3, respectively (Fig. 5.1, A and C). RzSeq3, like RzN8, contains known sequences immediately 5' of the ribozyme cleavage site and thus was amenable to selection with a structured stem loop primer for the hugely preferencial reverse transcription of only site-specifically cleaved transcripts.

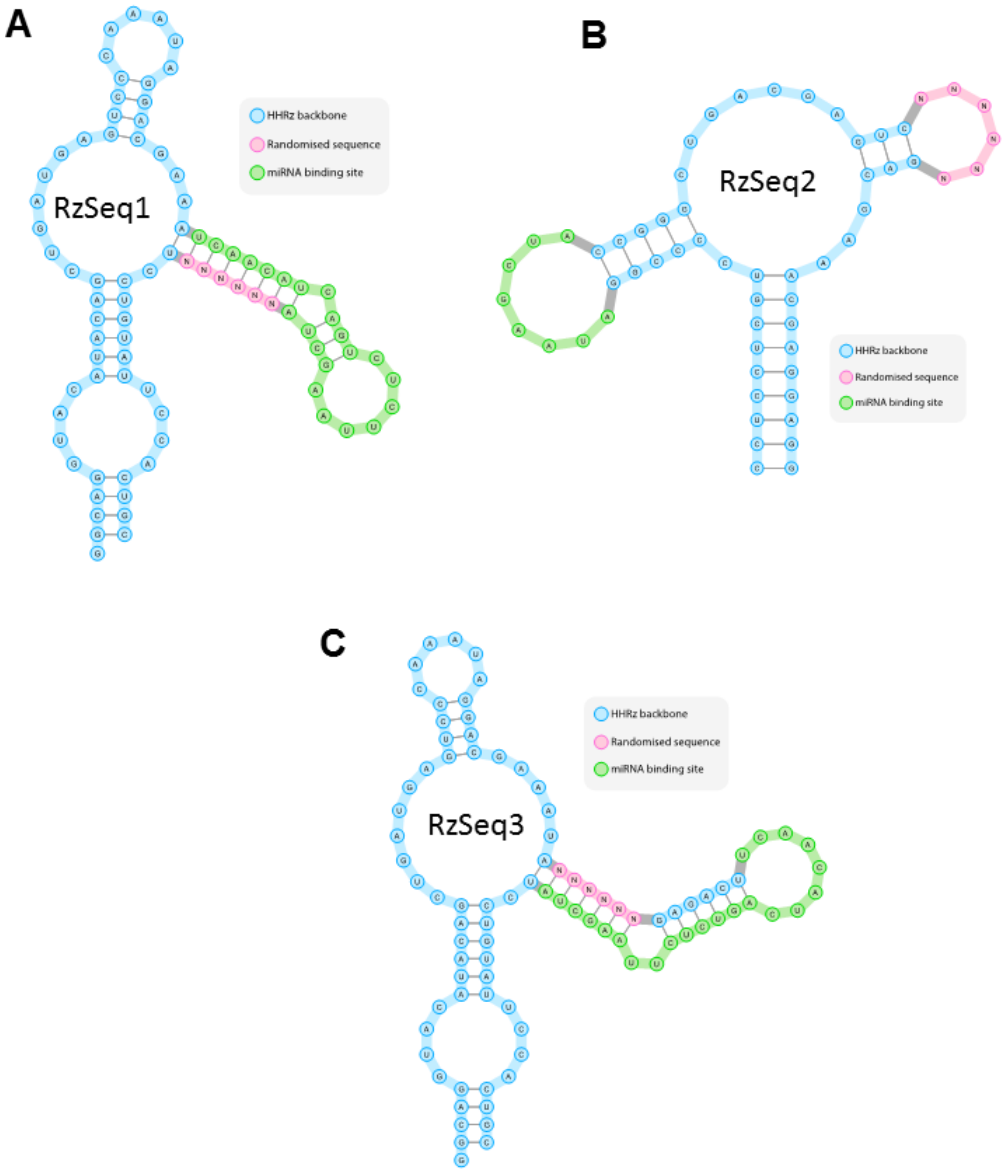



Figure 5.1: Predicted secondary structures of indicative representatives of the RzSeq1, RzSeq2 and RzSeq3 ribozyme pools showing miR-21 binding sites and randomised regions
*(A) The RzSeq1 pool is based on a type I hammerhead ribozyme. A miR-21 binding site has been incorporated into the 5' arm of stem III. Pool diversity was generated through replacement with six contiguous random nucleotides at the base of the 3' arm of stem III. Allostery would be predicted to occur via binding of miR-21 to stem III and disruption thereof. (B) The RzSeq2 pool is based on a type III hammerhead ribozyme. A truncated miR-21 binding site, corresponding to the seed sequence only, has been inserted into the loop at the end of stem I. The loop at the end of stem II has been randomised with six contiguous Ns. miR-21 binding would be expected to disrupt the interactions between the loops of stems I and II, thus allowing for allosteric modulation. Continued...*

The third ribozyme pool, RzSeq2, was designed to undergo allosteric modulation via a distinct mechanism. Self-cleavage of the hammerhead ribozyme in physiological magnesium ion concentrations is known to depend on tertiary interactions between loops on stems I and II [69, 70]. The sequences of these loops are variable and amenable to modifications and so long as these tertiary interactions are maintained and/or reconfigured appropriately ribozyme functionallity is likely to be retained. This principle has been exploited for the generation of allosteric ribozymes by replacing one loop with an aptamer sequence and introducing compensatory changes into the opposite loop to maintain tertiary interactions [151]. Upon ligand binding, the tertiary interaction is disrupted and ribozyme self-cleavage is inhibited. This mechanistic shift from the disruption of secondary structure to that of tertiary structure confers faster switching kinetics to these types of allosteric ribozymes [151, 152]. The RzSeq2 pool was designed based on these principles. A miR-21 target site was incorporated into the loop of stem I, and the putative loop of stem II was replaced by six random nucleotides (Fig. 5.1B). Given the typical size of stem I loop and biochemical data demonstrating similar binding affinities of various miRNAs to either full length or seed only target sites [95], the miR-21 target site on RzSeq2 was condensed to contain only the seed sequence (complementary to miRNA positions 2 - 7).

### 5.2.2 Population level analysis of ribozyme and/or miRNA related effects on dEGFP-coding constructs bearing individual ribozyme pools

Before attempting any selection it was necessary to assess the behaviour of the individual random ribozyme pools and their response to the cognate miR-21 at the population level. It was particularly important to investigate the susceptibility of ribozyme pools to miRNA-mediated repression since this can serve as a major confounding factor in the identification of miRNA-responsive allosteric ribozymes. Despite this caveat, any demonstration of miRNA-mediated repression would still serve as evidence in favour of an interaction between miR-21 and the ribozyme-borne target site.

We created a new plasmid vector for use in the following RNAseq selection experiments. This vector contained a bidirectional CMV promoter ($P_{BD\text{-}CMV}$) that simultaneously drove the transcription in opposite directions of both the mCherry and dEGFP genes (pBiDir-mCherry-dEGFP, Fig. 5.2A). The 3' UTR of the dEGFP gene had been engineered to contain conve-

*Figure 5.1 continued: (C) The RzSeq3 pool is based on a type I hammerhead ribozyme. A miR-21 binding site has been inserted into the 3' arm of stem III and six random bases have been inserted at the base of the 5' arm of stem III. Allosteric modulation would be predicted to occur as for RzSeq1.*

nient cloning sites for insertion of ribozyme sequences. The advantage of this plasmid is that mCherry expression can be used as a dEGFP normalisation control to account for both endogenous noise in gene expression and differences in transfection efficiency. Co-transfection of the empty bidirectional vector (pBiDir-mCherry-dEGFP) or the pooled vectors harbouring either RzSeq1 or RzSeq2-originating ribozymes in the 3' UTR of the dEGFP gene (pBiDir-mCherry-dEGFP-RzSeq1 or pBiDir-mCherry-dEGFP-RzSeq2, respectively; note: these are not unique plasmids but instead each comprises a diverse randomised set derived from the cognate source ribozyme pool), together with the control (pEM-empty) or miR-21 (pEM-21) expression vectors, enabled investigation of both the population-averaged ribozyme pool activity and the overall responsiveness of each pool to the cognate miRNA. These experiments were not performed with the RzSeq3-containing plasmid pool which was expected to behave similarly to the RzSeq1-containing pool given the presence of an equivalently configured identical miR-21 binding sequence that was instead located on the opposite arm of stem III (Fig. 5.1, A and C).

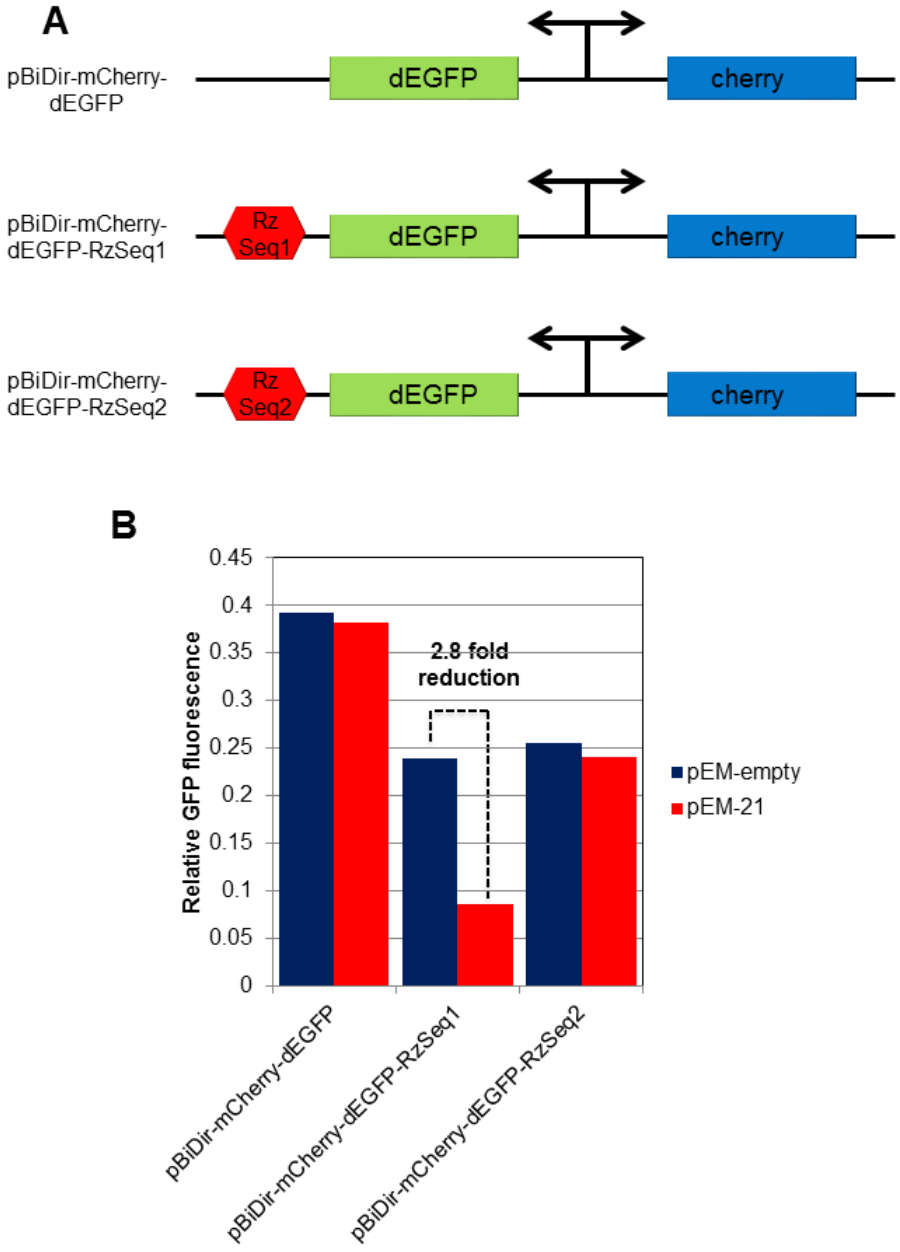



Figure 5.2: Population level analysis of individual ribozyme pools in response to the cognate miR-21

*(A) Schematic diagrams of the pBiDir based constructs with an empty UTR (pBiDir-mCherry-dEGFP), or a UTR carrying the RzSeq1 (pBiDir-mCherry-dEGFP-RzSeq1) or RzSeq2 (pBiDir-mCherry-dEGFP-RzSeq2) ribozyme pools; note: the latter two constructs are not unique plasmids but rather individual plasmid pools (populations of plasmids with each member distinguished only by the region of nucleotide randomisation) derived from the cognate ribozyme pool. (B) 5 × $10^5$ HEK293T cells were transfected with 500 ng of pBiDir-mCherry-dEGFP, pBiDir-mCherry-dEGFP-RzSeq1 or pBiDir-mCherry-dEGFP-RzSeq2 in combination with 500 ng of either a control (pEM-empty) or miR-21 (pEM-21) expression vector. Fourty-eight hours post-transfection, GFP and mCherry fluorescence signals were measured by flow cytometry to calculate the relative GFP fluorescence values (n = 1). Continued...*

When co-transfected with pEM-empty, both the RzSeq1 and RzSeq2-containing pools exhibited reduced GFP fluorescence by approximately 39 % and 35 % respectively, relative to the bidirectional empty vector in which the 3' UTR of the dEGFP gene is unstructured (Fig. 5.2B). This effect may partly be a result of ribozyme-mediated cleavage, although only a small fraction of each of the random pools would be expected to encode ribozymes capable of folding into catalytically active conformations. Alternatively, the presence of such highly structured regions within the 3' UTR of the dEGFP gene may have an inhibitory effect on translation. The distinction between these two mechanisms could have been investigated through the generation of equivalent catalytically-inactive (A14G substitution, [63, 73]) ribozyme libraries corresponding to both the RzSeq1 and RzSeq2 pools, and the subsequent generation of matched plasmid pools. However, since neither of the above mentioned mechanisms would preclude the discovery of allosteric ribozymes, further enquiry into this matter was considered to be a low priority.

Next, it was important to determine whether co-expression of the cognate miR-21 had any effect on the global expression profiles from these random ribozyme pools. All ribozymes within the RzSeq1 and RzSeq2 pools contained miR-21 binding sites designed using biochemical data to promote high affinity binding between this miRNA and its target [95]. In addition, mismatches were introduced into positions 8 and 9 of the target site to prevent Ago2-mediated cleavage. However, these target sites were still potentially susceptible to miRNA-mediated repression by either translational inhibition or mRNA destabilisation. As expected, since the bidirectional empty vector control does not contain any miR-21 target sites, the levels of GFP fluorescence corresponding to this experimental control were comparable in the presence of either a control miRNA or miR-21 (Fig. 5.2B). Likewise, based on the data obtained following transfection with its associated plasmid library (pBiDir-mCherry-dEGFP-RzSeq2), ribozymes within the RzSeq2 pool appeared relatively insensitive to miR-21 despite the presence of miR-21 seed only target sites (Fig. 5.2B). This result may reflect either the inability of miR-21 to bind the seed-only miRNA response elements contained within the RzSeq2 pool ribozymes, or ineffective miR-21 mediated repression arising from interaction at the RzSeq2-associated target site. In contrast, GFP fluorescence was inhibited by approximately 2.8-fold when cells harbouring the RzSeq1-containing pool were exposed to miR-21 relative to the control miRNA (Fig. 5.2B). Theoretically, the full-length miR-21 target site mismatched at positions 8 and

*Figure 5.2 continued: In the absence of miR-21, both the RzSeq1 and RzSeq2 encoding constructs exerted an inhibitory effect on GFP fluorescence relative to the matched empty plasmid. The addition of miR-21 elicited a strong inhibitory response as measured by GFP fluorescence from pBiDir-mCherry-dEGFP-RzSeq1 but resulted in little if any effect on the other two plasmids.*

9 contained within the RzSeq1 pool ribozymes cannot promote Ago2-catalysed cleavage. Therefore, the full extent of this repression can be most likely be attributed to translational inhibition and/or mRNA destabilisation.

Together, these results provided useful information that would aid in the interpretation of subsequent results. The population-level susceptibility of RzSeq1 pool ribozymes to miRNA-mediated repression may obscure detection of variants exhibiting allosteric modulation and thus this factor would need to be taken into account. Additionally, the potentially near global insensitivity of the RzSeq2 pool ribozymes to miR-21-mediated inhibition may reflect an absence of sufficient functional interaction, an observation which yet again provides important context for the interpretation of subsequent results.

### 5.2.3 Development of experimental platforms to select for ribozyme activity in the specific presence or absence of miR-21

RNAseq selection relies on comparative analysis of the abundance of individual ribozymes under conditions in which the cognate miRNA is either present or not. To maximise the chances of identifying miRNA-responsive allosteric ribozymes we decided to perform RNAseq selection under conditions of both miR-21 overexpression and endogenous expression. Since miR-21 expression cannot be detected endogenously in HEK293T cells [185], this cellular platform was used for the selection of ribozymes under miR-21 overexpression conditions. Conditions of low or high miR-21 expression could thus be created in HEK293T cells through the co-transfection of either a control (pEM-empty) or miR-21 (pEM-21) expression vector. In contrast, HeLa cells endogenously express miR-21 [185], thus making these a suitable platform for ribozyme selection under normal endogenous miR-21 expression levels. Equivalent, low or high (normal cellular level) miR-21 level conditions can be experimentally generated in HeLa cells by delivering a specific miR-21 inhibitor dissolved in DMSO or the matched vehicle control alone (DMSO), respectively.

Antisense competitive inhibitors (antagomiRs) have been used to great effect for the inhibition of miRNA activity in numerous experimental systems [186]. These molecules are oligonucleotides complementary to a miRNA which also contain modifications to the sugar phosphate backbone that promote high affinity binding and resistance to cellular nucleases. AntogomiR delivery in high doses leads to sequestration of a particular miRNA and derepression of its targets (Fig. 5.3A). For our experiments we chose a commercial miR-21 antagomiR (antagomiR-21) offered by Life Technologies. The optimal concentration of antagomiR required for complete inhibition must be determined empirically as it differs between cell lines and miRNAs. For this purpose, a miR-21 negative sensor comprising of a dEGFP

gene with two perfectly complementary miR-21 target sites within its 3' UTR was constructed (pcDNA-dEGFP-21T). The expression of dEGFP from this construct was predicted to display an inverse relationship to the amount of active miRNA and this sensor could thus be used to indirectly assess miR-21 activity. HeLa cells were transfected with the miR-21 negative sensor (pcDNA-dEGFP-21T), the piBlue-N1 transfection control and a series of increasing amounts of antagomiR-21 (Fig 5.3B). At the lowest concentration of antagomiR-21 (10 pM), GFP fluorescence was strongly repressed (Fig 5.3B). As the concentration of antagomiR-21 in the transfection mix was titrated up, marked derepression of GFP fluorescence was observed. AntagomiR-21-mediated inhibition of miR-21 and resultant derepression of GFP fluorescence appeared to saturate at antagomiR-21 concentrations above 20 pM. Based on these data, a final antagomiR-21 concentration of 30 pM was chosen for the subsequent RNAseq selection experiments.

### 5.2.4 RNAseq selection by way of depletion screening facilitated by quantification of uncleaved ribozymes

The first RNAseq selection strategy we pursued was depletion screening in which the abundance of uncleaved ribozymes was quantified by sequencing. Assuming that the transcription rates of all ribozyme variants within a pool are equal and that non-ribozyme associated transcript degradation is negligible, the steady state levels of uncleaved ribozymes should correlate inversely with the cleavage activity of each particular variant. Under this premise, the depletion screening approach would seek to identify ribozyme variants in which the abundance of the uncleaved forms is higher in the presence of miR-21 relative to the control non-specific miRNA. Any ribozymes that satisfy this criterion should constitute promising candidates for miR-21-responsive allosteric variants, which could then be used in the construction of a positive miRNA sensor circuit.

In order to differentiate between cleaved and uncleaved ribozymes, and specifically and selectively incorporate the latter into the sequencing library, primers spanning the cut site were used for the RT-PCR step (Fig. 5.4). This strategy had been employed in Chapter 4 during the ON state selection cycles. Incorporation of Illumina sequencing adapter sequences to the 5' and 3' ends of the ensuing amplicons was achieved during the RT-PCR amplification step by including the corresponding adapters within the overhanging tails of the forward and reverse PCR primers. This strategy allowed us to use RT-PCR to specifically capture uncleaved ribozymes in a conformation that was immediately amenable to high-throughput sequencing on the MiSeq Illumina platform.

The data output from the MiSeq platform was typically comprised of ~10 – 15 million

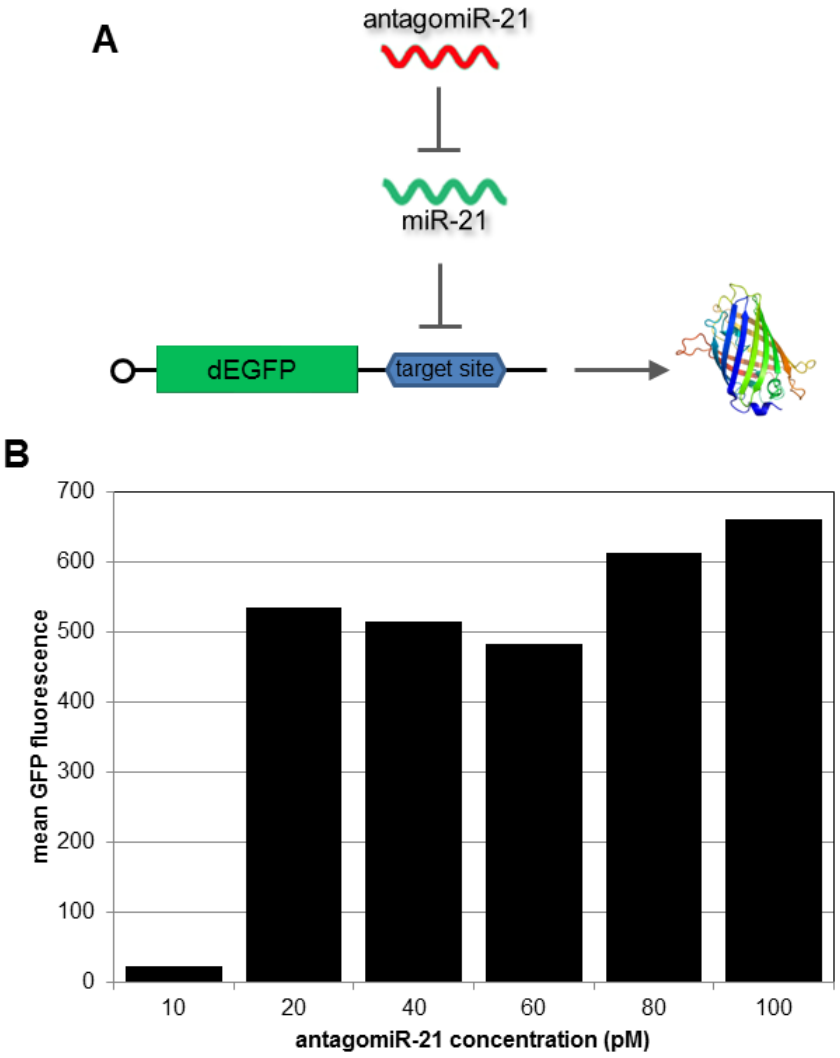



Figure 5.3: Titration to identify the minimum dose of antagomiR-21 required for inhibition of its cognate miRNA

*(A) The vector pcDNA-dEGFP-21T encodes the depicted schematic mRNA transcript encoding dEGFP and two 3' UTR copies of a perfectly complementary miR-21 binding site. dEGFP expression was inhibited upon interaction between miR-21 and its cognate target sites within the sensor transcript. It was postulated that this inhibition would be relieved via antagomiR-21 based sequestration of miR-21. (B) 5 × $10^5$ HEK293T cells were transfected with 500 ng of pcDNA-dEGFP-21T, 500 ng of piBlue-N1 (transfection control) and increasing quantities of antagomiR-21 at a final transfection mix concentration ranging from 10 - 100 pM. 48 hours post-transfection, GFP and iBlue fluorescence were measured by flow cytometry to calculate the relative GFP fluorescence values (n = 1). GFP fluorescence was derepressed by when ≥ 20 pM antagomiR-21 was present.*

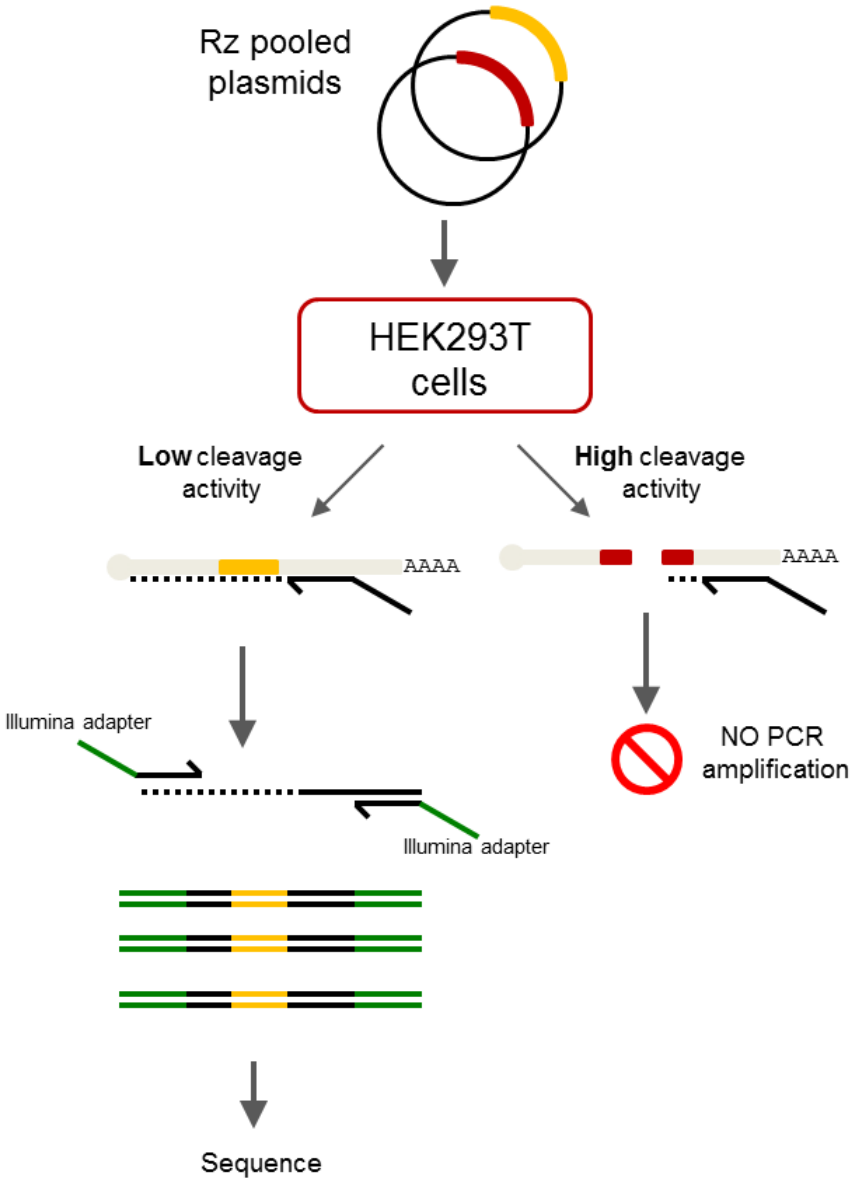



Figure 5.4: Schematic of the RNAseq depletion screening strategy utilised in this study *Following transfection of ribozyme plasmid pools into HEK293T cells, uncleaved ribozyme-containing transcripts can be captured by RT-PCR with cleavage site spanning primers and incorporated into a sequencing library. The PCR primers contain 5' extensions comprised of adapter sequences (Illumina P5 and P7 sequences, depicted in yellow) to enable the direct high throughput sequencing of amplicons.*

reads, each of which could be divided into four distinct regions: index, 5' invariant ribozyme sequence, random hexamer region, and 3' invariant ribozyme sequence (Fig. 5.5). These reads were subjected to an analysis pipeline which is schematically depicted in Fig 5.5. The index specified from which experimental condition a read had originated. This information was initially used to demultiplex the data into the various originating experimental groups (cells exposed to the control miRNA versus miR-21, HEK293T versus HeLa cells), and to segregate reads from different biological replicates. Next, a trimming process was used to remove invariant sequences corresponding to 5' and 3' regions of the ribozyme, leaving only the 6 nt region which had been randomised within the starting pool. Given that the sequence of this 6 nt region accounts for the only variability between constituent members of the original ribozyme pool, it would be reasonable to conclude that it captures all, or at the least sufficient, information required to explain any differences in the rates of ribozyme self-cleavage. The number of occurrences of each unique sequence within these 6 nt trimmed reads were tabulated and subjected to downstream comparative analysis for the identification of potential miR-21-regulated allosteric ribozymes.

#### 5.2.4.1 Pilot RNAseq selection using RzSeq1 pool ribozymes

To understand the parameters governing this novel selection strategy, a pilot experiment was conducted on the RzSeq1 pool ribozymes using the depletion screening method described above. An important aspect of this pilot experiment was to determine the extent to which principally PCR stochasticity during library preparation, and other potential minor process-related variability (DNA synthesis for the original randomisation, non-homogenous ribozyme and matched plasmid pool mixes, transfection) contributed to technical noise which would manifest as discordant quantifications of identical ribozyme species between experimental technical replicates. This type of noise was likely to represent a particular weakness of such a depletion screening strategy as ribozymes with the highest cleaveage activity would correspond to those represented in the library at the lowest levels. Quantification of these active ribozymes, several of which may be potential allosteric ribozymes of interest, would be the most susceptible to technical noise, given that the stochastic nature of PCR frequently results in lower amplification efficiencies of rare species as compared to more abundant ones. To assess the level of noise, correlation coefficients between technical replicate samples were calculated following parallel RT-PCR amplification, amplicon library preparation, sequencing and data analysis steps. Four experimental conditions were explored in this pilot experiment; the RzSeq1-bearing plasmid pool was co-transfected into two different cell lines, either exhibiting (HeLa) or not exhibiting (HEK293T) endogenous expression of miR-21 (Fig. 5.6). miR-21 expression was established in HEK293T cells which do not endogenously expresss this miRNA by co-

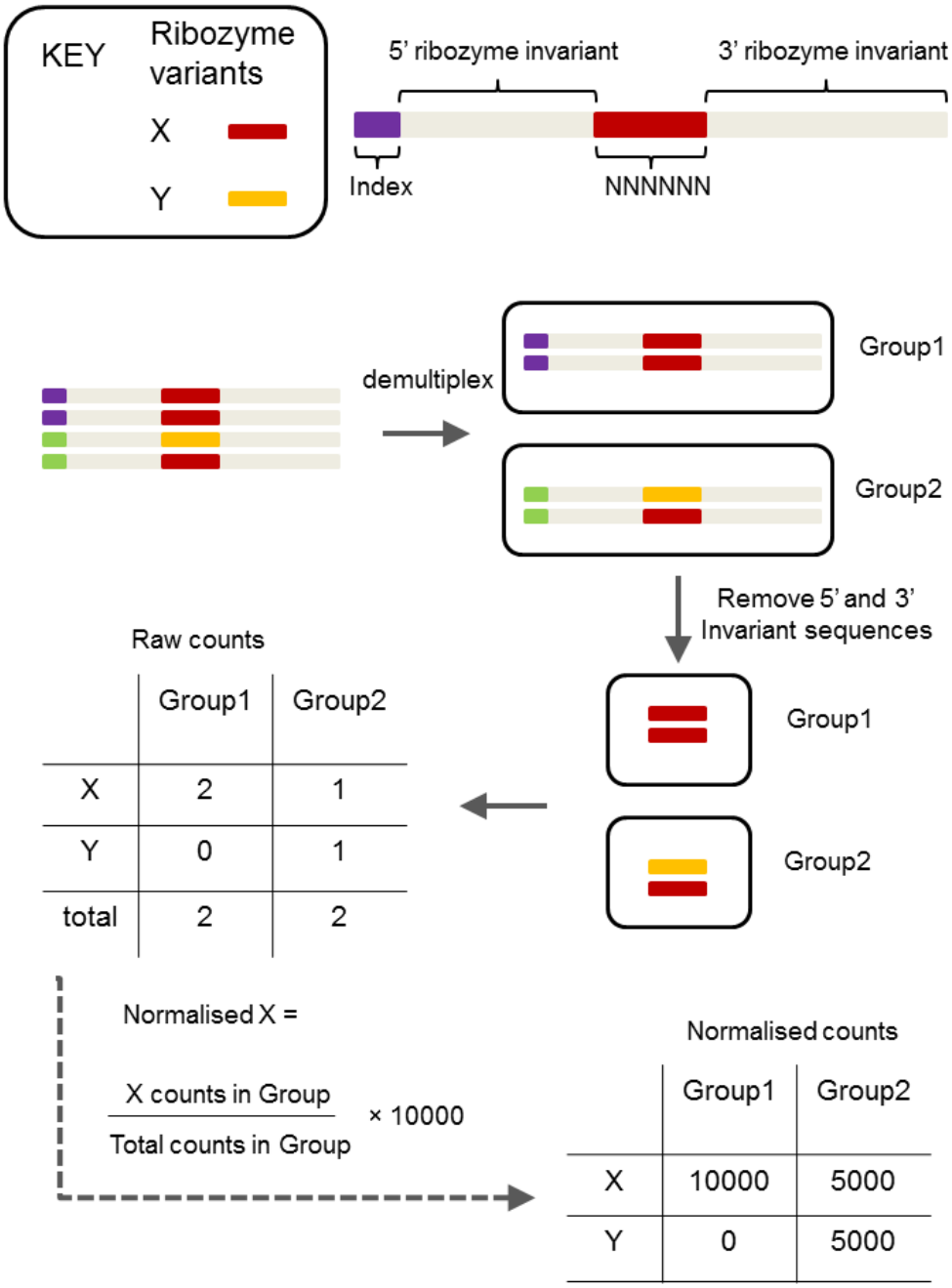



Figure 5.5: Schematic of the RNAseq data analysis pipeline implemented in this study *Ribozyme sequencing reads are first demultiplexed based on the identity of their index sequence. Next, sequences corresponding to the 5' and 3' invariant regions are removed. The frequency with which each ribozyme variant occurs is then tabulated. Finally, individual ribozyme counts within a group are normalised to the total number of reads encompassing all identified ribozyme variants within that group.*

transfection of a miR-21-encoding expression vector (pEM-21). Co-transfection of HEK293T cells with a control miRNA expression vector (pEM-empty) served as the negative control. Since HeLa cells endogenously express miR-21 at high levels, antagomiR-21 treatment was used to generate miR-21 negative conditions, while identical treatment with a DMSO vehicle control alone was used for miR-21 positive conditions. 48 h after transfection RNA was extracted and used to generate depletion screen amplicon libraries for sequencing.

The MiSeq data obtained was then subjected to the analysis pipeline described above (Fig. 5.5) and the read count corresponding to each unique ribozyme variant was compared between technical duplicates. RNAseq data obtained from technical replicates of HEK293T cells co-transfected with RzSeq1-bearing plasmid pool and either the control miRNA or miR-21-expressing plasmid were highly correlated with Pearson coefficients of 0.981 and 0.979 respectively (Fig. 5.7). The correlation was slightly reduced in HeLa cells, in which the Pearson coefficients between replicates were 0.942 and 0.859 in the presence of antagomiR-21 and the vehicle control, respectively (Fig. 5.7). Interestingly, a minor reduction in correlation was observed between technical replicates in both HEK293T and HeLa cells when miR-21 was present and active. Earlier observations had demonstrated a miR-21-dependent inhibition of GFP fluorescence in cells transfected with pBiDir-mCherry-dEGFP-RzSeq1 (Fig. 5.2), and we had proposed that this may have been a consequence of translational repression and/or mRNA destabilisation. A process of miR-21 mediated target mRNA destabilisation would lead to a global decrease in the abundance of RzSeq1-containing transcripts in the presence of miR-21. Such a mechanism may account for the increased technical noise observed between replicate samples under miR-21 positive conditions.

Inherent to the depletion screening strategy employed here, active ribozymes, a subset of which may constitute potential allosteric ribozymes, would be amongst the least represented species within the sequencing library. The pilot experiment described above revealed that ribozyme selection by RNAseq in HEK293T cells transfected with miR-21 was the experimental platform that displayed the least technical noise (highest correlation between technical replicate samples). Based on this consideration it was decided to continue pursuing RNAseq selection using only HEK293T cells to maximise the potential to identify statistically meaningful changes in ribozyme abundance levels. As a further means by which to minimise noise, an expression level threshold (normalised ribozyme count of 1) was employed such that only sequences present above this level were considered in downstream analysis.

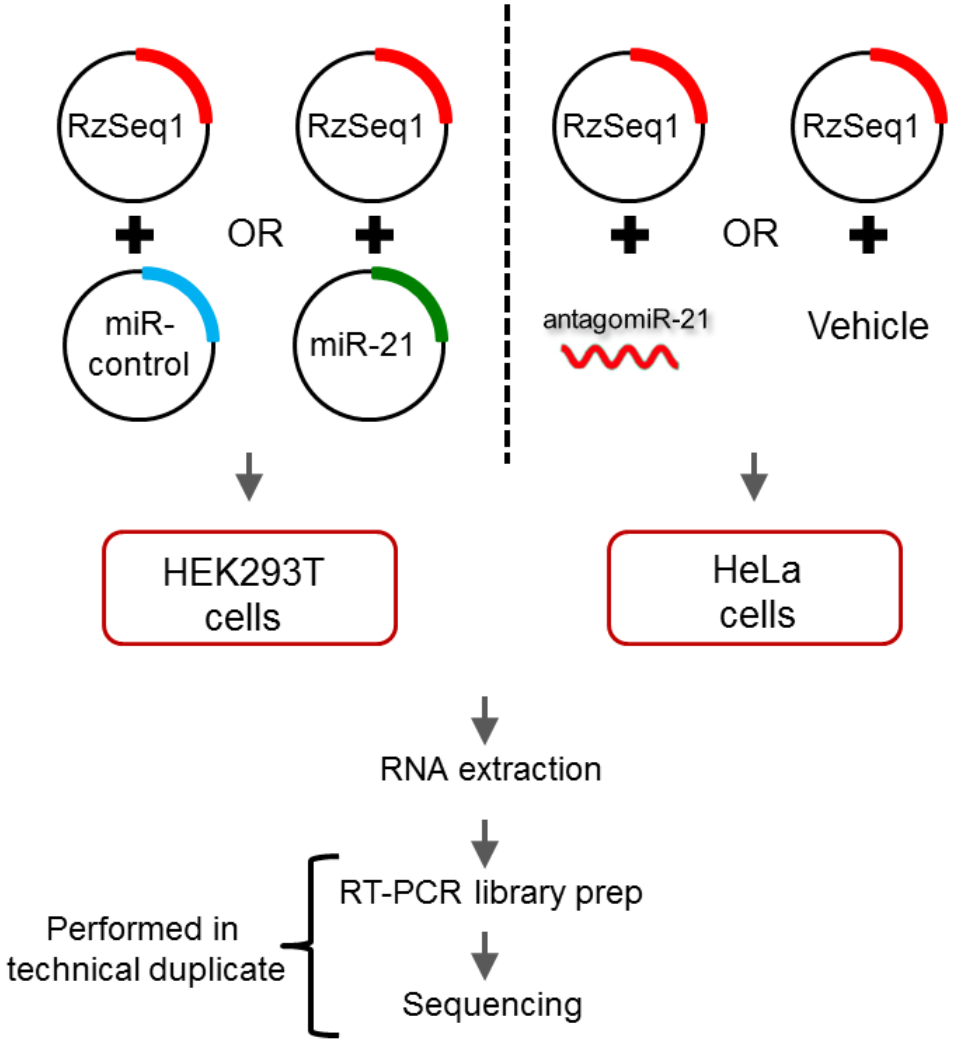



Figure 5.6: Overview of the pilot RNAseq depletion screening strategy pursued in this study *(A) 5 × $10^6$ HEK293T cells were transfected with 5 µg of pBiDir-mCherry-dEGFP-RzSeq1 in combination with 5 µg of either a control (pEM-empty) or miR-21 (pEM-21) expression vector (n = 1). (B) 5 × $10^6$ HELA cells were transfected with 5 µg of pBiDir-mCherry-dEGFP-RzSeq1 in combination with antagomiR-21 dissolved in DMSO (final concentration 30 pM) or the vehicle control only (DMSO) (n = 1). (C) RNA was extracted from cells, incorporated into sequencing libraries by RT-PCR, and sequenced on the Illumina MiSeq platform (technical replicates, n = 2).*

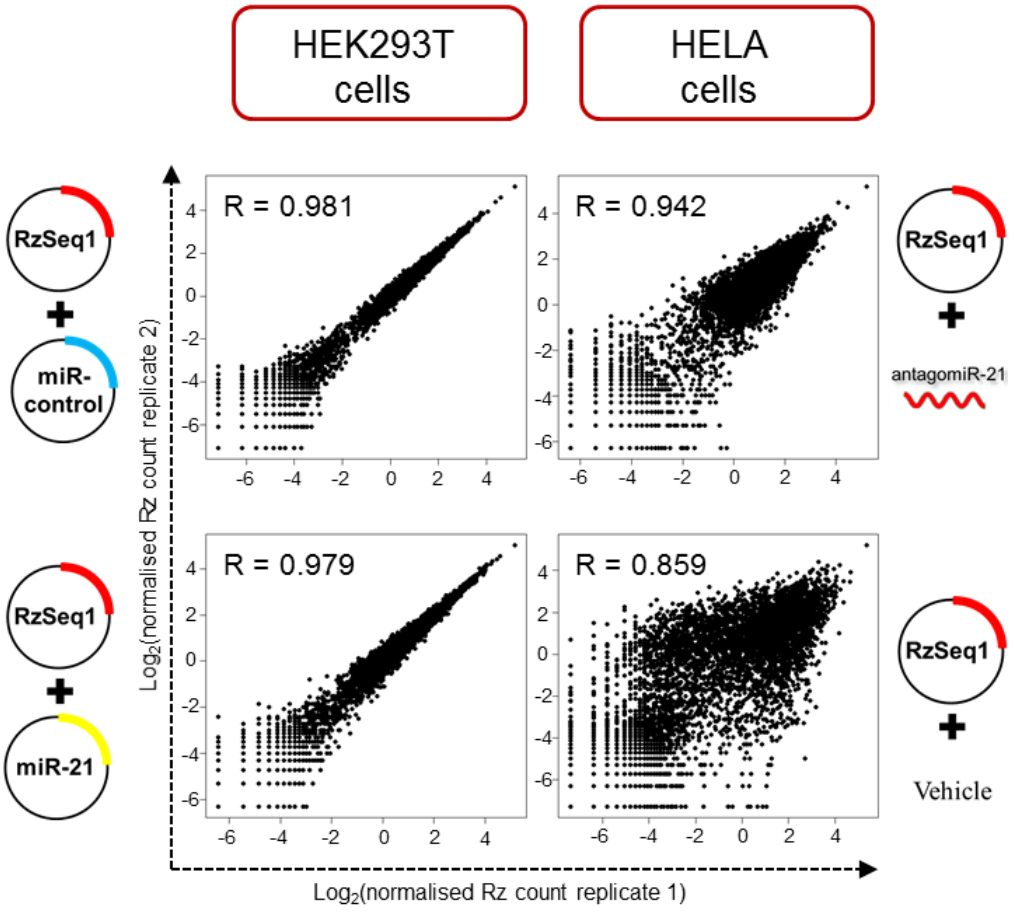



Figure 5.7: Correlation of data between technical replicates of the pilot RNAseq selection experiment

*Scatter plots charting the normalised read counts of each unique RzSeq1 pool ribozyme between the two technical replicate samples are shown. Comparison of data from technical replicates enabled estimation of the degree of technical noise. Replicate samples obtained from HEK293T cells demonstrated higher correlation coefficients than equivalent samples from HELA cells. Correlation between replicate samples was reduced for both HEK293T and HeLa cells when miR-21 was present and active. In HEK293T cells miR-21 was supplied transgenically (pEM-21), while in HeLa cells endogenously expressed miR-21 was left uninhibited by transfection with vehicle only (DMSO) in the absence of antagomiR-21.*

#### 5.2.4.2 Characterisation of candidate allosteric variants within the RzSeq1 and RzSeq2 ribozyme pools revealed the likely occurance of both miR-21 mediated inhibition of ribozyme cleavage and repression of output expression from the associated ribozyme-bearing transcript

RNAseq selection from the RzSeq1 ribozyme pool in the absence or presence of miR-21 was carried out in biological triplicates in HEK293T cells using the methods discussed above. Although equivalent data was generated for the RzSeq2 ribozyme pool, in this case only a single biological sample was examined. Following data processing according to our standard pipeline (Fig. 5.5), the abundance levels of individual ribozymes were tabulated, normalised, averaged across replicate samples, filtered on a threshold normalised expression level (1), and finally compared between the two conditions as defined by the absence (miR-control, empty vector control (pEM-empty)) or presence of the cognate miR-21 (Fig. 5.8).

The most striking immediate observation from this analysis is the stark contrast between the number of variant ribozymes exhibiting differential abundance levels of uncleaved forms in the RzSeq1 ribozyme pool in the presence of the control miRNA as compared to miR-21 itself, relative to the number of such ribozymes observed within the RzSeq2 ribozyme pool (Fig. 5.8). While 100 unique ribozyme variants from the RzSeq1 pool were found to be expressed at levels $\geq$ 2-fold higher in the presence of miR-21 relative to the control miRNA (Fig. 5.8, blue points), only a single sequence variant within the RzSeq2 pool demonstrated this behaviour. The Pearson correlation coefficient between abundance levels amongst RzSeq2 variants under conditions with or without miR-21 was 0.973. Despite unavoidable biological noise impacting differentially on individual experimental samples, this is approaching the correlation that was observed between technical replicates (Fig. 5.7), and thus would imply minimal biological variation in the abundance of RzSeq2 pool ribozymes. This is in agreement with earlier observations showing that GFP fluorescence from cells transfected with the RzSeq2-derived plasmid pool was not inhibited in the presence of miR-21 as compared to miR-control (Fig. 5.2). Together, these results could be interpreted as a failure of miR-21 to bind to its target site contained within RzSeq2 ribozymes and therefore an inability to elicit any allosteric modulation of cleavage activity. Alternatively, though rather less likely, it is possible that the short introduced randomised sequence within the RzSeq2 pool ribozymes led to an impairment of cleavage activity that was independent of miR-21 binding amongst the vast majority of RzSeq2 variants within this pool.

Considering the direct link between a ribozyme sequence, structure, and catalytic activity, an attempt was made to identify sequence parameters that may be predictive of ribozyme catalytic and/or allosteric activity using the data generated from the RzSeq1 pool RNAseq experiment. Ribozyme sequences could be classified according to their cleavage activity (ex-

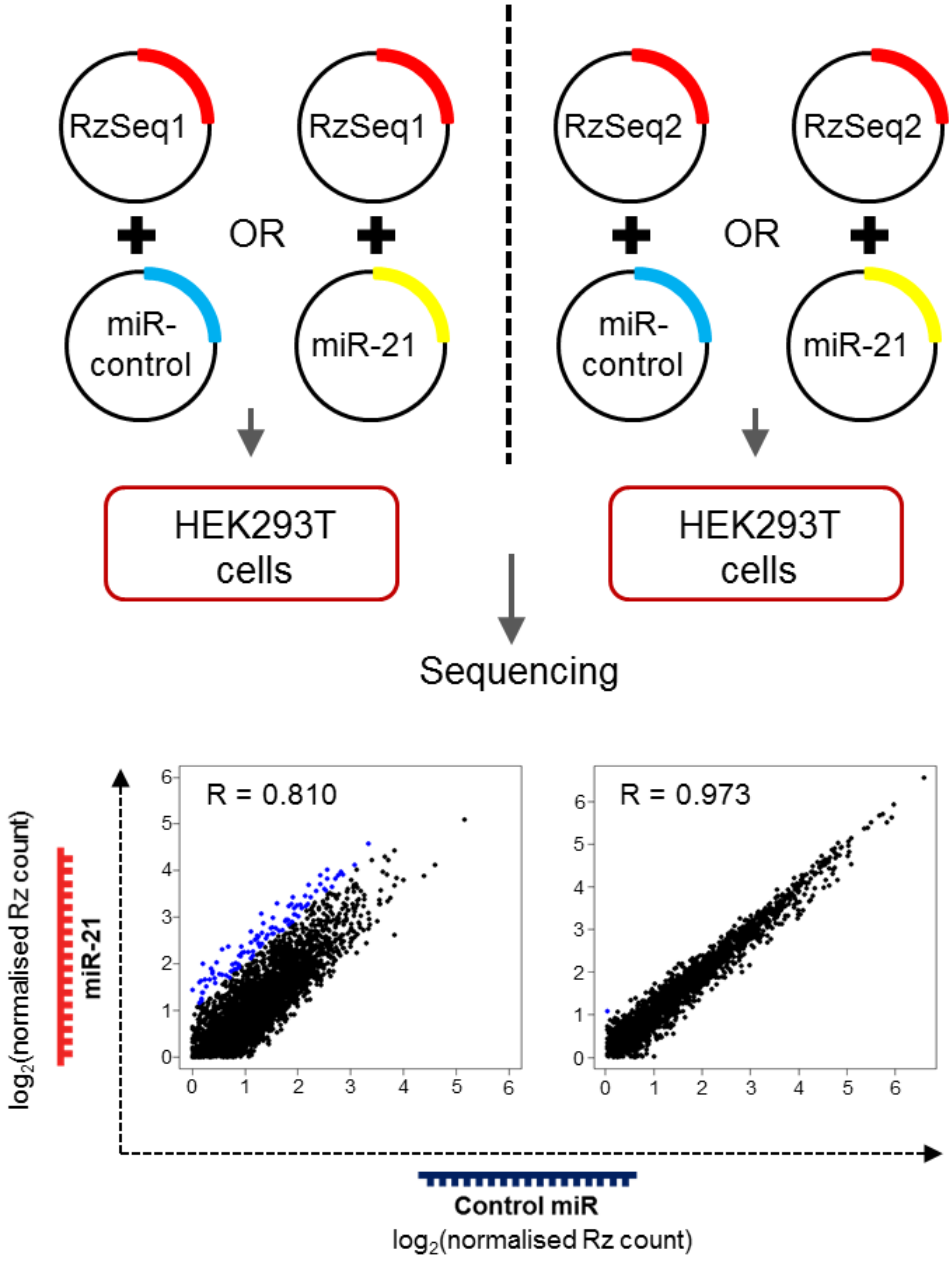



Figure 5.8: Analysis of differential expression levels of individual ribozyme variants in the presence and absence of miR-21 reveals candidate allosteric variants
*$5 \times 10^6$ HEK293T cells were transfected with 5 µg of either pBiDir-mCherry-dEGFP-RzSeq1 or pBiDir-mCherry-dEGFP-RzSeq2 in combination with 5 µg of either a control (pEM-empty) or miR-21 (pEM-21) expression vector (biological replicates, n = 3). 48 hours post-transfection RNA was extracted, sequenced, and analysed according to the depletion screening scheme (Fig. 5.5). Differential ribozyme abundance in the absence or presence of miR-21 was visualised in the scatter plot (blue, $\geq$ 2-fold upregulation).*

pression level) or their ability to undergo allosteric modulation (differential expression in the absence or presence of miR-21). Twenty-five candidate ribozyme sequences each were chosen from the extremes at both ends of these two parameters and subsequently clustered by sequence similarity (Fig. 5.9). If sequence identity is the main determinant of function, it was hypothesised that ribozyme sequences within each group of 25 would cluster together more closely and exhibit a clear distinction from the remaining 25 sequences at the opposite end of the matching spectrum; the latter would themselves be expected to cluster broadly as a separate grouping.

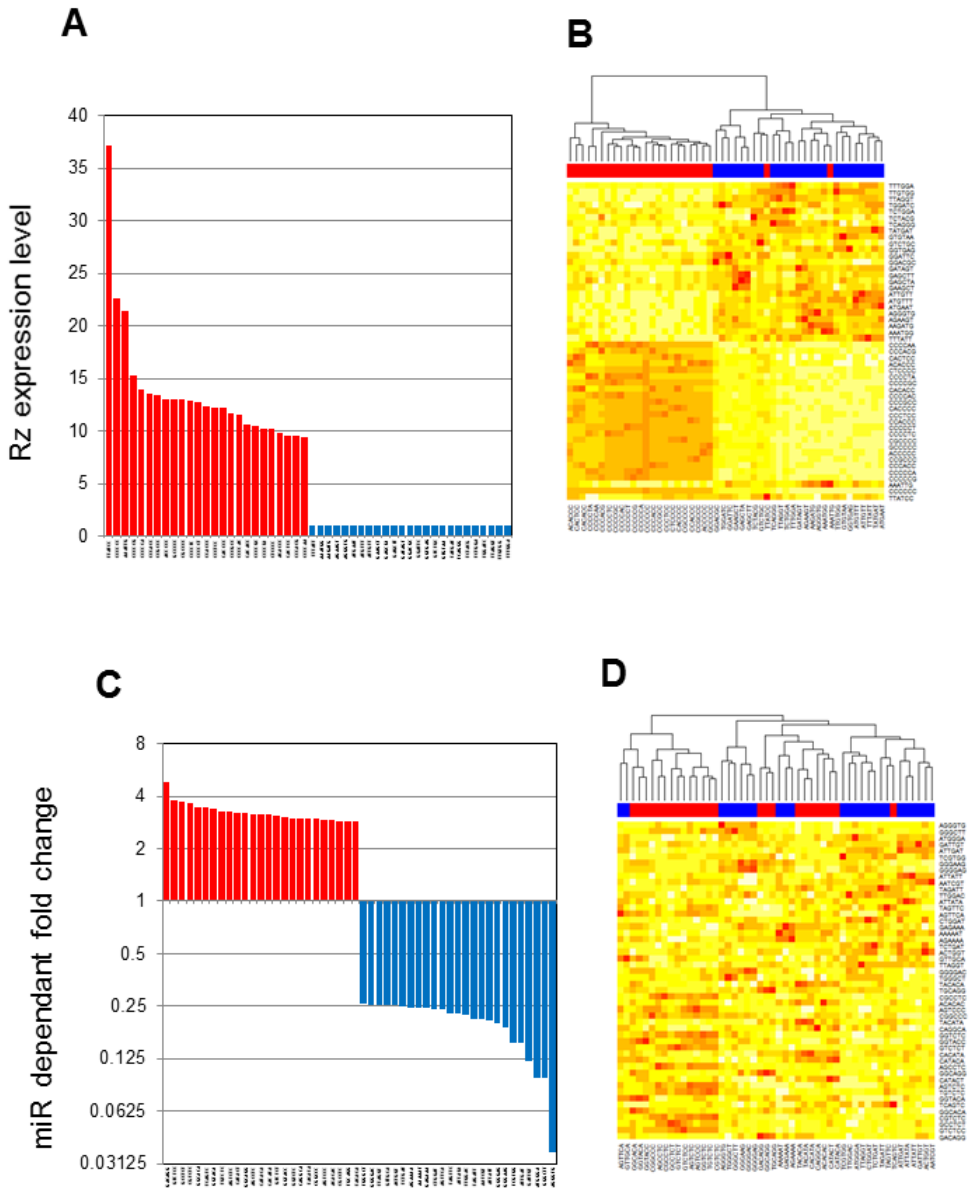



Figure 5.9: Use of Levenshtein edit distance clustering and heatmap visualisation to explore potential relationships between the sequence of the 6 nt randomised region of RzSeq1 and the observed ribozyme function

*(A) The normalised read counts of the 25 most (red) and 25 least (blue) abundant ribozymes found following transfection of the RzSeq1 pool constructs into HEK293T cells in the presence of the control miRNA expression vector (pEM-empty) are shown. These two groups were predicted to correspond to the least (red) and most (blue) catalytically active ribozymes, respectively. (B) Similarity between the 50 chosen sequences was calculated by the Levenshtein edit distance method, used to perform hierarchical clustering, and visualised by dendrogram and heatmap images. The two main branches of the dendrogram are well differentiated and congruent with classification into abundant and rare ribozyme species, thus supporting a correlation with regards to this trait between sequence and function. Continued...*

For this analysis, the expression levels (normalised counts) of the 25 most and 25 least abundant ribozymes following co-transfection of the RzSeq1-derived plasmid pool and an empty miRNA expression vector (pEM-empty) were considered. A large gulf in the levels of expression of the most abundant (red) and least abundant (blue) ribozymes was observed (Fig. 5.9A). Indeed, the top three most abundant ribozymes exhibited counts exceeding 20-fold that of the most prevalent least abundant ribozyme and even the lowest ranked most abundant ribozyme was present at about 9-fold the level of the top ranking least abundant ribozyme. The Levenshtein edit distance (number of insertions, deletions, or substitutions required to change one sequence into another) was then used as a measure of similarity between all 50 ribozyme sequences on which hierarchical clustering was performed. Based on this analysis, sequence similarity visualised as a heatmap revealed two distinct groups. These were also highlighted as the two major branches of the associated hierarchical clustering derived dendrogram. The final branches have been colour coded based on whether these correspond to the high or low expression group (Fig. 5.9B). This sequence based clustering approach successfully classified 96 % (48/50) of the ribozymes studied into the correct normalised count group. A similar analysis was performed on the 25 most upregulated and 25 most downregulated RzSeq1 pool ribozymes, in the presence of miR-21 (pEM-21) as compared to miR-control (pEM-empty) in HEK293T cells (Fig. 5.9, C and D). In this case, the heatmap and dendrogram obtained did not differentiate as clearly between ribozyme sequences that were allosterically inhibited or activated by miR-21. Assuming the primacy of the ribozyme sequence over the randomised region in the RzSeq1 pool in determining variations in allosteric activity, the credibility of the top 25 most upregulated hits as candidate positive miR-21 sensors would appear to be diminished. On the other hand it is possible that diverse distint permutations of prefered nucleotides at multiple, potentially varied, positions of the randomised region of an individual ribozyme, rather than convergence towards a single optimal sequence characterised by a WebLogo type signature, govern modulation of allosteric activity. If so, the Levenshtein edit distance analysis approach taken would have been unlikely to yield accurate clustering.

Finally, a subset representing the most promising ribozyme variants was selected for fur-

*Figure 5.9 continued: (C) The relative expression ratios of based on normalised read counts (miR-21-present: miR-21-absent) of the 25 most upregulated (red) and 25 most downregulated (blue) ribozymes found following transfection of the RzSeq1 pool constructs into HEK293T cells with pEM-empty or pEM-21 are shown. (D) The 50 selected sequences mapping to one or other end of the expression spectrum were clustered as outlined above, and the data visualised as a dendrogram and heatmap. Unlike the analysis with regards to catalytic activity, the analysed upregulated and downregulated ribozymes were interspersed to a considerable degree within the dendrogram obscuring any obvious segregation into the two original classification groups.*

ther functional analysis. A volcano plot was used to visualise ribozyme variants from within RzSeq1 pool, which underwent a statistically significant change in expression level in the presence of miR-21 (Fig. 10A, *p* value calculated by *t*-test between triplicate datasets). Ten ribozymes were selected from those demonstrating a statistically significant upregulation (Fig. 5.10A, blue dots), and individually characterised. The expression levels of each candidate, termed RzSeq1.1 to RzSeq1.10, were then considered. When measured by RNAseq, all 10 variants were found to be present as uncleaved entities at higher levels in the presence of miR-21 as opposed to a control miRNA (Fig. 5.10B). However, paradoxical results were obtained when the same ribozymes were cloned into the 3'UTR of the dEGFP gene and their effects on GFP fluorescence in the absence or presence of miR-21 were measured by flow cytometry (Fig. 5.10C). In this case, GFP fluorescence was reduced in the presence of miR-21 for all 10 candidate allosteric ribozymes tested. Since flow cytometry measures the levels of protein expression, these results do not directly contradict the RNAseq data which reflects mRNA transcript levels instead. Indeed, miRNA mediated allosteric modulation of ribozyme cleavage activity may result in an increased abundance of ribozyme containing transcripts in the presence of miR-21. At the same time however, the bound miRNA may exert repressive effects on the target transcript via a mechanism of translational repression. Thus, it is possible that a miRNA-ribozyme interaction both increases the abundance of output transcript mRNA whilst simultaneously reducing its competence to direct protein translation.

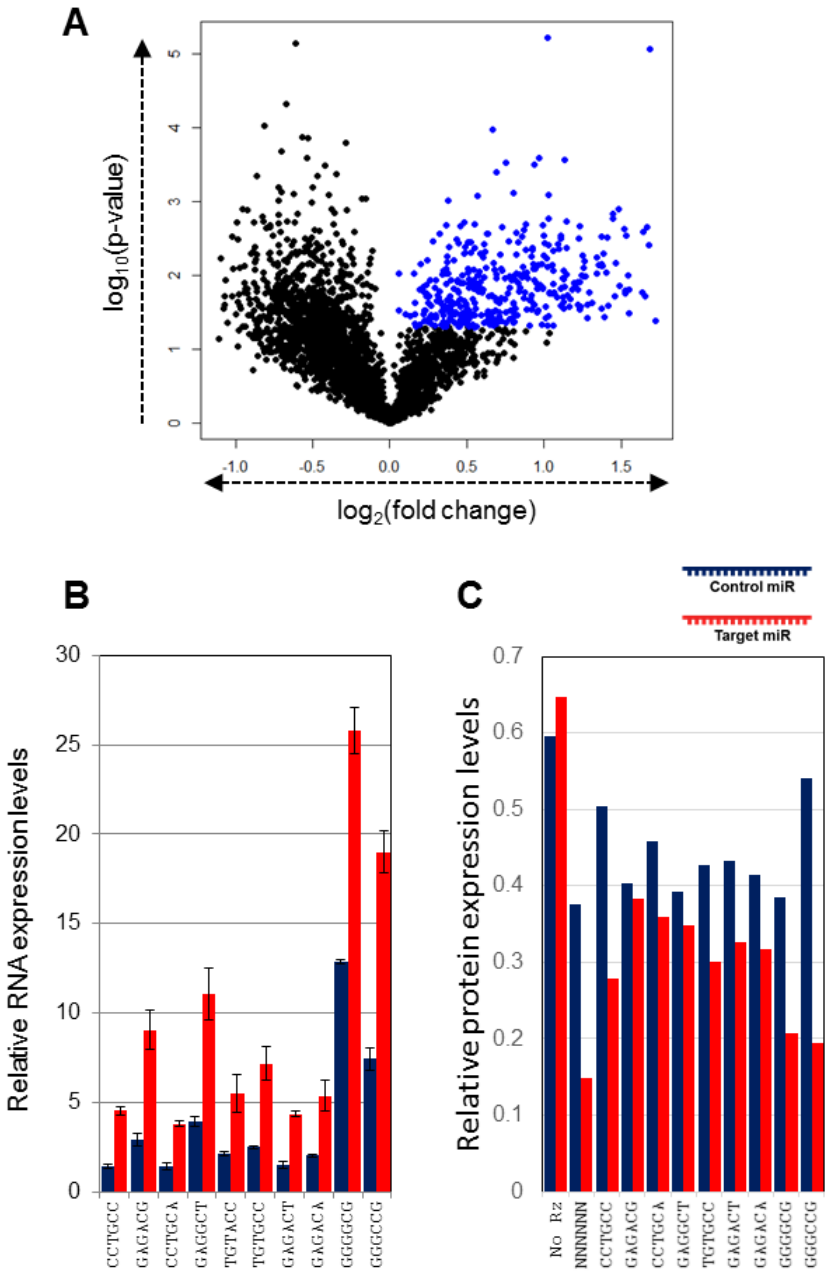



Figure 5.10: Characterisation of RzSeq1-sourced individual candidate allosteric ribozymes identified by depletion screening analysis
*(A) The differential expression of RzSeq1 pool ribozyme variants in the presence relative to absence of miR-21 was visualised using the volcano plot shown. p-values have been calculated between the two triplicate groups by the t-test. Significantly ($p < 0.05$) upregulated ribozymes are coloured blue. (B) Normalised read counts as measured by RNAseq selection of candidate RzSeq1 ribozymes (named according to the identity of the randomised region) in the presence of miR-21 (blue) and that of the control miRNA (red) are shown ($n = 3$, error bars represent standard deviation). Continued...*

### 5.2.5 RNAseq enrichment screen for miRNA-responsive allosteric ribozymes

The RNAseq depletion screening strategy described above that was based on amplification of full length ribozymes faced an interlinked problem - ribozymes with sufficient self-cleavage activity to be considered viable allosteric candidates were likely to be found at low abundance within the RNAseq library and consequently measures of their abundance were prone to elevated technical noise. In an attempt to address this issue we have developed a modified RNAseq selection protocol by which active ribozyme variants are enriched in the RNAseq library and a digital sequencing method is employed to minimise technical noise associated with the PCR amplification step.

As demonstrated by the RT-PCR selection approach, structured stem loop primers can enable specific and selective reverse transcription of the 5' by-product of ribozyme self-cleavage (see Chapter 4). In that case, five iterative cycles of selection using structured stem loop primers were sufficient to markedly/almost exclusively enrich for active ribozyme variants from within a largely inactive random ribozyme pool, as measured by the ability of active ribozymes to inhibit GFP fluorescence as a consequence of cleavage of the associated output transcript (Chapter 4, Fig. 4.10). For this part of the study, active ribozymes from the RzSeq3 pool were selectively incorporated into an RNAseq library using a structured stem loop primer-based reverse transcription step. RzSeq3 ribozymes were amenable to such selection because they contain known sequences immediately upstream of the cleavage site against which a structured primer was designed. The amount of cleavage product incorporated into this library was expected to be determined both by its rate of production, a function of the pace of generation of the ribozyme-bearing transcript itself and the ribozyme cleavage rate, and the rate of decay of its cleavage product. Assuming that the rates of transcription of the ribozyme-bearing mRNA and of cleavage product decay are relatively constant regardless of sequence variation within the randomised region, the quantity of a specific ribozyme variant captured by this approach can be used to infer its cleavage rate. However, the ribozyme cleavage by-product is expected to be very short lived in cells, due to its lack of a protective poly(A) tail and consequent rapid degradation by the exosome complex. Therefore, the absolute abundance of active ribozyme species available for capture into the sequencing library is likely to

*Figure 5.10 continued: (C) 5 × $10^5$ HEK293T cells were transfected with 500 ng of empty vector (No Rz, pBiDir-mCherry-dEGFP), RzSeq1-derived plasmid pool (NNNNNN, pBiDir-mCherry-dEGFP-RzSeq1), or individual RzSeq1 variants (the same as quantified in subfigure B) cloned into the 3' UTR of pBiDir-mCherry-dEGFP, in combination with 500 ng of either the control (pEM-empty) or miR-21 (pEM-21) expression vector. 48 hours post-transfection, GFP and mCherry fluorescence was measured by flow cytometry to calculate the relative GFP fluorescence values (n = 1).*

be substantially reduced compared to libraries prepared by amplification of full-length inactive ribozymes. This required the implementation of additional means by which to reduce noise.

In order to compensate for additional noise arising as a result of the short lived nature and therefore low abundance of the RNA species of interest, a digital sequencing methodology was incorporated into the enrichment screening protocol utilised [187–190]. Digital RNA sequencing is a recently described technique in which individual RNA transcripts are labelled during reverse transcription, each with a unique molecular index (UMI), before the subsequent library preparation PCR amplification step. Following sequencing, it is possible not only to map each read to a particular ribozyme variant but, using UMI tags, one can also determine from which single RNA molecule the read has originated. The number of distinct UMI tags associated with sequencing reads corresponding to a single ribozyme variant can be used to count the number of occurrences of that particular variant within the original sample. As exemplified in Fig. 5.11, both the more abundant RzA-containing transcript and the less abundant RzB-containing transcript are faithfully converted to cDNA by reverse transcription and tagged with UMIs. Following PCR amplification, RzA-associated cDNA molecules are amplified with perfect efficiency over two PCR cycles to give eight final amplicons and an equal number of sequencing reads. Information regarding the expression level of RzB is however distorted by the stochasticity of PCR. The rarer RzB cDNA variant fails to undergo replication during one of the PCR cycles and it is therefore amplified with less than 100 % efficiency, ultimately producing only two amplicons and a matching number of sequencing reads. Therefore, based on only the number of sequencing reads, the relative expression ratio of RzA:RzB would be incorrectly calculated as 4:1. In contrast, implementation of digital sequencing and the counting of unique UMI tags associated with sequencing reads from each ribozyme variant leads to an accurate measurement of 2:1 for the expression ratio of RzA:RzB. Consequently, on the basis of theory at least, technical noise originating from the PCR amplification step can be almost entirely eliminated by this method.

Based on the rationale discussed above we concluded that the combination of a structured stem loop RT-PCR enrichment screening strategy and a digital RNAseq readout step should further facilitate the discovery of miRNA-responsive allosteric ribozymes.

#### 5.2.5.1 Application of an experimental and analytical digital RNAseq protocol reduced technical noise associated with the read count output of a ribozyme enrichment screen

To carry out a digital RNAseq enrichment screen on RzSeq3 ribozymes, the encoding plasmid pool was co-transfected in biological triplicates into HEK293T cells with either the control miRNA or miR-21 expression plasmid, followed by RNA extraction, library preparation and

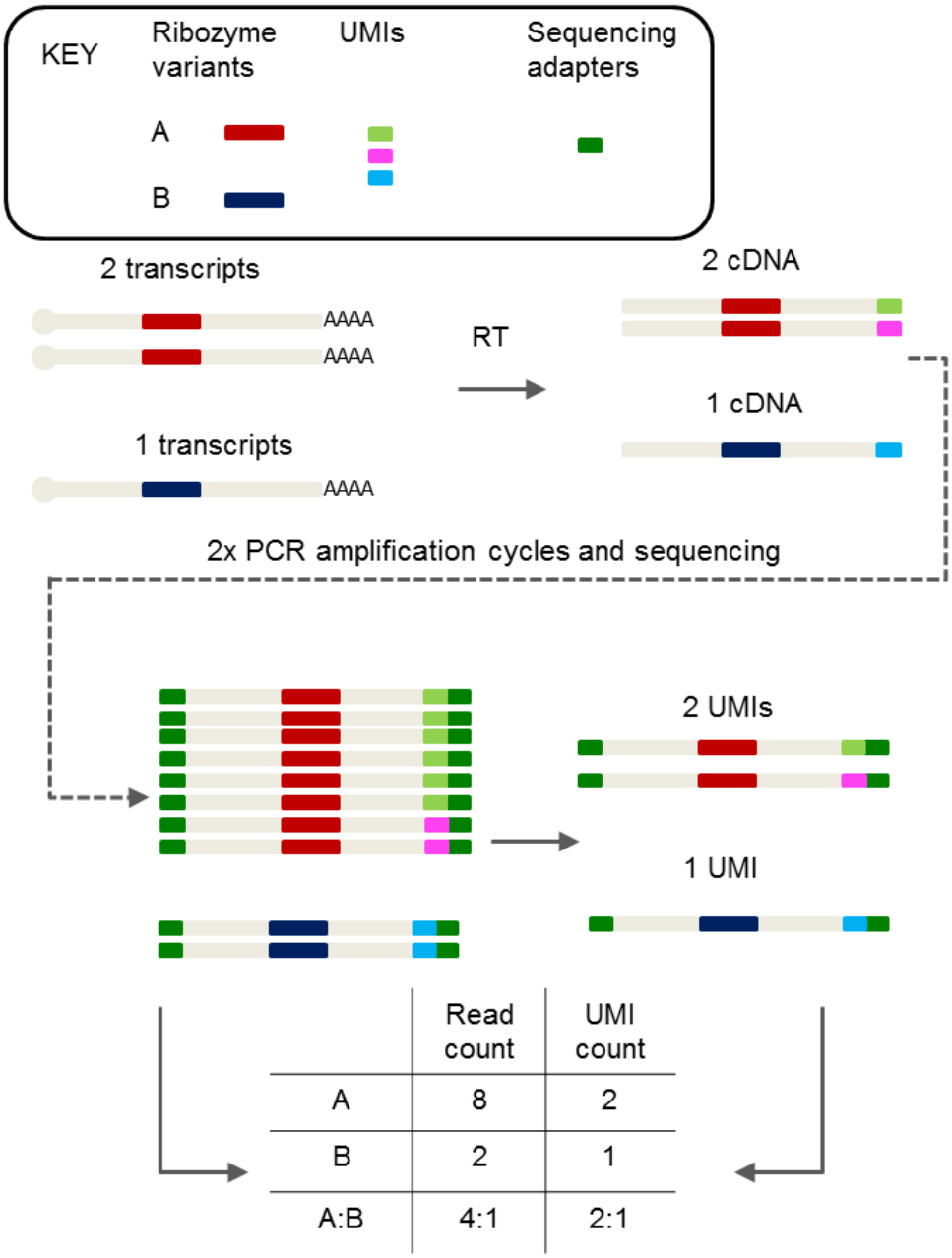



Figure 5.11: Schematic outline of the UMI based digital RNAseq analysis pipeline utilised in this study

*UMIs are appended to cDNA molecules derived from each individual ribozyme-bearing mRNA transcript. PCR is then used to incorporate sequencing adapter sequences and amplify the library for high throughput sequencing. Finally, bioinformatic analysis is performed by simple read counting or by counting the number of unique UMIs with which a single ribozyme species is tagged. In the event that stochastic factors during PCR lead to an amplification efficiency below 100 %, digital RNAseq analysis will to a potentially much greater degree of accuracy retain information regarding the relative expression ratios of transcripts.*

sequencing (Fig. 5.12). Data analysis for digital RNAseq followed a similar pipeline as described earlier (Fig. 5.5) with one important additional step. Following demultiplexing into individual experimental conditions, reads were collapsed such that each distinct UMI-tagged sequence only occurred once; each unique ribozyme sequence tagged with a different UMI was counted as a separate distinct sequence. Finally, the same data was analysed using both the conventional and digital RNAseq pipelines, thus enabling a direct assessment of the noise cancelling effect associated with incorporation of RNAseq.

The correlation between the abundance of each individual ribozyme variant as measured by RNAseq in biological replicate libraries could be used as a combined measure of both technical and biological noise. Assuming that the abundance is determined by ribozyme sequence only and that non-PCR stochasic elements have no influence, the number of filtered reads associated with each ribozyme variant should be identical across replicate experiments. Perfectly correlated data on a scatter plot would fall along the line X = Y, and would be expected to have a Pearson's correlation coefficient of 1 (Fig. 5.13A). Any deviation from this line would represent a discordance among measurements of ribozyme expression between replicate libraries, and would result in a Pearson's correlation coefficient < 1, and approaching 0 for completely uncorrelated data (Fig. 5.13B).

A qualitatively increased correlation was observed under both the control miRNA and miR-21-present conditions as on average the points on the scatter plot shifted towards the X = Y line when sequence read data were analysed using the digital RNAseq pipeline as compared to that which would be applied to conventional RNAseq derived data, (Fig. 5.13, C - F). Indeed, the improved correlations noted above were quantifiably validated by shifts in the Pearson's correlation coefficients from 0.725 to 0.819 and 0.684 to 0.783 under the control miRNA and miR-21-present conditions respectively, thus clearly demonstrating the utility of UMI barcoding and the digital analysis technique as a means of reducing experimental noise.

#### 5.2.5.2 Application of a digital RNAseq-supported enrichment screen of the RzSeq3 pool led to identification of 12 candidate allosteric ribozymes meeting the nominated cut-off thresholds

Given the above described rationale, it was hypothesised that candidate miR-21-responsive allosteric ribozymes could be identified by a comparative analysis of average expression levels in the presence of miR-21 versus a control miRNA (Fig. 5.14). Despite the noise reduction afforded by the digital sequencing technique, an expression threshold of 1 was still applied to identify candidate sequences whose levels were sufficiently tightly correlated between triplicate samples. The levels of expression of ribozyme variants in the RNAseq enrichment screen were expected to be positively correlated to ribozyme cleavage activity. Thus, candidate al-

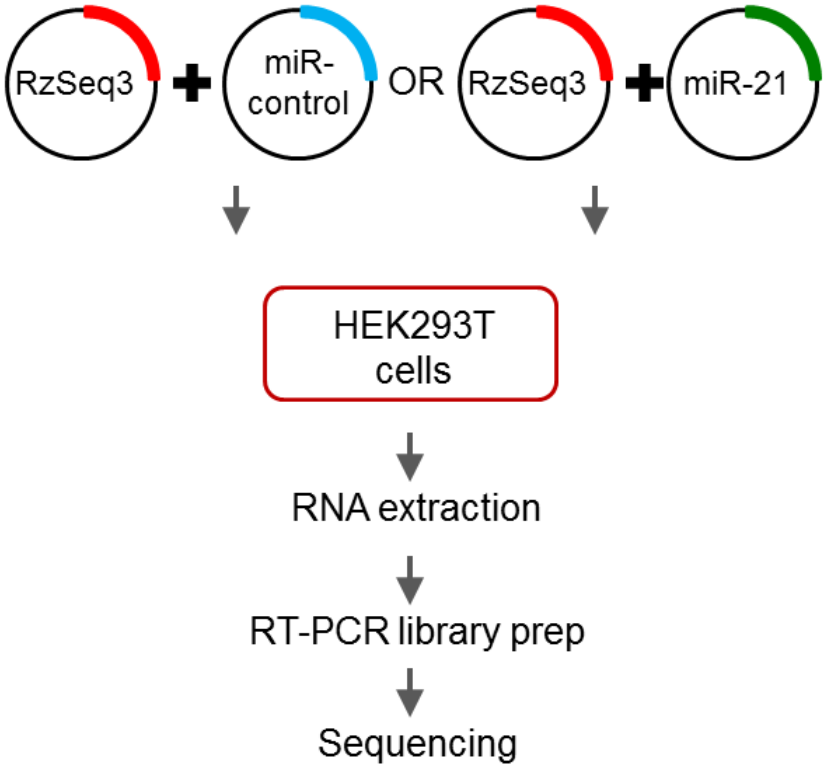



Figure 5.12: Overview of the RNAseq enrichment screening protocol to which theRzSeq3 ribozyme pool was subjected
*5 × $10^6$ HEK293T cells were transfected with 5 µg of pBiDir-mCherry-dEGFP-RzSeq3 in combination with 5 µg of either the control (pEM-empty) or miR-21 (pEM-21) expression vector (n = 3). 48 hours post-transfection RNA was extracted, incorporated into an RNAseq library, sequenced and analysed.*

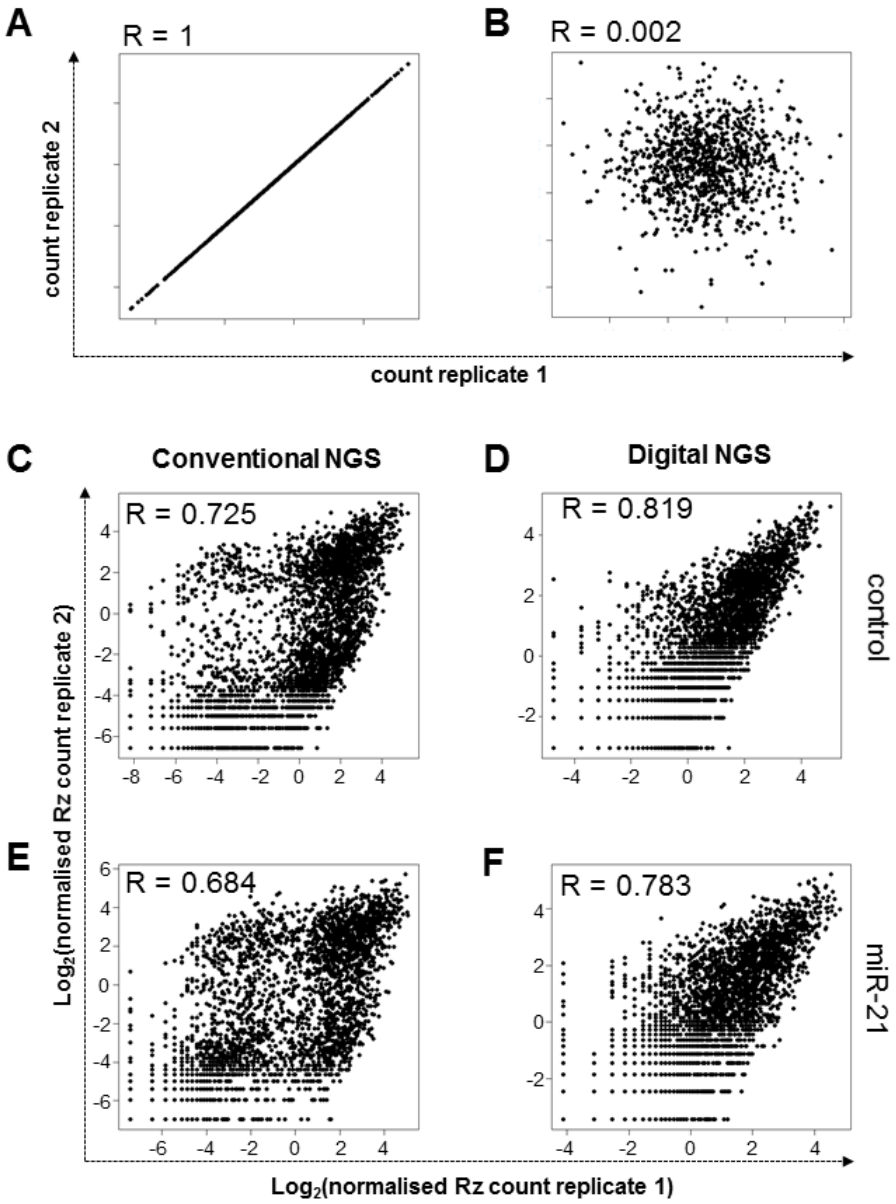



Figure 5.13: Digital RNAseq reduces technical noise and thus increases the correlation between data obtained using replicate samples

*(A) Hypothetical perfectly correlated data. (B) Hypothetical uncorrelated data. (C - D) Differential expression of individual RzSeq3 variants between biological replicate samples in the presence of the control miRNA expression vector (pEM-empty). The data obtained was separately analysed using both the conventional read count pipeline (C) and the digital RNAseq pipeline (D), clearly demonstrating that conventional next generation sequencing (NGS) led to more variability than analysis of the same data using the digital pipeline . (E - F) Similar analysis of differential expression levels between biological replicate samples for individual RzSeq3 variants in the presence of miR-21 (pEM-21). As above, a tighter correlation is observed between replicates analysed via a digital pathway (F) than a conventional one (E).*

losteric ribozymes could be identified by a reduction in their frequency of occurrence in the presence of the cognate miR-21 compared to its absence. Before threshold implementation, 435 candidate ribozymes were identified which underwent a $\geq$ 2-fold reduction in the presence of miR-21. Following threshold based filtering the number of candidates was reduced to 136 (Fig. 5.14).

A volcano plot was used to visualise differentially expressed ribozymes amongst the RzSeq3 pool and the corresponding levels of statistical significance of these calls (Fig. 5.15A). Fifty-one ribozyme variants appeared to undergo a significant inhibition of their catalytic activity in the presence of miR-21 ($p < 0.05$). However, only 12 of these ribozymes exhibited $\geq$ 2-fold downregulation (Fig. 5.15B). It is proposed that these 12 candidate ribozymes are to be subjected to in-depth characterisation during future studies.

## 5.3 Discussion

### 5.3.1 mRNA-mediated inhibition has the potential to contribute to global reduction in ribozyme expression across a randomised pool and consequent normalisation-related masking and/or indentification of false hits

At the population level, as reflected by repression of the dEGFP output emenating from the matched plasmid pool, the RzSeq1 ribozyme pool appeared to be susceptible to miR-21-mediated repression (Fig. 5.2B). Given that all ribozymes within this pool contained an identical miR-21 binding site it can be assumed that the levels of repression exerted on all ribozyme-bearing variants are similar. If we assume that the vast majority/most ribozyme variants underwent similar levels of miRNA mediated inhibition, then we would expect to see a reduction in expression across an equivalent proportion of RzSeq1 variants in the presence of miR-21. However, this effect would be masked through normalisation where the number of unique counts of a particular RzSeq1 variant (variant X) are divided by the total number of RzSeq1 counts all sequence variants. As the percentage of ribozyme variants affected by miR-21 mediated inhibition (mRNA destabilisation) approaches 100 % and assuming the above generalisations, both the numerator and denominator read counts would be reduced by similar fractions, "cancelling out" any variant-specific repression effect. Under this premise, it is plausible that ribozymes showing an apparent increase in abundance in the presence of miR-21 may merely represent those ribozymes with structural conformations that are less susceptible to miRNA binding and consequent repression. This hypothesis could in future be further explored by searching for a correlation between target site accessibility and allosteric

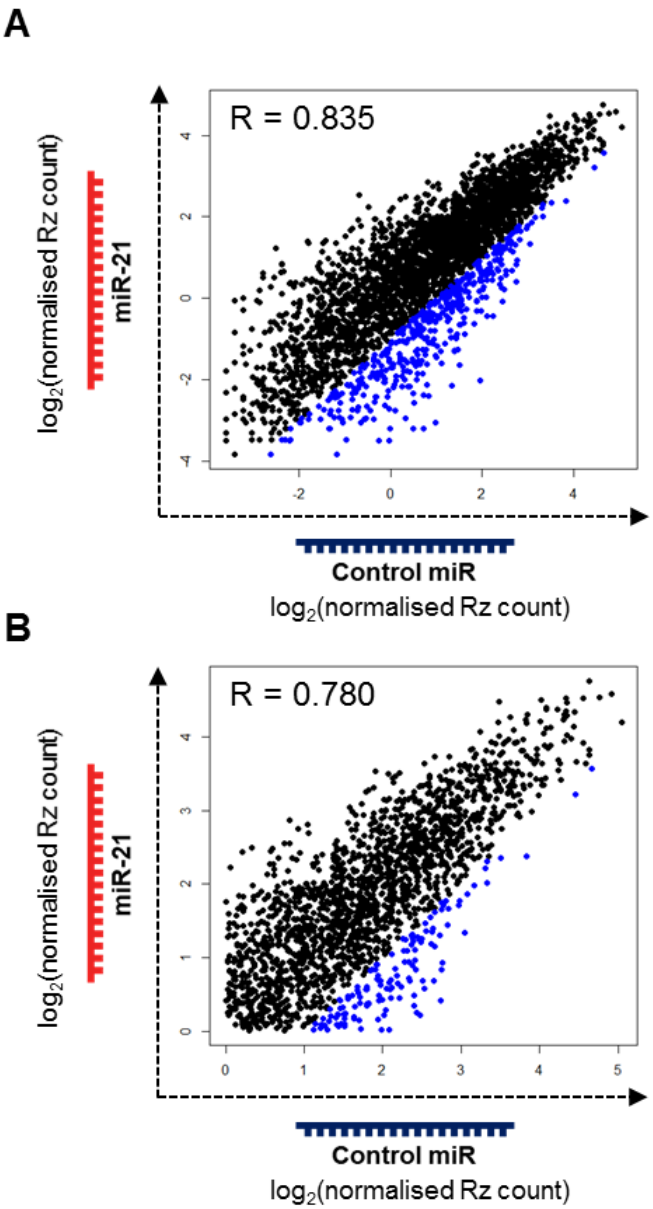



Figure 5.14: Differential expression analysis of RzSeq3 ribozyme variants in the presence and absence of miR-21 reveals candidate allosteric variants responsive to this miRNA
*Data was obtained as per the protocol shown in Fig. 5.12 and analysed via a digital RNAseq pipeline as per Fig. 5.11. Differential levels of ribozyme abundance in the absence or presence of miR-21 are visualised in the scatter plots (blue, $\leq$ 2-fold downregulation), before thresholding (A) and after threshold filtering (B).*

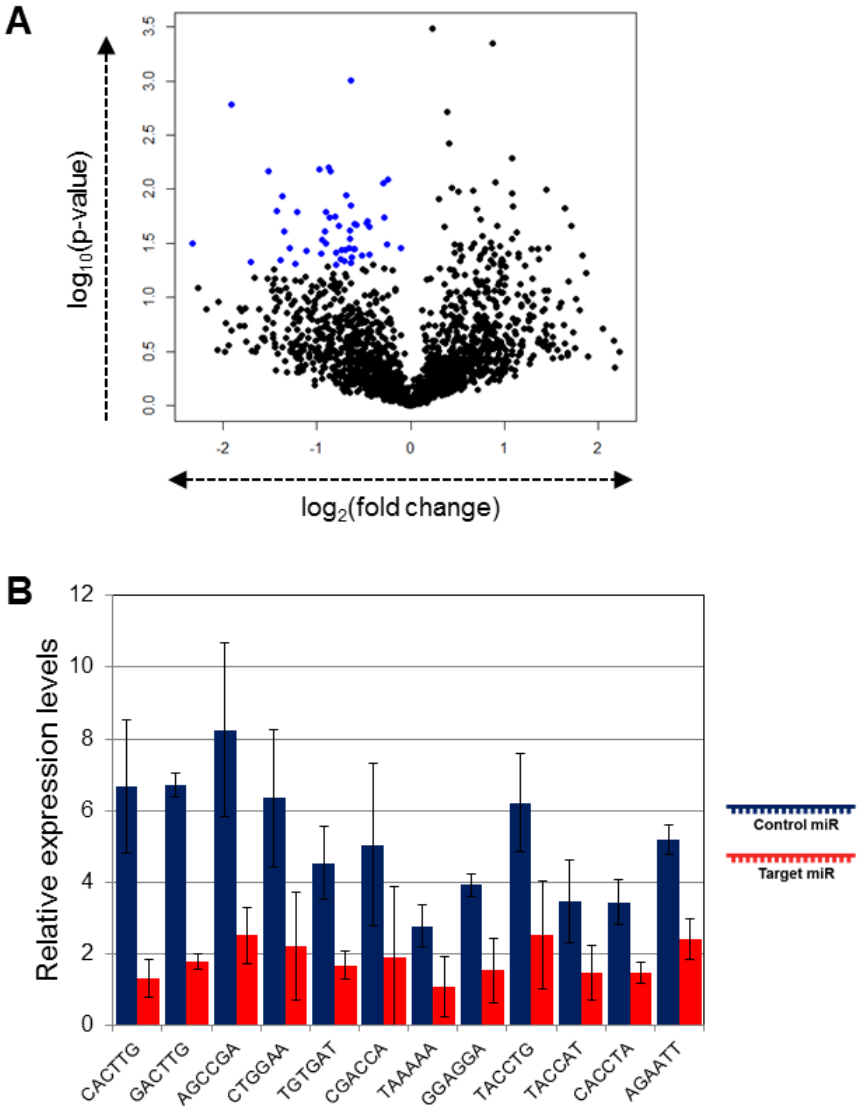



Figure 5.15: Characterisation of individual candidate allosteric RzSeq3 ribozymes identified by enrichment screening

*(A) The differential expression of RzSeq3 pool ribozymes in the presence relative to absence of miR-21 is visualised in this volcano plot. p-values have been calculated between the two triplicate groups by the t-test. Significantly ($p < 0.05$) downregulated ribozymes are coloured blue. (B) Normalised read counts as measured by RNAseq selection of candidate ribozymes, named by the sequence of their randomised region, in the presence of miR-21 (blue) and the control miRNA (red) are shown (n = 3, error bars represent standard deviation).*

ribozyme modulation using a miRNA target prediction algorithm, such as PITA, that takes into account both primary sequence and RNA secondary structure.

### 5.3.2 Functional allosteric ribozymes may be still hidden amongst the technical noise even when a noise reduction digital RNAseq protocol is utilised

The RNAseq selection procedure applied to the RzSeq1 and RzSeq2 ribozyme pools attributed increased catalytic activity to the most depleted variants. A high level of catalytic activity in the absence of its target miRNA is required for any potential ribozyme based positive sensor and hence the ribozyme sequences of interest would correspond to the most highly depleted species. This is however problematic as technical noise is most prominent with regard to analysis and interpretation of the least abundant sequences. This limitation precluded the use of data obtained from constructs introduced into HeLa cells, as these experiments consistently demonstrated the highest levels of noise. Furthermore, sequences potentially representative of the highest catalytic activity ribozymes may be lost after the data is filtered on a minimum expression threshold in an attempt to eliminate noise.

The parallel analysis of data arising from a single library generated for digital RNAseq processing by transcript counts versus read counts revealed the extent to which digital RNAseq can reduce technical noise as compared to standard high throughput sequencing (Fig. 5.13). However despite the clearly evident reduction in noise there still remained significant variability in transcript counts between biological triplicates, particularly amongst the least abundant transcripts. In digital RNAseq, library preparation requires a discrete reverse transcription step followed by PCR amplification. This is because individual RT primer molecules, each containing a unique molecular index (UMI), must be used for reverse transcription. Subsequently, these UMI-containing primers will need to be removed by purification before the PCR amplification step to prevent the spurious incorporation of UMIs into amplicons which would interfere with the accurate counting of input RNA molecules. The specific RT-PCR method applied in this study to prepare sequencing libraries for RzSeq1 and RzSeq2 pool ribozymes utilised a single reverse primer shared between both the RT and PCR steps, and was therefore not amenable to digital RNAseq adaptation. Consequently, digital RNAseq analysis was only performed on libraries generated using structured stem loop RT primers for the enrichment screen of the RzSeq3 ribozyme pool, as the preparation of these libraries involved a discrete RT reaction which was readily amenable to use of a UMI-containing RT primer.

Despite implementation of the digital RNAseq strategy outlined above, noise within RzSeq3 sequencing data was still sufficient to preclude detection of the least abundant ribozymes. In

digital RNAseq it is assumed that all PCR amplicons labelled by a common unique barcode are derived from a single cDNA molecule. This assumption would not be valid if RT primers were carried over and went on to participate in the PCR amplification step where these would spuriously label amplicons, thus adding and extra layer of noise. Notably, unlike the approach taken in this study, previously published methods for the preparation of digital RNAseq libraries often involved a comprehensive clean up between RT and PCR steps in order to ensure removal of RT primers. Such a methodological adaptation could be utilised in the future to attempt to further minimise noise.

### 5.3.3 Parallel DNA/RNA sequencing and potential future innovations supported by massively expanded sequencing offers powerful further opportunities for identification of allosteric ribozymes

The analysis of RNA sequencing data undertaken in this study was based on the assumption that read count was either negatively correlated with ribozyme cleavage activity for the RzSeq1 and RzSeq2 ribozyme pools, or positively correlated for the RzSeq3 pool. This however, is an oversimplification since the read count will also be affected by the proportion of each ribozyme sequence variant within the original DNA pool. Hence, we hypothesised that simultaneous collection and dual sequencing of DNA and RNA from a population of cells would enable the normalisation of counts for each ribozyme variant at the RNA level to corresponding values of each variant in the input plasmid DNA pool.

Such a technique was performed as part of the first sequencing run with samples from the RzSeq1 pool harvasted from both HEK293T cells and HeLa cells in the presence and absence of active miR-21 (data not shown). Given that the expected dynamic range within the DNA sample pools was expected to be lower than in the matched RNA pools, as the latter would frequently manifest much reduced lower limit counts as a result of ribozyme self-cleavage, fewer sequencing reads were devoted to DNA as opposed to RNA samples. Unfortunately, as a result of insufficient depth, this decision came at the cost of increased technical noise within the DNA read counts. Consequently, given the high level of DNA read count noise observed, it was decided that RNA counts would not be normalised to DNA levels. Given cost and logistical contraints, it was also decided that DNA sequencing would be excluded from subsequent runs as it was assessed that the benefits from DNA normalisation were likely to be outweighed by the loss of depth in RNA sequencing reads and the resulting increase in noise levels associated with this principal experimental output. Dual DNA/RNA sequencing would ideally be incorporated into future experiments in which a greater number of reads can be obtained from a sequencing platform with higher throughput than the Illumina MiSeq platform, such as for

example thethe Illumina NextSeq instrument capable of > 100-fold greater read capacity per run. Alternatively, a shorter region of randomisation within the ribozyme would reduce the number of unique species and thus increase the sequencing depth per unique ribozyme using the current methods. However, this would come at the substantial cost of much-to-massively reduced starting ribozyme diversity and the consequent reduced probability of identifying candidate allosteric ribozymes meeting the desired criteria. Indeed, given the power of *in vitro* evolution and the exponential gains in sequencing capacity that have been achieved, often at minimal added cost, it could be argued that future selection experiments should be initiated from a massively more diverse initial ribozyme pool carrying longer and/or multiple (rationally and/or empirically selected) discrete regions of randomisation. If so, such experiments would need to be supported by massively-scaled up sequencing to allow for capture of high read depth DNA and RNA sequence data, and potentially time course-related ribozyme count information as well. Consideration could also be given to identifying putative highly active ribozymes, not on the basis of relative quantification of cleavage products but rather on the basis of the rate of extinction of these species from the RNA time-course data or even the absence of an initial time-point RNA signal despite a cognate adequate strength DNA signal.

### 5.3.4 Use of a modified reverse transcription procedure minimises exposure to post RNA extraction *in vitro* conditions favouring ribozyme autocleavage

The RNA sequencing selection strategy applied in this study relied on the assumption that the ribozyme self-cleavage events that were being measured occurred in the cell and not post-extraction. However, besides occuring within the cell, RNA cleavage could potentially occur anywhere between RNA extraction and the cDNA-generating reverse transcription step. Furthermore, post-extraction cleavage is unlikely to be under the same constraints as would be the case in a cellular environment. miRNA RISC complexes are removed during the RNA purification process and therefore unable to inhibit ribozyme cleavage post RNA extraction. Additionally, small RNAs such as miRNAs are preferentially depleted during the RNA purification technique that was used in this study (Qiagen RNeasy mini kit) and consequently these entities would also have been unavailable post-extraction to modulate ribozyme cleavage activity.

The DNA digestion step following RNA extraction is particularly problematic. Although the RNA extraction process has specificity for RNA it is known to also result in co-purification of miniscule quantities of DNA. If this carried over DNA is not removed it has the potential to be amplified during the subsequent PCR step, since PCR would not differenti-

ate between the RNA-derived cDNA and any equivalent contaminating DNA. As ribozyme-encoding DNA does not exhibit the catalytic activity associated with the matching RNA species, contaminating DNA may mask some of the expected patterns of differential expression of the cleaved/uncleaved target ribozyme. Post RNA extraction DNA digestion is typically conducted in a buffer containing a high concentration of $Mg^{2+}$ ions at 37°C, conditions that are almost optimal for *in vitro* ribozyme self-cleavage. To overcome this issue, Kobori *et al*. conducted their DNA digestion step at 4°C in order to minimise ribozyme catalytic activity [184]. In contrast, for our study a more elegant solution was implemented to entirely avoid the DNA digestion step. Briefly, the reverse transcription primers utilised could prime cDNA synthesis through regions of 3' base pair complementarity but also contained extended 5' non-complementary regions. The novel sequence appended to the 5' end of the cDNA transcript could then be used as a binding site for a PCR primer, enabling direct differentiation between cDNA and the parental plasmid DNA from which it originated.

### 5.3.5 Apparent lack of target miRNA binding to the majority of RzSeq2 pool randomised ribozymes potentially accounts for the absence of a miR-21-related allosteric effect

Population level analysis of the RzSeq2 ribozyme pool activity facilitated by transfection of HEK293T cells with the pBiDir-mCherry-dEGFP-RzSeq2 plasmid pool revealed almost no difference in relative GFP fluorescence when transfected with an empty (pEM-empty) or miR-21 (pEM-21) expression vector (Fig. 5.2B). This result was in contrast to that corresponding to the RzSeq1-containing plasmid pool (pBiDir-mCherry-dEGFP-RzSeq1) which demonstrated clearly evident susceptibility to miR-21 mediated inhibition (Fig. 5.2B). This difference may have resulted from an inability of miR-21 to bind to its cognate seed sequence only target site in RzSeq2. Alternatively, it is possible that despite the intended miR-21 binding still occuring, miRNA-ribozyme interactions at this locus may have been insuffient and/or not appropriately configured to elicit repression of the corresponding transcript. Theoretically, any combination thereof could also account for this outcome. The argument in support of a lack of binding between miR-21 and RzSeq2 ribozymes is further strengthened by the fact that there is minimal difference between the levels of RzSeq2 self-cleavage as assessed by RNAseq in the absence or presence of miR-21 (Fig. 5.8). However, this would be contrary to biochemical data generated by the Zamore laboratory showing that miRNA-target transcript interactions based on mRNA complementarity with bases within the miRNA seed region alone is sufficient for near optimal binding affinity [95]. Notably though, it is known that RNA molecules can fold into stable secondary structures that are less accessible to miRNA binding,

an effect most prominent when a miRNA target site is located within a base-paired stem region of an RNA structure [168, 191]. RzSeq2 was designed such that the miRNA target seed sequence was located in a theoretically accessible loop region. However, because RzSeq2 ribozymes also contain a region of randomisation, it is possible that certain variants exhibit intramolecular interactions affecting the miRNA seed site. Although such an effect may inhibit miR-21 binding, it seems implausible that this would apply to the vast majority of variants within the ribozyme pool.

In order to directly investigate binding interactions between a miRNA and allosteric ribozymes, it may be advantageous to use an Argonaute CLIP approach [192]. Ago2 pulldown of cross-linked cellular lysates could be probed by RT-PCR for the presence of ribozyme-containing transcripts providing a direct readout for such interactions. This would enable more rational design of stem-loop structures in order to ensure that structural constraints do not preclude binding of miRNA complexes.

### 5.3.6 Using barcodes to enable unrestricted randomisation within the ribozyme sequence

The RNAseq selection techniques used in this chapter restricted the regions at which randomised sequences could be incorporated into the ribozyme backbones of the RzSeq1, RzSeq2 and RzSeq3 pools. In order to identify ribozyme variants and correctly assess their cleavage activity, the randomised region was required to fall within the 5' cleavage product. This was further restricted for the RzSeq3 selection experiment which utilised a structured stem-loop primer, thus imposing the additional requirement for a invariant sequence immediately upstream of the cleavage site.

An elegant solution to address these and related limitations would be to use barcode-tagging and parallel DNA/RNA sequencing. A random barcode would be located within the region of RNA to be sequenced, for example within the 3' UTR immediately upstream of the ribozyme insertion site. Under these circumstances, sequences within the ribozyme could be randomised freely with no associated data capture-related restrictions. Sequencing of the plasmid pool encoding the random ribozymes across a region encompassing both the barcode and the ribozyme, could then be used to generate a table directly linking a given unique barcode with a single variant ribozyme sequence. By this strategy, it would be possible to indirectly link ribozyme cleavage activity to the random sequence identity via a barcode.

# Chapter 6

# Conclusions

## 6.1 Achievements of the current study: new tools, novel strategies and a more compact miRNA-responsive positive sensor circuit

RNA is considered to be a malleable material and as such it has enabled synthetic biologists to create a diverse set of biologically active tools, taking cues from the abundant naturally occurring examples of functional RNA. The complexity of these synthetic systems has increased as modules have been built to increase the repertoire of molecules that can serve as input signals to such systems. At the same time, the wiring of multiple genetic components into compelex circuits has enhanced the downstream signal processing of which these are capable. Such circuits are now finding utility in a wide variety of bio-engineering applications from production of fuels to therepeutics [3]. There is now a great demand for the addition of new components to the synthetic biology toolbox to continue driving this creative process. Here we have attempted to generate modules that would enable the integration of information on the cellular abundance of an miRNA input signal of choice into a wider synthetic circuit. The rational behind this goal was the proposed technical limitations (see below) of available synthetic miRNA sensors and the huge and increasingly recognised importance of this class of regulatory molecule in governing numerous cellular processes involved in both health and disease [193].

Currently published tools for the detection of miRNA activity fall under two main classes; those by which the output signal is negatively correlated (negative sensor) and those by which it is positively correlated (positive sensor) with the input miRNA signal. The negative sensors exploit the endogenous repressive activity of miRNAs and function via a direct and simple mechanism whereby miRNA target sites are engineered into the transcript of a reporter gene

[99, 105, 145, 146]. Thus the expression of the reporter negatively correlates with the level of miRNA activity. Positive sensors utilise a more complex and indirect circuit architecture to enable a miRNA to positively regulate the expression of an output reporter gene. This is achieved through a two-component circuit whereby the miRNA acts to repress an intermediate regulator module which in turn functions to repress the final output [135, 148, 149]. Published implementations of miRNA positive sensor circuits have utilised regulation at both the transcriptional and post-transcriptional level and as a result of their complex input-to-output paths are likely to be slow to respond to changing signals. Here we sought to apply principles of RNA engineering to the design and construction of positive miRNA sensor circuits that function via a more condensed and direct architecture and in so doing have contributed two novel design frameworks for miRNA-responsive circuits.

In Chapter 3 we have employed a multi-component circuit architecture consisiting of an input miRNA, an intermediate regulator and a reporter gene output to build a positive miRNA sensor. In contrast to prior work, we have utilised a regulator (L7Ae) which acts to inhibit the translation of the output reporter (dEGFP) transcript to more direcly couple the input and output signals (Fig. 1.5). Thus all regulation in our circuit occurs at the post-transcriptional level; miRNA-mediated cleavage of the L7Ae regulator mRNA transcript, and L7Ae-mediated translational repression of the output protein. This circuit was fabricated through the engineering of two genetic NOT gates governing each of the post-transcriptional regulatory steps just mentioned. The performance of these NOT gates was subjected to optimisation by way of modulation of protein stability (destabilisation of L7Ae, Fig. 3.5B-C), and titration of the circuit component controlling the strength of the input signal (Fig. 3.5D). Finally the individual NOT gates were assembled in series and it was demonstrated that together these could function to regulate the expression of a dEGFP reporter in response to a specific miRNA input signal (Fig. 3.8A). Furthermore, we have shown that it is possible to simultaneously monitor the state of both coupled NOT gates using an ECFP tagged regulator (ECFP-L7Ae) and a dEGFP output. Such a circuit would serve as both a negative and positive sensor of the target miRNA activity and posses the qualities associated with each of these architectures.

In Chapter 4 and Chapter 5 we have described several processes for the construction of the most conceptually condensed circuit proposed to date for the positive detection of miRNA activity. These circuits employed a fused regulator/ouput module, consisiting of a candidate miRNA-regulated allosteric self-cleaving ribozyme located within a non-coding region of the output transcript. Firstly, a published allosteric ribozyme (NOT-1, [64]) was used to optimise the repression of a reporter gene coding for EGFP or its destabilised variant (dEGFP) in our chosen HEK293T cell culture system. Then techniques of *in vitro* selection were used to identify candidate of miRNA-responsive allosteric ribozymes given the amenable nature of

ribozymes to *in vitro* based selection methods for the identification of *de novo* generated novel allosteric variants [60, 150, 184]. This required the adaptation of existing techniques to precisely serve our purpose of massively enriching for miRNA-responsive allosteric ribozymes from within rationally designed ribozyme pools. These starting pools carried huge numbers of variants produced through randomisation of a short stretch of the ribozyme sequence.

In Chapter 4 an RT-PCR based selection approach was devised in which ribozymes carrying a short and specified randomised region were expressed in the context of a larger transcript within living cells in the presence or abscence the cognate miRNA ligand before RNA purification and *in vitro* based ribozyme selection. In essence, individual ribozyme variants within the randomised pool may or may not have have been able to undergo cleavage, and even those capable of auto-cleavage may have had this activity allosterically modified in the presence of the matching miRNA. This process was modelled on one described by Chen *et al.* [150] however a key difference being the way in which we selected for cleaved ribozymes. For this we repurposed structured stem-loop primers, which had previously been described for RT-qPCR quantification of miRNAs [158, 180], to discriminate between cleaved and uncleaved ribozymes and specifically prime the former for reverse transcriptase-driven conversion to cDNA. This novel application of stem-loop primers has enabled us to simplify the process required for the capture and regeneration of active ribozymes during selection such that the entire process could be completed by RT-PCR, as was the case for the selection of inactive ribozymes. The specificity of this stem-loop primer technique for the amplification of cleaved ribozymes was further validated through examination of our data as nine of nine randomly selected ribozymes that were characterised following enrichment were shown to be capable of repression of GFP fluorescence encoded by the coupled transcript (Fig. 4.10A). Indeed, given the nine nucleotide region of the ribozyme that was subjected to randomisation, it was expected that the vast majority of ribozyme variants within the starting pool would lack auto-cleavage activity, and hence would not be able to mediate repression of GFP fluorescence. Although the RT-PCR selection experiment described in Chapter 4 was unsuccessful in the identification of miRNA-responsive allosteric ribozymes, the approach described is versatile in nature and it can be envisioned that it may be used for the selection of allosteric ribozymes that respond to different types of ligand. In particular the novel application of structured stem-loop primers which we have described serves as a useful additional technique for the synthetic biology toolbox.

In Chapter 5 techniques were devised to enable the selection of miRNA-responsive allosteric ribozymes by analysis of high-throughput sequencing data based on a method described by Kobori *et al.* [184]. At their core, these methods used the same RT-PCR techniques described in the previous chapter for the generation of sequencing libraries containing

specifically either cleaved or uncleaved ribozyme variants. First a depletion screening technique aimed at selectively sequencing uncleaved ribozymes was used to select for allosteric ribozymes. The data obtained either in the presence of the cognate miRNA or its absence was subjected to comparative analysis to identify potential allosteric ribozyme candidates. Numerous such examples were identified. Analysis of 25 candidates mapping to each end of the catalytic activity spectrum demonstrated a good correlation between ribozyme sequence and catalytic activity, as inferred from the abundance of cleavage product measured by sequencing (Fig. 5.9B). This is consistent with the known close association between ribozyme sequence, structure and function. A similar level of correlation was however not observed between ribozyme sequence and the differential levels of ribozyme activities in the presence or absence of the cognate miRNA. Furthermore, individual characterisation of the 10 candidates exhibiting the most statistically robust decrease in cleavage activity (inceased counts of uncleaved ribozymes) in the presence of miR-21 failed to reveal ribozymes capable of regulating dEGFP expression in the desired manner (Fig. 5.10). On the contrary, reporter constructs carrying all 10 candidates were associated with reduced GFP fluorescence in the presence of miR-21 as compared to the control miRNA. In a final attempt, another selection method was devised for the enrichment screening of ribozymes by RNAseq to selectively identify variants exhibiting cleavage activity. This utilised a structured stem-loop primer modified to incorporate an UMI to enable analysis via a digital RNAseq pipeline. Consistent with previous descriptions of digital RNAseq [187–190], a marked reduction in noise was observed between technical replicate samples following application of digital RNAseq analysis as compared to processing of data obtained by conventional next generation sequencing (Fig. 5.13). Through RT-PCR facilitated enrichment screening coupled with digital RNAseq analysis we have identified 12 candidate miRNA-responsive allosteric ribozymes that are awaiting further testing. The techniques described in this chapter are all generally applicable to the selection of ribozymes responsive to any chosen ligand and therefore constitute important tools for the research community.

In summary, with the L7Ae based system (Chapter 3) we have fulfilled our goal of generating a more compact miRNA-responsive positive sensor circuit. Additionally, whilst we have not been successful in the generation of miRNA-responsive allosteric ribozymes (Chapters 4 and 5), we have identified 12 candidates worthy of further investigation, and have described a number of novel adaptations to *in vitro* selection techniques which may have important applications in future endeavours.

## 6.2 Future development

### 6.2.1 Scaling up circuit complexity to allow for multiple inputs, Boolean logic of choice, compact design and/or memory of an input signal

The modular devices described throughout this thesis lend themselves to the assembly of more complex circuits. This can be achieved through the use of one or more of multiple different miRNA sensitive modules, as have been described here, or the incorporation of alternative existing components from the synthetic biology toolkit.

An attractive next step would involve the fabrication of multi-input miRNA responsive devices. In certain situations, the detection of a single miRNA may not be sufficient to discriminate between cell types or states. In these cases greater resolving power can be attained by integrating the information from multiple independent miRNA input signals. Such multi-input miRNA sensitive circuits can be configured using various different circuit architectures to enable the processing of different forms of Boolean logic. For example, two different miRNA-responsive ribozymes could be linked in series within the 3' UTR of the transcript of the same reporter gene such that the transcript only becomes stabilised when both miRNA inputs allosterically inhibit their respective ribozymes (AND logic). Alternatively, two parallel reporter genes could each carry a single unique miRNA-responsive ribozyme such that together these would function to compute OR gate logic. For the identification of cancerous cells, in which the expression levels of many miRNA are only subtly altered, the multi-input sensors described above may be of particular importance.

Further complexity can be introduced through the use of pre-existing RNA based synthetic components. One highly attractive module to incorporate would be a perfect miRNA target site for the negative detection of miRNA activity. A transcript possessing such a target site would be subject to high-level miRNA-mediated repression as a consequence of an elevated rate of transcript degradation. Using such a module, a single molecule sensor consisting of a transcript with both a miRNA-responsive ribozyme and a perfectly complementary miRNA target site versus a distinct miRNA entity could be built for the simultaneous detection of both the over-expression of one miRNA and the under-expression of a second unrelated miRNA. The highly compact nature of this circuit and its correspondingly small genetic footprint would make it highly amenable to scaling. In addition, certain delivery vectors such as lentivirus are bound by restrictive packaging limits and the field would therefore benefit greatly from the development of more compact biological circuits.

### 6.2.2 miRNA-directed apoptosis: an example of a therapeutically relevant outcome

The ultimate aim of this work was to construct synthetic circuits within the cell that are able to rewire cellular behaviour in response to specific miRNA input signals. These could act to rewire metabolic or cellular pathways, or more drastically to induce apoptosis in specific populations of cells. In addition, exchanging an effector module from one of the above conceptual circuits for a biomarker-producing component could render the device with diagnostic capabilities.

As an example, if the desired end result was directed apoptosis of the miRNA-positive cell population, the dEGFP output protein used throughout this study would be replaced by an output protein capable of inducing cell death. Endogenous proteins that function within the apoptotic signaling pathway could be exploited for this purpose. The overexpression of hBax, Bad or Bid can be used to transgenically induce apoptotic cell death. Alternatively, bacterial protein toxins can be used to achieve the same end goal. The Diphtheria toxin A chain (DTA) is one of the most potent cellular toxins, a single molecule of which is capable of inducing cell death. Through preliminary work with this toxin, it was found to be too potent such that even minimal expression from a leaky promoter resulted in fatal consequences for the cells involved. Given our interest in synthetic biological circuits capable of inducing directed apoptosis and our awareness of the importance of Boolean logic-conferred safeguards, we propose that HSV-tk, or an alternative conditionally toxic protein, would be a strong candidate output protein for an eventual therapeutically active circuit. HSV-tk alone is not toxic, however this viral enzyme can specifically phosphorylate the exogenously supplied pro-drug ganciclovir, converting it into a toxic form. Thus, if HSV-tk serves as the output from a miRNA-responsive circuit, targeted cellular apoptosis would result from AND gate logic, with any triggered cell death arising only from the simultaneous obligatory presence of both the cognate miRNA and exogenously delivered ganciclovir.

# References


[1] Encode Project Consortium, Ian Dunham, Anshul Kundaje, Shelley F Aldred, Patrick J Collins, Carrie A Davis, Francis Doyle, Charles B Epstein, Seth Frietze, Jennifer Harrow, Rajinder Kaul, Jainab Khatun, Bryan R Lajoie, Stephen G Landt, Burn-Kyu K Lee, Florencia Pauli, Kate R Rosenbloom, Peter Sabo, Alexias Safi, Amartya Sanyal, Noam Shoresh, Jeremy M Simon, Lingyun Song, Nathan D Trinklein, Robert C Altshuler, Ewan Birney, James B Brown, Chao Cheng, Sarah Djebali, Xianjun Dong, Ian Dunham, Jason Ernst, Terrence S Furey, Mark Gerstein, Belinda Giardine, Melissa Greven, Ross C Hardison, Robert S Harris, Javier Herrero, Michael M Hoffman, Sowmya Iyer, Manolis Kelllis, Jainab Khatun, Pouya Kheradpour, Anshul Kundaje, Timo Lassman, Qunhua Li, Xinying Lin, Georgi K Marinov, Angelika Merkel, Ali Mortazavi, Stephen C Parker, Timothy E Reddy, Joel Rozowsky, Felix Schlesinger, Robert E Thurman, Jie Wang, Lucas D Ward, Troy W Whitfield, Steven P Wilder, Weisheng Wu, Hualin S Xi, Kevin Y Yip, Jiali Zhuang, Bradley E Bernstein, Ewan Birney, Ian Dunham, Eric D Green, Chris Gunter, Michael Snyder, Michael J Pazin, Rebecca F Lowdon, Laura A Dillon, Leslie B Adams, Caroline J Kelly, Julia Zhang, Judith R Wexler, Eric D Green, Peter J Good, Elise A Feingold, Bradley E Bernstein, Ewan Birney, Gregory E Crawford, Job Dekker, Laura Elinitski, Peggy J Farnham, Mark Gerstein, Morgan C Giddings, Thomas R Gingeras, Eric D Green, Roderic Guigó, Ross C Hardison, Tomothy J Hubbard, Manolis Kellis, W J Kent, Jason D Lieb, Elliott H Margulies, Richard M Myers, Michael Snyder, and John A Starnatoyannopoulos. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, 2012.

[2] R J Taft, M Pheasant, and J S Mattick. The relationship between non-protein coding DNA and eukaryotic complexity. *BioEssays*, 29(3):288–299, 2007.

[3] D Ewen Cameron, Caleb J Bashor, and James J Collins. A brief history of synthetic biology. *Nature Reviews Microbiology*, 12(5):381–390, 2014.

[4] Thomas R. Cech and Joan A. Steitz. The Noncoding RNA Revolution - Trashing Old Rules to Forge New Ones. *Cell*, 157(1):77–94, 2014.

[5] Mark G Caprara and Timothy W Nilsen. RNA: versatility in form and function. *Nature Structural & Molecular Biology*, 7(10):831–833, 2000.

[6] C Tuerk and L Gold. Systematic Evolution of Ligands By Exponential Enrichment - Rna Ligands To Bacteriophage-T4 Dna-Polymerase. *Science*, 249(4968):505–510, 1990.

[7] A D Ellington and J W Szostak. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287):818–822, 1990.

[8] David P Bartel and Jack W Szostak. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261:1411–8, 1993.

[9] Walter Gilbert. Origin of life: The RNA world. *Nature*, 319(6055):618, 1986.

[10] Maumita Mandal and Ronald R Breaker. Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology*, 5(6):451–463, 2004.

[11] Lin He and Gregory J Hannon. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5(7):522–531, 2004.

[12] Hirohide Saito, Tetsuhiro Kobayashi, Tomoaki Hara, Yoshihiko Fujita, Karin Hayashi, Rie Furushima, and Tan Inoue. Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nature Chemical Biology*, 6(1):71–78, 2010.

[13] Hirohide Saito, Yoshihiko Fujita, Shunnichi Kashida, Karin Hayashi, and Tan Inoue. Synthetic human cell fate regulation by protein-driven RNA switches. *Nature Communications*, 2(1):160, 2011.

[14] Franz Narberhaus, Torsten Waldminghaus, and Saheli Chowdhury. RNA thermometers. *FEMS Microbiology Reviews*, 30(1):3–16, 2006.

[15] J Johansson, P Mandin, A Renzoni, C Chiaruttini, M Springer, and P Cossart. An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. *Cell*, 110(5):551–561, 2002.

[16] Saheli Chowdhury, Christophe Maris, Frédéric H-T Allain, and Franz Narberhaus. Molecular basis for temperature sensing by an RNA thermometer. *The EMBO journal*, 25(11):2487–97, 2006.

[17] Athanasios Saragliadis, Stefanie S Krajewski, Charlotte Rehm, Franz Narberhaus, and Jörg S Hartig. Thermozymes: Synthetic RNA thermometers based on ribozyme activity. *RNA Biology*, 10(6):1010–6, 2013.

[18] Alexander Serganov and Dinshaw J Patel. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Reviews Genetics*, 8(10):776–790, 2007.

[19] Peggy J Farnham and Terry Platt. Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic Acids Research*, 9(3):563–577, 1981.

[20] Maumita Mandal and Ronald R Breaker. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural & Molecular Biology*, 11(1):29–35, 2004.

[21] Wade C Winkler, Ali Nahvi, Narasimhan Sudarsan, Jeffrey E Barrick, and Ronald R Breaker. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structural & Molecular Biology*, 10(9):701–707, 2003.

[22] Wade Winkler, Ali Nahvi, and Ronald R Breaker. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419(6910):952–6, 2002.

[23] Ming T Cheah, Andreas Wachter, Narasimhan Sudarsan, and Ronald R Breaker. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143):497–500, 2007.

[24] M. Mandal. A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science*, 306(5694):275–279, 2004.

[25] Rüdiger Welz and Ronald R Breaker. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *RNA*, 13(4):573–82, 2007.

[26] N. Sudarsan, M. C. MC Hammond, KF F. Block, R. Welz, JE E. Barrick, A. Roth, and RR R. Breaker. Tandem riboswitch architectures exhibit complex gene control functions. *Science*, 314(5797):300–4, 2006.

[27] Thomas R. Cech, Arthur J. Zaug, and Paula J. Grabowski. In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell*, 27(3 PART 2):487–496, 1981.

[28] C Guerriertakada, K Gardiner, T Marsh, N Pace, and S Altman. The Rna Moiety of Ribonuclease-P Is the Catalytic Subunit of the Enzyme. *Cell*, 35(3):849–857, 1983.

[29] Gerry A Prody, John T Bakos, Jamal M Buzayan, Irving R Schneider, and G Bruening. Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231(1975):1577–1580, 1986.

[30] Jonathan Perreault, Zasha Weinberg, Adam Roth, Olivia Popescu, Pascal Chartrand, Gerardo Ferbeyre, and Ronald R. Breaker. Identification of Hammerhead Ribozymes in All Domains of Life Reveals Novel Structural Variations. *PLoS Computational Biology*, 7(5):e1002031, 2011.

[31] Adam Roth, Zasha Weinberg, Andy G Y Chen, Peter B Kim, Tyler D Ames, and Ronald R Breaker. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nature Chemical Biology*, 10(1):56–60, 2014.

[32] Zasha Weinberg, Peter B Kim, Tony H Chen, Sanshu Li, Kimberly A Harris, Christina E Lünse, and Ronald R Breaker. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nature Chemical Biology*, 11(8):606–610, 2015.

[33] Wade C Winkler, Ali Nahvi, Adam Roth, Jennifer a Collins, and Ronald R Breaker. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428(6980):281–286, 2004.

[34] Jeffrey E Barrick, Keith a Corbino, Wade C Winkler, Ali Nahvi, Maumita Mandal, Jennifer Collins, Mark Lee, Adam Roth, Narasimhan Sudarsan, Inbal Jona, J Kenneth Wickiser, and Ronald R Breaker. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6421–6426, 2004.

[35] Sławomir Milewski. Glucosamine-6-phosphate synthase-the multi-facets enzyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1597(2):173–192, 2002.

[36] C Flamm, W Fontana, I L Hofacker, and P Schuster. RNA folding at elementary step resolution. *RNA*, 6(3):325–338, 2000.

[37] Ivo L Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Research*, 31(13):3429–3431, 2003.

[38] Rhiju Das and David Baker. Automated de novo prediction of native-like RNA tertiary structures. *Proceedings of the National Academy of Sciences of the United States of America*, 104(37):14664–9, 2007.

[39] Marc Parisien and Francois Major. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, 452(7183):51–55, 2008.

[40] Joseph N Zadeh, Conrad D Steenberg, Justin S Bois, Brian R Wolfe, Marshall B Pierce, Asif R Khan, Robert M Dirks, and Niles A Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, 2011.

[41] Geoffrey Werstuck and Michael R. Green. Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions. *Science*, 282(5387):296–298, 1998.

[42] Jin Tang and Ronald R. Breaker. Rational design of allosteric ribozymes. *Chemistry & Biology*, 4(6):453–459, 1997.

[43] G A Soukup and R R Breaker. Engineering precision RNA molecular switches. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7):3584–9, 1999.

[44] J Ciesiolka and M Yarus. Small RNA-divalent domains. *RNA*, 2(8):785–793, 1996.

[45] D Nieuwlandt, M Wecker, and L Gold. In-Vitro Selection of RNA Ligands to Substance-P. *Biochemistry*, 34(16):5651–5659, 1995.

[46] S Ringquist, T Jones, E E Snyder, T Gibson, I Boni, and L Gold. High-affinity RNA ligands to Escherichia coli ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry*, 34(11):3640–3648, 1995.

[47] W Pan, R C Craven, Q Qiu, C B Wilson, J W Wills, S Golovine, and J F Wang. Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 92(25):11509–13, 1995.

[48] Kevin N Morris, Kirk B Jensen, Carol M Julin, Michael Weil, and Larry Gold. High affinity ligands from in vitro selection : Complex targets. *RNA*, 95:2902–2907, 1998.

[49] Niles Lehman and Gerald F Joyce. Evolution in vitro of an RNA enzme with altered metal dependence. *Nature*, 361(14):182–185, 1993.

[50] J Tsang and G F Joyce. Evolutionary optimization of the catalytic properties of a DNA-cleaving ribozyme. *Biochemistry*, 33(19):5966–5973, 1994.

[51] Karen B. Chapman and Jack W. Szostak. Isolation of a ribozyme with 5 prime-5 prime ligase activity. *Chemistry & Biology*, 2(5):325–333, 1995.

[52] J Tsang and G F Joyce. Specialization of the DNA-cleaving activity of a group I ribozyme through in vitro evolution. *Journal of Molecular Biology*, 262(1):31–42, 1996.

[53] A J Hager and J W Szostak. Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chemistry & Biology*, 4(8):607–617, 1997.

[54] Thomas Tuschl, Phillip A. Sharp, and David P. Bartel. Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. *The EMBO journal*, 17(9):2637–2650, 1998.

[55] Ronald R. Breaker and Gerald F. Joyce. A DNA enzyme that cleaves RNA. *Chemistry & Biology*, 1(4):223–229, 1994.

[56] Ronald R. Breaker and Gerald F. Joyce. A DNA enzyme with Mg2+-dependent RNA phosphoesterase activity. *Chemistry & Biology*, 2(10):655–660, 1995.

[57] James Attwater, Shunsuke Tagami, Michiko Kimoto, Kyle Butler, Eric T Kool, Jesper Wengel, Piet Herdewijn, Ichiro Hirao, and Philipp Holliger. Chemical fidelity of an RNA polymerase ribozyme. *Chemical Science*, 4(7):2804–2814, 2013.

[58] Alexander I Taylor, Vitor B Pinheiro, Matthew J Smola, Alexey S Morgunov, Sew Peak-Chew, Christopher Cozens, Kevin M Weeks, Piet Herdewijn, and Philipp Holliger. Catalysts from synthetic genetic polymers. *Nature*, 518(7539):427–430, 2014.

[59] Noam Prywes, Yale S Michaels, Ayan Pal, Seung Soo Oh, and Jack W Szostak. Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying. *Chemical Communications*, 52(39):6529–6532, 2016.

[60] Benedikt Klauser, Janina Atanasov, Lena K. Siewert, and Jörg S. Hartig. Ribozyme-Based Aminoglycoside Switches of Gene Expression Engineered by Genetic Selection in S. cerevisiae. *ACS Synthetic Biology*, 4(5):516–525, 2015.

[61] Mary G Wallis, Uwe von Ahsen, Renee Schroeder, and Michael Famulok. A novel RNA motif for neomycin recognition. *Chemistry & Biology*, 2(8):543–552, 1995.

[62] Julia E Weigand, Martin Sanchez, Ewald-Bernd Gunnesch, Sabrina Zeiher, Renee Schroeder, and Beatrix Suess. Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, 14(1):89–97, 2008.

[63] Laising Yen, Jennifer Svendsen, Jeng-shin Lee, and John T Gray. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature*, 431:471–476, 2004.

[64] Robert Penchovsky and Ronald R Breaker. Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nature Biotechnology*, 23(11):1424–1433, 2005.

[65] Samuel G. Rouleau, Rachel Jodoin, Martin Bisaillon, and Jean Pierre Perreault. Programming a highly structured ribozyme into complex allostery using RNA oligonucleotides. *ACS Chemical Biology*, 7(11):1802–1806, 2012.

[66] Benedikt Klauser and Jörg S. Hartig. An engineered small RNA-mediated genetic switch based on a ribozyme expression platform. *Nucleic Acids Research*, 41(10):5542–5552, 2013.

[67] Andres A Rojas, Alejandro Vazquez-Tello, Gerardo Ferbeyre, Federica Venanzetti, Lutz Bachmann, Bruno Paquin, Valerio Sbordoni, and Robert Cedergren. Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. *Nucleic Acids Research*, 28(20):4037–4043, 2000.

[68] Gerardo Ferbeyre, James M Smith, and Robert Cedergren. Schistosome Satellite DNA Encodes Active Hammerhead Ribozymes. *Molecular and Cellular Biology*, 18(7):3880–3888, 1998.

[69] Anastasia Khvorova, Aurélie Lescoute, Eric Westhof, and Sumedha D Jayasena. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature Structural & Molecular Biology*, 10(9):708–712, 2003.

[70] Monika Martick and William G. Scott. Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell*, 126(2):309–320, 2006.

[71] Birte Vester and Jesper Wengel. LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary RNA and DNA. *Biochemistry*, 43(42):13233–13241, 2004.

[72] J S McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7):1105–1119, 1990.

[73] Markus Wieland and Jörg S. Hartig. Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angewandte Chemie - International Edition*, 47(14):2604–2607, 2008.

[74] Thomas Franch, M Petersen, E Gerhart H Wagner, J P Jacobsen, and Kenn Gerdes. Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *Journal of Molecular Biology*, 294(5):1115–25, 1999.

[75] Yoontae Lee, Kipyoung Jeon, Jun-Tae Lee, Sunyoung Kim, and V Narry Kim. MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO Journal*, 21(17):4663–4670, 2002.

[76] Antony Rodriguez, Sam Griffiths-Jones, Jennifer L Ashurst, and Allan Bradley. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research*, 14(10a):1902–1910, 2004.

[77] Tuan Anh Nguyen, Myung Hyun Jo, Yeon-Gil Choi, Joha Park, S Chul Kwon, Sungchul Hohng, V Narry Kim, and Jae-Sung Woo. Functional Anatomy of the Human Microprocessor. *Cell*, 161(6):1374–87, 2015.

[78] Ahmet M Denli, Bastiaan B J Tops, Ronald H a Plasterk, René F Ketting, and Gregory J Hannon. Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231–5, 2004.

[79] Richard I Gregory, Kai-Ping Yan, Govindasamy Amuthan, Thimmaiah Chendrimada, Behzad Doratotaj, Neil Cooch, and Ramin Shiekhattar. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014):235–240, 2004.

[80] Jinju Han, Yoontae Lee, Kyu Hyun Yeom, Young Kook Kim, Hua Jin, and V. Narry Kim. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development*, 18(24):3016–3027, 2004.

[81] Markus Landthaler, Abdullah Yalcin, and Thomas Tuschl. The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis. *Current Biology*, 14(23):2162–2167, 2004.

[82] Jaroslaw Blaszczyk, Joseph E. Tropea, Mikhail Bubunenko, Karen M. Routzahn, David S. Waugh, Donald L. Court, and Xinhua Ji. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure*, 9(12):1225–1236, 2001.

[83] Carole Gwizdek, Batool Ossareh-Nazari, Amy M. Brownawell, Alain Doglio, Edouard Bertrand, Ian G. Macara, and Catherine Dargemont. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *Journal of Biological Chemistry*, 278(8):5505–5508, 2003.

[84] Yan Zeng and Bryan R. Cullen. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Research*, 32(16):4776–4785, 2004.

[85] G Hutvágner, J McLachlan, A E Pasquinelli, E Bálint, T Tuschl, and P D Zamore. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, 293(5531):834–838, 2001.

[86] René F Ketting, Sylvia E J Fischer, Emily Bernstein, René F Ketting, Sylvia E J Fischer, Emily Bernstein, Titia Sijen, Gregory J Hannon, and Ronald H A Plasterk. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C . elegans. *Genes & Development*, 15(20):2654–2659, 2001.

[87] S W Knight and B L Bass. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science*, 293(5538):2269–71, 2001.

[88] Jong-Eun Park, Inha Heo, Yuan Tian, Dhirendra K Simanshu, Hyeshik Chang, David Jee, Dinshaw J Patel, and V Narry Kim. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 475(7355):201–205, 2011.

[89] Elad Elkayam, Claus-D Kuhn, Ante Tocilj, Astrid D Haase, Emily M Greene, and Gregory J Hannon. The structure of human Argonaute-2 in complex with miR-20a. *Cell*, 150(1):100–110, 2012.

[90] Ji-Joon Song, Stephanie K Smith, Gregory J Hannon, and Leemor Joshua-Tor. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, 305(5689):1434–1437, 2004.

[91] Jidong Liu, Michelle a Carmell, Fabiola V Rivas, Carolyn G Marsden, J Michael Thomson, Ji-Joon Song, Scott M Hammond, Leemor Joshua-Tor, and Gregory J Hannon. Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, 305(5689):1437–1441, 2004.

[92] Anastasia Khvorova, Angela Reynolds, and Sumedha D. Jayasena. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115(2):209–216, 2003.

[93] Dianne S Schwarz, György Hutvágner, Tingting Du, Zuoshang Xu, Neil Aronin, and Phillip D Zamore. Asymmetry in the assembly of the RNAi enzyme complex. - PubMed - NCBI. *Cell*, 115(2):199–208, 2003.

[94] John G Doench and Phillip A Sharp. Specificity of microRNA target selection in translational repression. *Genes*, 504(5):504–511, 2004.

[95] Liang Meng Wee, C. Fabián Flores-Jasso, William E. Salomon, and Phillip D. Zamore. Argonaute Divides Its RNA Guide into Domains with Distinct Functions and RNA-Binding Properties. *Cell*, 151(5):1055–1067, 2012.

[96] J B Ma, Y R Yuan, G Meister, Y Pei, T Tuschl, and D J Patel. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature*, 434(7033):666–670, 2005.

[97] James S Parker, S Mark Roe, and David Barford. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, 434(7033):663–6, 2005.

[98] Wei Yang and Thomas A. Steitz. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*, 3(2):131–134, 1995.

[99] György Hutvágner and Phillip D Zamore. A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, 297(5589):2056–60, 2002.

[100] Benjamin P Lewis, I-hung Shih, Matthew W Jones-Rhoades, David P Bartel, and Christopher B Burge. Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7):787–798, 2003.

[101] Stephen W Eichhorn, Huili Guo, Sean E Mcgeary, Ricard A Rodriguez-mias, and Chanseok Shin. mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. *Mollecular Cell*, 56(1):104–115, 2014.

[102] Sergej Djuranovic, Ali Nahvi, and Rachel Green. miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. *Science*, 336(6078):237 LP – 240, 2012.

[103] Ariel A Bazzini, Miler T Lee, and Antonio J Giraldez. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science*, 336(6078):233 LP – 237, 2012.

[104] Brenda J Reinhart, Frank J Slack, Michael Basson, Amy E Pasquinelli, Jill C Bettinger, Ann E Rougvie, H Robert Horvitz, and Gary Ruvkun. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403(6772):901–906, 2000.

[105] Julius Brennecke, David R Hipfner, Alexander Stark, Robert B Russell, and Stephen M Cohen. bantam Encodes a Developmentally Regulated microRNA that Controls Cell Proliferation and Regulates the Proapoptotic Gene hid in Drosophila. *Cell*, 113(1):25–36, 2003.

[106] Yan Zeng, Rui Yi, and Bryan R Cullen. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proceedings of the National Academy of Sciences*, 100(17):9779–9784, 2003.

[107] John G Doench, Christian P Petersen, and Phillip A Sharp. siRNAs can function as miRNAs. *Genes & Development*, 17(4):438–442, 2003.

[108] Nicholas T Ingolia. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, 15(3):205–213, 2014.

[109] P H Olsen and V Ambros. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental biology*, 216(2):671–680, 1999.

[110] C P Petersen, M E Bordeleau, J Pelletier, and P A Sharp. Short RNAs Repress Translation after Initiation in Mammalian Cells. *Molecular Cell*, 4(4):533–542, 2006.

[111] R. S. Pillai. Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells. *Science*, 309(5740):1573–1576, 2005.

[112] Richard J Jackson, Christopher U T Hellen, and Tatyana V Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2):113–127, 2010.

[113] H A Meijer, Y W Kong, W T Lu, A Wilczynska, R V Spriggs, S W Robinson, J D Godfrey, A E Willis, and M Bushell. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science*, 340(6128):82–85, 2013.

[114] S Bagga, J Bracht, S Hunter, K Massirer, J Holtz, R Eachus, and A E Pasquinelli. Regulation by let-7 and lin-4 miRNAs Results in Target mRNA Degradation. *Cell*, 122(4):553–563, 2005.

[115] A. J. Giraldez. Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs. *Science*, 312(5770):75–79, 2006.

[116] Jan Rehwinkel, Pavel Natalin, Alexander Stark, Julius Brennecke, Stephen M Cohen, and Elisa Izaurralde. Genome-Wide Analysis of mRNAs Regulated by Drosha and Argonaute Proteins in Drosophila melanogaster. *Molecular and Cellular Biology*, 26(8):2965–2975, 2006.

[117] Daniela Schmitter, Jody Filkowski, Alain Sewer, Ramesh S Pillai, Edward J Oakeley, Mihaela Zavolan, Petr Svoboda, and Witold Filipowicz. Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Research*, 34(17):4801–4815, 2006.

[118] Matthias Selbach, Bjorn Schwanhausser, Nadine Thierfelder, Zhuo Fang, Raya Khanin, and Nikolaus Rajewsky. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63, 2008.

[119] Janina Pfaff, Janosch Hennig, Franz Herzog, Ruedi Aebersold, Michael Sattler, Dierk Niessing, and Gunter Meister. Structural features of Argonaute-GW182 protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 110(40):E3770–9, 2013.

[120] Marina Chekulaeva, Hansruedi Mathys, Jakob T Zipprich, Jan Attig, Marija Colic, Roy Parker, and Witold Filipowicz. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nature Structural & Molecular Biology*, 18(11):1218–1226, 2011.

[121] Mary Christie, Andreas Boland, Eric Huntzinger, Oliver Weichenrieder, and Elisa Izaurralde. Structure of the PAN3 Pseudokinase Reveals the Basis for Interactions with the PAN2 Deadenylase and the GW182 Proteins. *Molecular Cell*, 51(3):360–373, 2013.

[122] Nicole L Garneau, Jeffrey Wilusz, and Carol J Wilusz. The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology*, 8(2):113–126, 2007.

[123] Jeff Coller and Roy Parker. Eukaryotic mRNA Decapping. *Annual Review of Biochemistry*, 73(1):861–890, 2004.

[124] Rebecca L Skalsky and Bryan R Cullen. Viruses, microRNAs, and Host Interactions. *Annual Review of Microbiology*, 64(1):123–141, 2010.

[125] Yoshimasa Saito and Hidetsugu Saito. MicroRNAs in cancers and neurodegenerative disorders, 2012.

[126] Michael P Nicoll, João T Proença, and Stacey Efstathiou. The molecular basis of herpes simplex virus latency. *FEMS Microbiology Reviews*, 36(3):684 LP – 705, 2012.

[127] Lawrence S Young and Paul G Murray. Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene*, 22(33):5108–5121.

[128] Hong-Jie Yang, Tie-Jun Huang, Chang-Fu Yang, Li-Xia Peng, Ran-Yi Liu, Guang-Da Yang, Qiao-Qiao Chu, Jia-Ling Huang, Na Liu, Hong-Bing Huang, Zhen-Yu Zhu, Chao-Nan Qian, and Bi-Jun Huang. Comprehensive profiling of Epstein-Barr virus-encoded miRNA species associated with specific latency types in tumor cells. *Virology Journal*, 10(1):1–13, 2013.

[129] Stephanie Barth, Thorsten Pfuhl, Alfredo Mamiani, Claudia Ehses, Klaus Roemer, Elisabeth Kremmer, Christoph Jaeker, Julia Hoeck, Gunter Meister, and Friedrich A. Graesser. Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Research*, 36(2):666–675, 2008.

[130] George Adrian Calin, Cinzia Sevignani, Calin Dan Dumitru, Terry Hyslop, Evan Noch, Sai Yendamuri, Masayoshi Shimizu, Sashi Rattan, Florencia Bullrich, Massimo Negrini, and Carlo M Croce. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2999–3004, 2004.

[131] Lin He, J Michael Thomson, Michael T Hemann, Eva Hernando-Monge, David Mu, Summer Goodson, Scott Powers, Carlos Cordon-Cardo, Scott W Lowe, Gregory J Hannon, and Scott M Hammond. A microRNA polycistron as a potential human oncogene. *Nature*, 435(7043):828–33, 2005.

[132] Li Ma, Julie Teruya-Feldstein, and Robert a Weinberg. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*, 449(7163):682–8, 2007.

[133] Kati Pulkkinen, Tarja Malm, Mikko Turunen, Jari Koistinaho, and Seppo Yl??-Herttuala. Hypoxia induces microRNA miR-210 in vitro and in vivo. Ephrin-A3 and neuronal pentraxin 1 are potentially regulated by miR-210. *FEBS Letters*, 582(16):2397–2401, 2008.

[134] Stefanie S Jeffrey. Cancer biomarker profiling with microRNAs. *Nature Biotechnology*, 26(4):400–401, 2008.

[135] Zhen Xie, Liliana Wroblewska, Laura Prochazka, Ron Weiss, and Yaakov Benenson. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science*, 333(6047):1307–1311, 2011.

[136] Mariana Lagos-Quintana, Reinhard Rauhut, Abdullah Yalcin, Jutta Meyer, Winfried Lendeckel, and Thomas Tuschl. Identification of Tissue-Specific MicroRNAs from Mouse. *Current Biology*, 12(9):735–739, 2002.

[137] Catherine L Jopling, Sylvia Schütz, and Peter Sarnow. Position-Dependent Function for a Tandem MicroRNA miR-122-Binding Site Located in the Hepatitis C Virus RNA Genome. *Cell Host & Microbe*, 4(1):77–85, 2008.

[138] Jura Inga Henke, Dagmar Goergen, Junfeng Zheng, Yutong Song, Christian G Schüttler, Carmen Fehr, Christiane Jünemann, and Michael Niepmann. microRNA-122 stimulates translation of hepatitis C virus RNA. *The EMBO Journal*, 27(24):3300 LP – 3310, 2008.

[139] Kara L Norman and Peter Sarnow. Modulation of Hepatitis C Virus RNA Abundance and the Isoprenoid Biosynthesis Pathway by MicroRNA miR-122 Involves Distinct Mechanisms. *Journal of Virology*, 84(1):666–670, 2010.

[140] Joacim Elmen, Morten Lindow, Sylvia Schutz, Matthew Lawrence, Andreas Petri, Susanna Obad, Marie Lindholm, Maj Hedtjarn, Henrik Frydenlund Hansen, Urs Berger, Steven Gullans, Phil Kearney, Peter Sarnow, Ellen Marie Straarup, and Sakari Kauppinen. LNA-mediated microRNA silencing in non-human primates. *Nature*, 452(7189):896–899, 2008.

[141] Harry L A Janssen, Hendrik W Reesink, Eric J Lawitz, Stefan Zeuzem, Maribel Rodriguez-Torres, Keyur Patel, Adriaan J van der Meer, Amy K Patick, Alice Chen, Yi Zhou, Robert Persson, Barney D King, Sakari Kauppinen, Arthur A Levin, and Michael R Hodges. Treatment of HCV Infection by Targeting MicroRNA. *New England Journal of Medicine*, 368(18):1685–1694, 2013.

[142] Stefano Volinia, George A Calin, Chang-Gong Liu, Stefan Ambs, Amelia Cimmino, Fabio Petrocca, Rosa Visone, Marilena Iorio, Claudia Roldo, Manuela Ferracin, Robyn L Prueitt, Nozumu Yanaihara, Giovanni Lanza, Aldo Scarpa, Andrea Vecchione, Massimo Negrini, Curtis C Harris, and Carlo M Croce. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2257–2261, 2006.

[143] Pedro P Medina, Mona Nolde, and Frank J Slack. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature*, 467(7311):86–90, 2010.

[144] K Musilova and M Mraz. MicroRNAs in B-cell lymphomas: how a complex biology gets more complex. *Leukemia*, 29(5):1004–1017, 2015.

[145] Yan Zeng, Eric J. Wagner, and Bryan R. Cullen. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Molecular Cell*, 9(6):1327–1333, 2002.

[146] Jennifer H Mansfield, Brian D Harfe, Robert Nissen, John Obenauer, Jalagani Srineel, Aadel Chaudhuri, Raphael Farzan-Kashani, Michael Zuker, Amy E Pasquinelli, Gary Ruvkun, Phillip A Sharp, Clifford J Tabin, and Michael T McManus. MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nature Genetics*, 36(10):1079–1083, 2004.

[147] G Mullokandov, A Baccarini, A Ruzo, A D Jayaprakash, N Tung, B Israelow, M J Evans, R Sachidanandam, and B D Brown. High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nature Methods*, 9(8):840–846, 2012.

[148] Mario Amendola, Alice Giustacchini, Bernhard Gentner, and Luigi Naldini. A double-switch vector system positively regulates transgene expression by endogenous microRNA expression (miR-ON vector). *Molecular therapy : the journal of the American Society of Gene Therapy*, 21(5):934–46, 2013.

[149] S Ezzine, G Vassaux, B Pitard, B Barteau, J M Malinge, P Midoux, C Pichon, and P Baril. RILES, a novel method for temporal analysis of the in vivo regulation of miRNA expression. *Nucleic Acids Research*, 41(20):e192, 2013.

[150] H C Chen, G H Chen, Y H Chen, W L Liao, C Y Liu, K P Chang, Y S Chang, and S J Chen. MicroRNA deregulation and pathway alterations in nasopharyngeal carcinoma. *British Journal of Cancer*, 100, 2009.

[151] Simon Ausländer, Pascal Stücheli, Charlotte Rehm, David Ausländer, Jörg S Hartig, and Martin Fussenegger. A general design strategy for protein-responsive riboswitches in mammalian cells. *Nature Methods*, 11(11):1154–60, 2014.

[152] Brent Townshend, Andrew B Kennedy, Joy S Xiang, and Christina D Smolke. High-throughput cellular RNA device engineering. *Nature Methods*, 12(10):989–994, 2015.

[153] Andreas Untergasser, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C Faircloth, Maido Remm, and Steven G Rozen. Primer3 - new capabilities and interfaces. *Nucleic Acids Research*, 40(15):e115–e115, 2012.

[154] Roman Rydzanicz, X. Sharon Zhao, and Philip E. Johnson. Assembly PCR oligo maker: A tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. *Nucleic Acids Research*, 33(SUPPL. 2):W521–W525, 2005.

[155] O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, and F. Zhang. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*, 343(6166):84–87, 2014.

[156] Douglas A Rubinson, Christopher P Dillon, Adam V Kwiatkowski, Claudia Sievers, Lili Yang, Johnny Kopinja, Dina L Rooney, Mingdi Zhang, Melanie M Ihrig, Michael T McManus, Frank B Gertler, Martin L Scott, and Luk Van Parijs. A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genetics*, 33(3):401–406, 2003.

[157] A. Radu Aricescu, Weixian Lu, and E. Yvonne Jones. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D: Biological Crystallography*, 62(10):1243–1250, 2006.

[158] Zsolt Czimmerer, Julianna Hulvely, Zoltan Simandi, Eva Varallyay, Zoltan Havelda, Erzsebet Szabo, Attila Varga, Balazs Dezso, Maria Balogh, Attila Horvath, Balint Domokos, Zsolt Torok, Laszlo Nagy, and Balint L. Balint. A Versatile Method to Design Stem-Loop Primer-Based Quantitative PCR Assays for Detecting Small Regulatory RNA Molecules. *PLoS ONE*, 8(1):e55168, 2013.

[159] Khalid K Alam, Jonathan L Chang, and Donald H Burke. FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Molecular Therapy-Nucleic Acids*, 4(August 2014):e230, 2015.

[160] M Kozak. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 83(9):2850–2854, 1986.

[161] Simon Ausländer, David Ausländer, Marius Müller, Markus Wieland, and Martin Fussenegger. Programmable single-cell mammalian biocomputers. *Nature*, 487(7405):123–7, 2012.

[162] Mario Roederer. Compensation in flow cytometry. *Current protocols in cytometry / editorial board, J. Paul Robinson, managing editor ... [et al.]*, Chapter 1:Unit 1.14, 2002.

[163] Ryan J Bloom, Sally M Winkler, and Christina D Smolke. A quantitative framework for the forward design of synthetic miRNA circuits. *Nature Methods*, 11(11):1147–53, 2014.

[164] T. Shimakami, D. Yamane, R. K. Jangra, B. J. Kempf, C. Spaniel, D. J. Barton, and S. M. Lemon. Stabilization of hepatitis C virus RNA by an Ago2-miR-122 complex. *Proceedings of the National Academy of Sciences*, 109(3):941–946, 2012.

[165] L Ghoda, T van Daalen Wetters, M Macrae, D Ascherman, and P Coffino. Prevention of rapid intracellular degradation of ODC by a carboxyl-terminal truncation. *Science*, 243(4897):1493–1495, 1989.

[166] Scotr Rogers, Rodney Wells, and Martin Rechsteiner. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, 234(4774):364–368, 1986.

[167] Xianqiang Li, Xiaoning Zhao, Yu Fang, Tommy Duong, Connie Fan, Chiao-Chain Huang, Steven R Kain, and Xin Jiang. Generation of destabilized green fluorescent protein as a transcription reporter. *The Journal of Biological Chemistry*, 273(52):34970–34975, 1998.

[168] Michael Kertesz, Nicola Iovino, Ulrich Unnerstall, Ulrike Gaul, and Eran Segal. The role of site accessibility in microRNA target recognition. TL - 39. *Nature Genetics*, 39 VN - r(10):1278–1284, 2007.

[169] Kei Endo, James A. Stapleton, Karin Hayashi, Hirohide Saito, and Tan Inoue. Quantitative and simultaneous translational control of distinct mammalian mRNAs. *Nucleic Acids Research*, 41(13), 2013.

[170] Rachel E Haurwitz, Samuel H Sternberg, and Jennifer A Doudna. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *The EMBO journal*, 31(12):2824–32, 2012.

[171] Laura a Banaszynski, Ling-Chun Chen, Lystranne a Maynard-Smith, a G Lisa Ooi, and Thomas J Wandless. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*, 126(5):995–1004, 2006.

[172] Maung Nyan Win and Christina D Smolke. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences*, 104(36):14283–14288, 2007.

[173] S Auslander, P Ketzer, and J S Hartig. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Molecular bioSystems*, 6(5):807–814, 2010.

[174] Yoko Nomura, Linlin Zhou, Anh Miu, and Yohei Yokobayashi. Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes. *ACS Synthetic Biology*, 2(12):684–689, 2013.

[175] Kim Beilstein, Alexander Wittmann, Manuel Grez, and Beatrix Suess. Conditional control of mammalian gene expression by tetracycline- dependent hammerhead ribozymes. *ACS synthetic biology*, 4(5):526–534, 2014.

[176] Nicolas Piganeau. *Selecting Allosteric Ribozymes*, pages 317–328. Humana Press, Totowa, NJ, 2012.

[177] N Piganeau, V Thuillier, and M Famulok. In vitro selection of allosteric ribozymes: theory and experimental validation. *Journal of Molecular Biology*, 312(5):1177–90, 2001.

[178] P Corish and C Tyler-Smith. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein engineering*, 12(12):1035–1040, 1999.

[179] Jim Haseloff and Wayne L. Gerlach. Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus. In Marlene Belfort and David A Shub, editors, *Gene*, volume 82, pages 43–52. Elsevier, Amsterdam, 1989.

[180] Caifu Chen, Dana A Ridzon, Adam J Broomer, Zhaohui Zhou, Danny H Lee, Julie T Nguyen, Maura Barbisin, Nan Lan Xu, Vikram R Mahuvakar, Mark R Andersen, Kai Qin Lao, Kenneth J Livak, Karl J Guegler, Applied Biosystems, Lincoln Centre Drive, and Foster City. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33(20):1–9, 2005.

[181] Ya'ara Ben-Ari, Yehuda Brody, Noa Kinor, Amir Mor, Toshiro Tsukamoto, David L Spector, Robert H Singer, and Yaron Shav-Tal. The life of an mRNA in space and time. *Journal of Cell Science*, 123(10):1761–1774, 2010.

[182] Jeremy E Wilusz, Courtney K Jnbaptiste, Laura Y Lu, William F Marzluff, Claus-d Kuhn, Leemor Joshua-tor, and Phillip a Sharp. A triple helix stabilizes the 3 9 ends of long noncoding RNAs that lack poly ( A ) tails. *Genes & Development*, 26(21):2392–2407, 2012.

[183] D R Gallie, N J Lewis, and W F Marzluff. The histone 3'-terminal stem-loop is necessary for translation in Chinese hamster ovary cells. *Nucleic Acids Research*, 24(10):1954–1962, 1996.

[184] Shungo Kobori, Yoko Nomura, Anh Miu, and Yohei Yokobayashi. High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Research*, 43(13):e85, 2015.

[185] Pablo Landgraf, Mirabela Rusu, Robert Sheridan, Alain Sewer, Nicola Iovino, Alexei Aravin, Sébastien Pfeffer, Amanda Rice, Alice O Kamphorst, Markus Landthaler, Carolina Lin, Nicholas D Socci, Leandro Hermida, Valerio Fulci, Sabina Chiaretti, Robin Foà, Julia Schliwka, Uta Fuchs, Astrid Novosel, Roman-Ulrich Müller, Bernhard Schermer, Ute Bissels, Jason Inman, Quang Phan, Minchen Chien, David B Weir, Ruchi Choksi, Gabriella De Vita, Daniela Frezzetti, Hans-Ingo Trompeter, Veit Hornung, Grace Teng, Gunther Hartmann, Miklos Palkovits, Roberto Di Lauro, Peter Wernet, Giuseppe Macino, Charles E Rogler, James W Nagle, Jingyue Ju, F Nina Papavasiliou, Thomas Benzing, Peter Lichter, Wayne Tam, Michael J Brownstein, Andreas Bosio, Arndt Borkhardt, James J Russo, Chris Sander, Mihaela Zavolan, and Thomas Tuschl. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell*, 129(7):1401–1414, 2007.

[186] Bruno R Steinkraus, Markus Toegel, and Tudor A Fulga. Tiny giants of gene regulation: experimental strategies for microRNA functional studies. *Wiley Interdisciplinary Reviews: Developmental Biology*, 5(3):311–362, 2016.

[187] Teemu Kivioja, Anna Vähärautio, Kasper Karlsson, Martin Bonke, Martin Enge, Sten Linnarsson, and Jussi Taipale. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*, 9(1):72–74, 2011.

[188] Katsuyuki Shiroguchi, Tony Z Jia, Peter A Sims, and X Sunney Xie. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized

single-molecule barcodes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(4):1347–52, 2012.

[189] Saiful Islam, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg, and Sten Linnarsson. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(1):163–166, 2014.

[190] Glenn K. Fu, Julie Wilhelmy, David Stern, H. Christina Fan, and Stephen P A Fodor. Digital encoding of cellular mRNAs enabling precise and absolute gene expression measurement by single-molecule counting. *Analytical Chemistry*, 86(6):2867–2870, 2014.

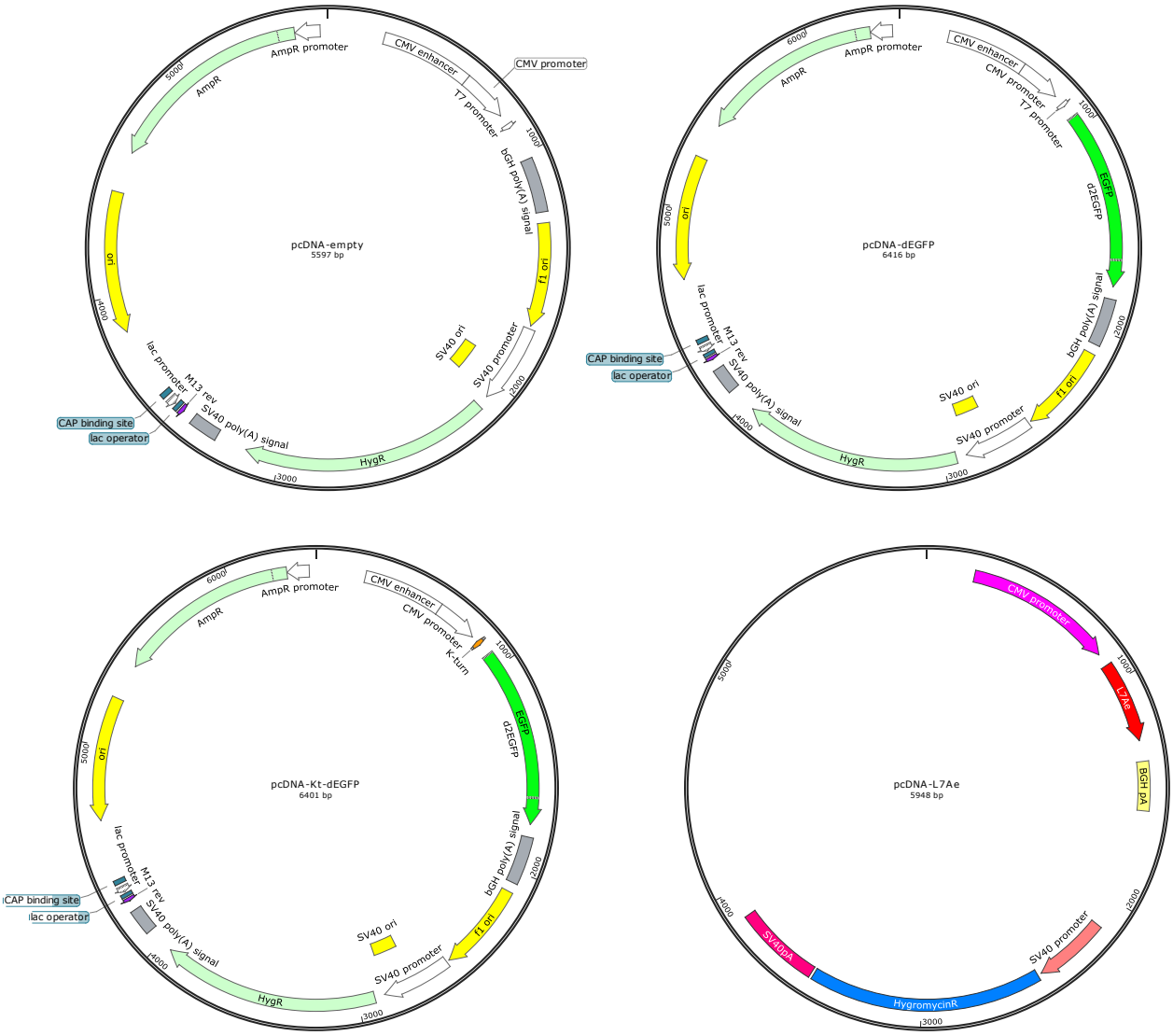
[191] Dang Long, Rosalind Lee, Peter Williams, Chi Yu Chan, Victor Ambros, and Ye Ding. Potent effect of target structure on microRNA function. *Nature Structural & Molecular Biology*, 14(APRIL):287–294, 2007.

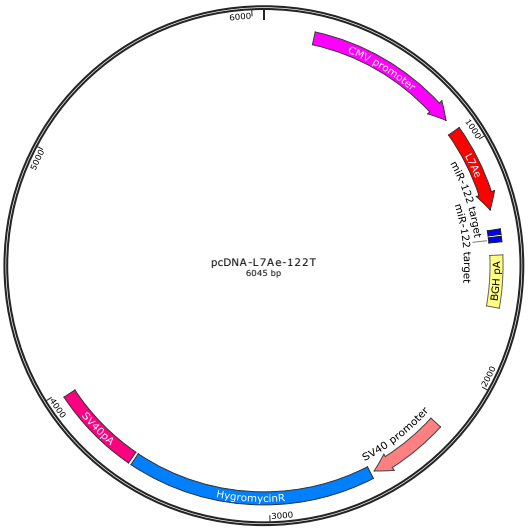
[192] Sung Wook Chi, Julie B Zang, Aldo Mele, and Robert B Darnell. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 460(7254):479–486, 2009.

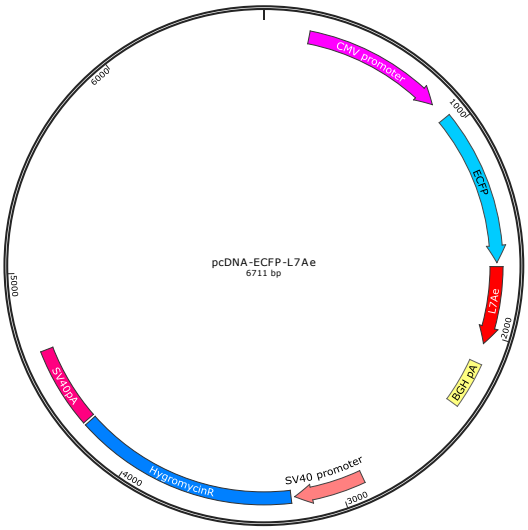
[193] Danish Sayed and Maha Abdellatif. MicroRNAs in Development and Disease. *Physiological Reviews*, 91(3):827 LP – 887, 2011.

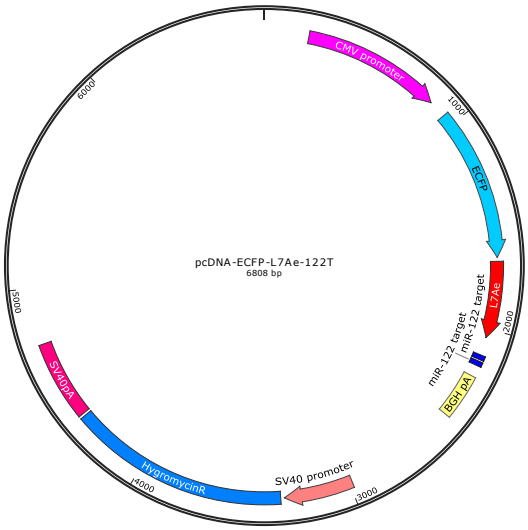
# Appendix A

# Plasmid maps

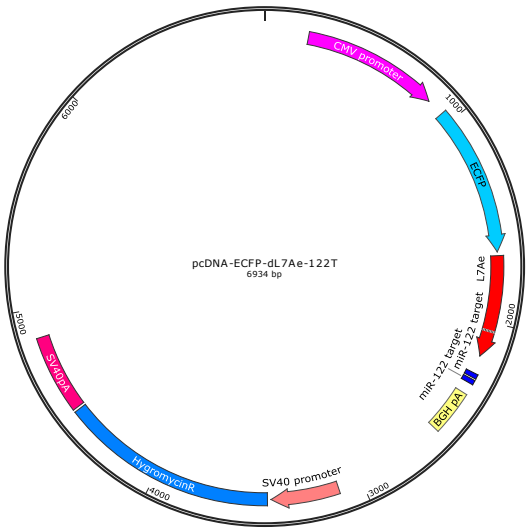


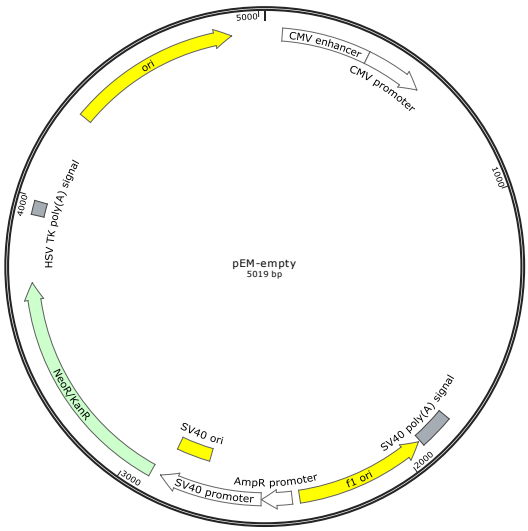

pcDNA-L7Ae-122T
6045 bp
CMV promoter
L7Ae
miR-122 target
miR-122 target
BGH pA
SV40 promoter
HygromycinR
SV40pA

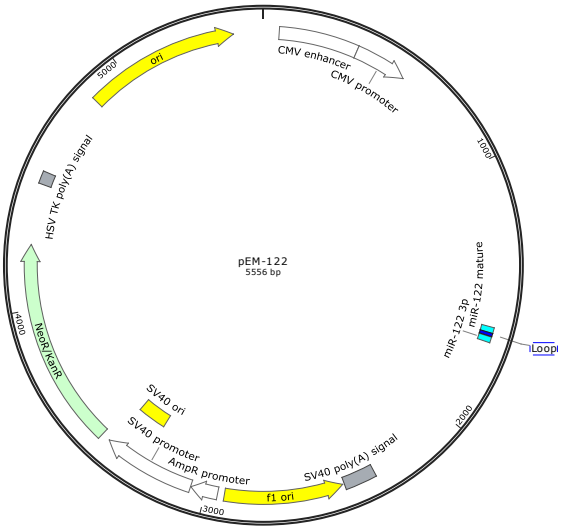

pcDNA-ECFP-L7Ae
6711 bp
CMV promoter
ECFP
L7Ae
BGH pA
SV40 promoter
HygromycinR
SV40pA

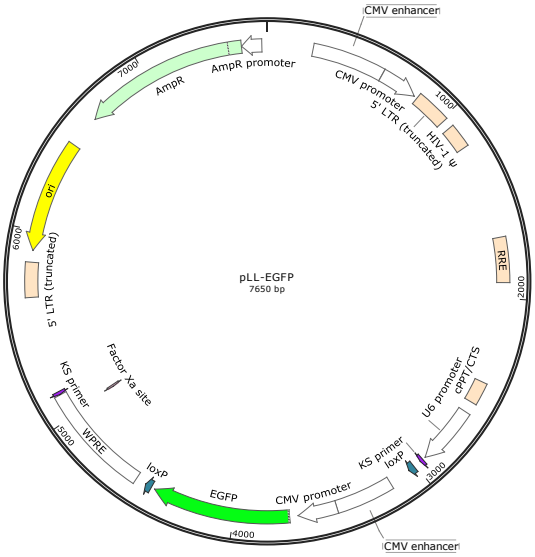

pcDNA-ECFP-L7Ae-122T
6808 bp
CMV promoter
ECFP
L7Ae
miR-122 target
miR-122 target
BGH pA
SV40 promoter
HygromycinR
SV40pA

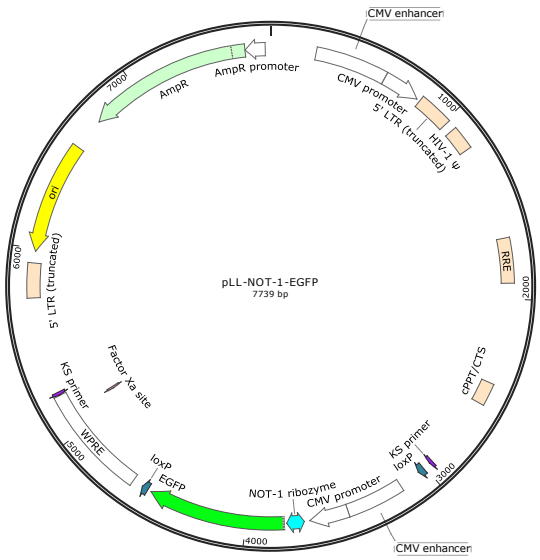

pcDNA-ECFP-dL7Ae
6837 bp
CMV promoter
ECFP
L7Ae
BGH pA
SV40 promoter
HygromycinR
SV40pA

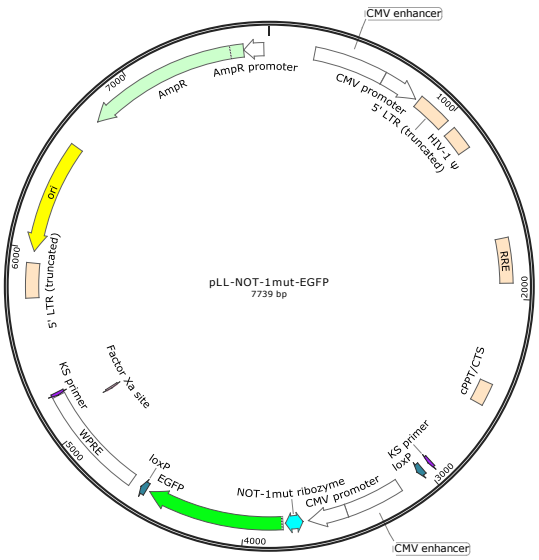

pcDNA-ECFP-dL7Ae-122T
6934 bp
CMV promoter
ECFP
L7Ae
miR-122 target
miR-122 target
BGH pA
SV40 promoter
HygromycinR
SV40pA

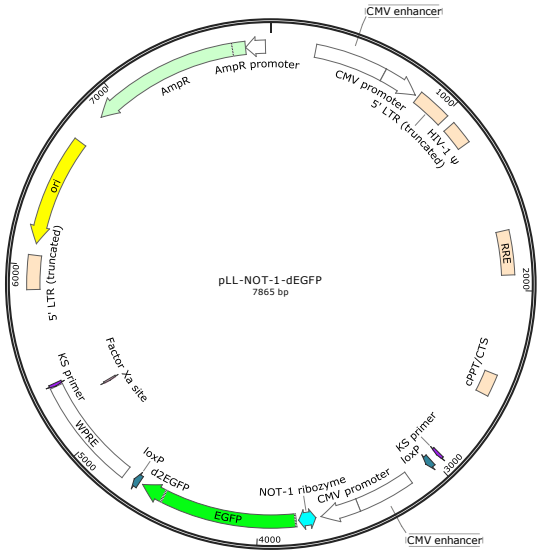

pEM-empty
5019 bp
ori
CMV enhancer
CMV promoter
HSV Tk poly(A) signal
NeoR/KanR
SV40 ori
SV40 poly(A) signal
AmpR promoter
f1 ori
SV40 promoter

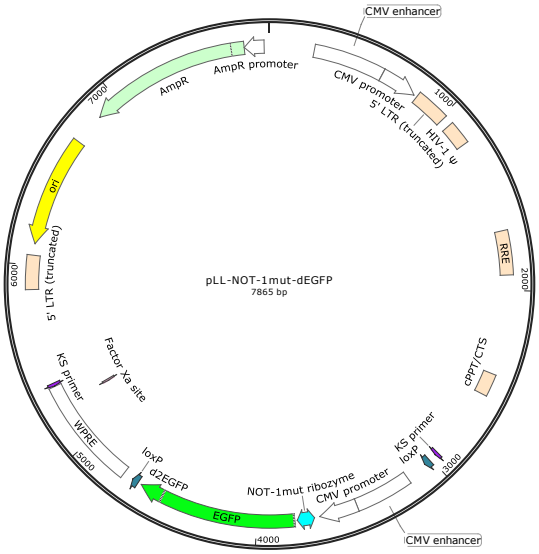

pEM-122
5556 bp
ori
CMV enhancer
CMV promoter
HSV TK poly(A) signal
miR-122 3p
miR-122 mature
Loop
NeoR/KanR
SV40 ori
SV40 promoter
AmpR promoter
SV40 poly(A) signal
f1 ori

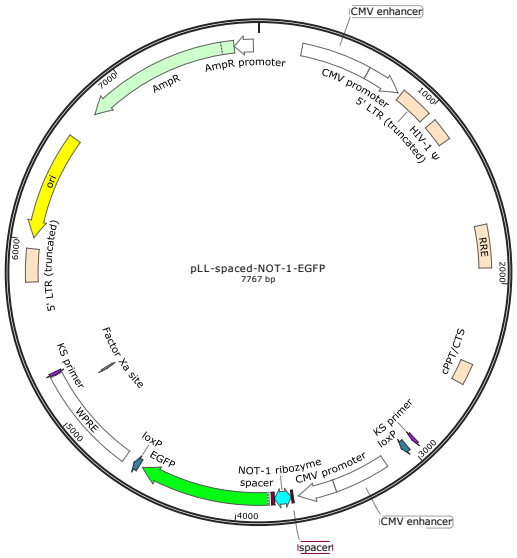

pLL-EGFP
7650 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
U6 promoter
KS primer
loxP
Factor Xa site
WPRE
EGFP
ori

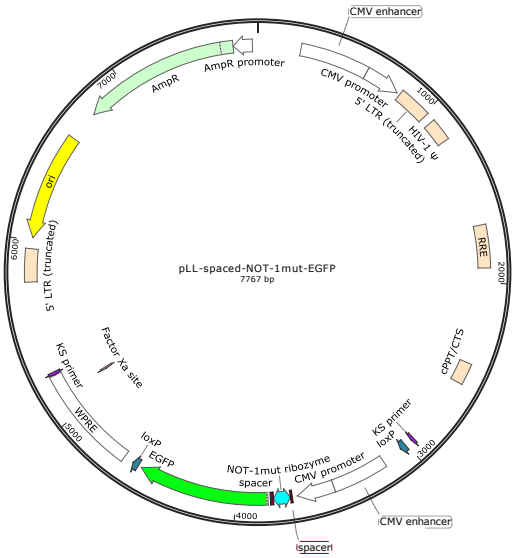

pLL-NOT-1-EGFP
7739 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
KS primer
loxP
Factor Xa site
WPRE
EGFP
NOT-1 ribozyme
ori

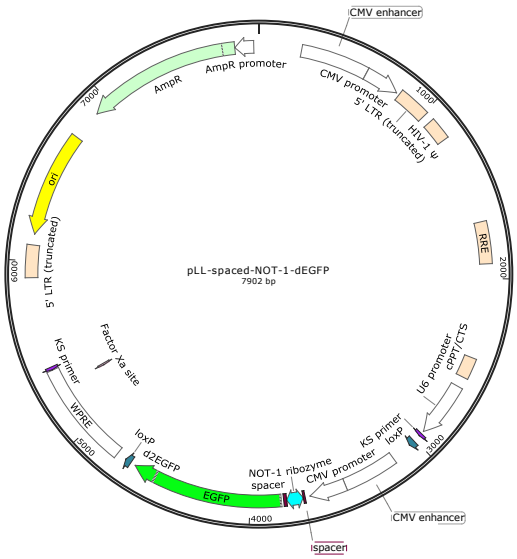

pLL-NOT-1mut-EGFP
7739 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
KS primer
loxP
Factor Xa site
WPRE
EGFP
NOT-1mut ribozyme
ori

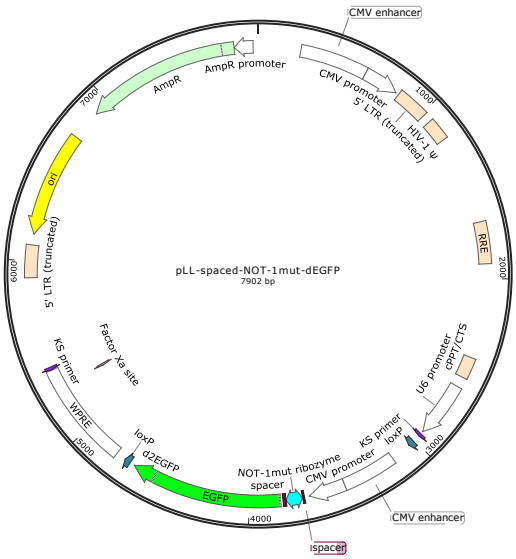

pLL-NOT-1-dEGFP
7865 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
KS primer
loxP
Factor Xa site
WPRE
d2EGFP
EGFP
NOT-1 ribozyme
ori

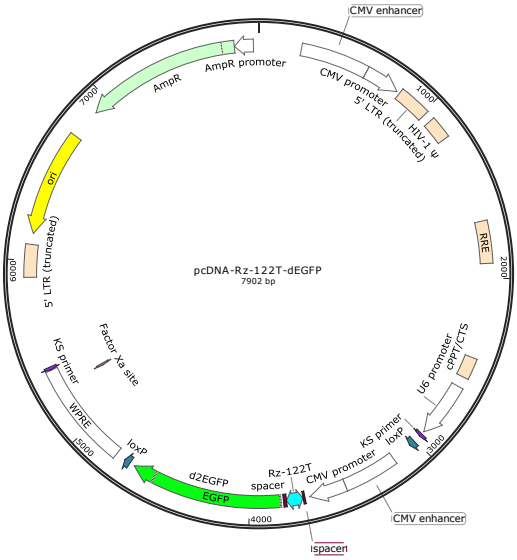

pLL-NOT-1mut-dEGFP
7865 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
KS primer
loxP
Factor Xa site
WPRE
d2EGFP
EGFP
NOT-1mut ribozyme
ori

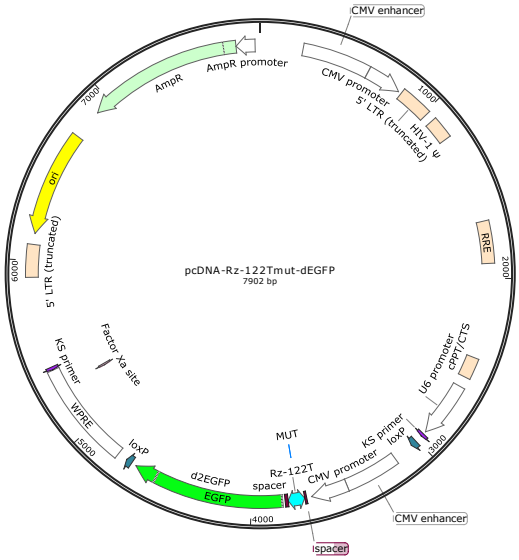

pLL-spaced-NOT-1-EGFP
7767 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
KS primer
Factor Xa site
WPRE
loxP
EGFP
NOT-1 ribozyme
spacer
CMV promoter
KS primer
loxP
CMV enhancer
ori
5' LTR (truncated)
spacer

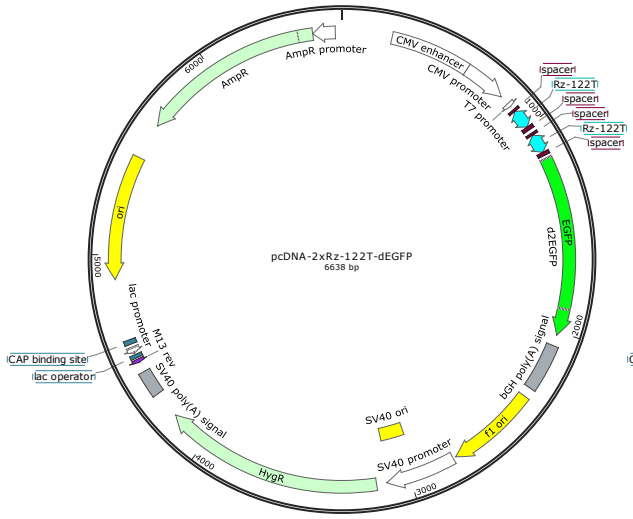

pLL-spaced-NOT-1mut-EGFP
7767 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
cPPT/CTS
KS primer
Factor Xa site
WPRE
loxP
EGFP
NOT-1mut ribozyme
spacer
CMV promoter
KS primer
loxP
CMV enhancer
ori
5' LTR (truncated)
spacer

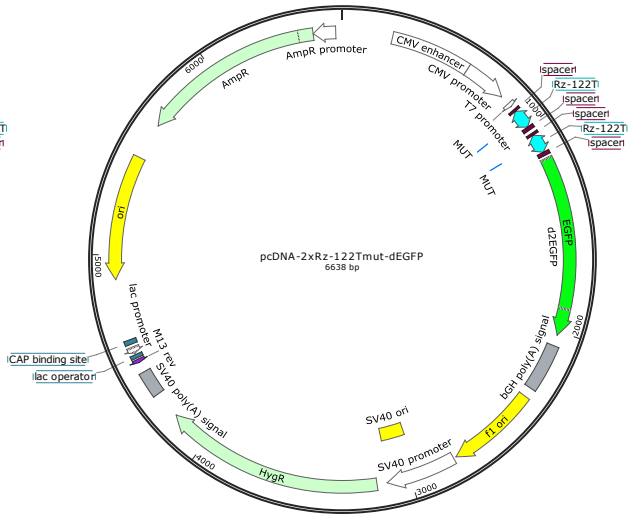

pLL-spaced-NOT-1-dEGFP
7902 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
U6 promoter
cPPT/CTS
KS primer
Factor Xa site
WPRE
loxP
d2EGFP
EGFP
NOT-1 ribozyme
spacer
CMV promoter
KS primer
loxP
CMV enhancer
ori
5' LTR (truncated)
spacer

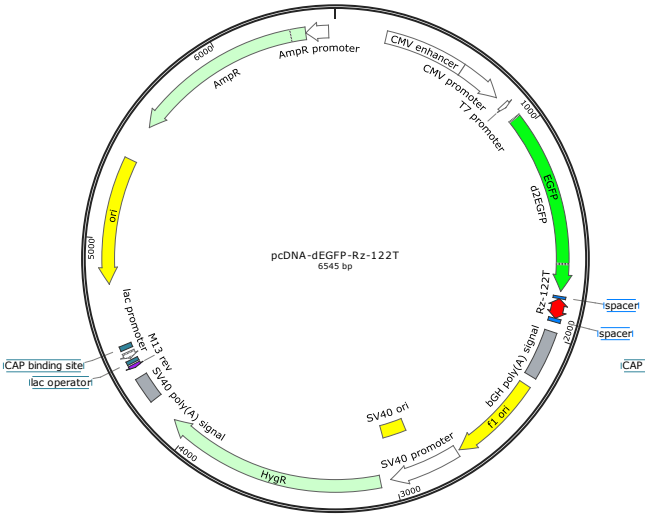

pLL-spaced-NOT-1mut-dEGFP
7902 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
U6 promoter
cPPT/CTS
KS primer
Factor Xa site
WPRE
loxP
d2EGFP
EGFP
NOT-1mut ribozyme
spacer
CMV promoter
KS primer
loxP
CMV enhancer
ori
5' LTR (truncated)
spacer

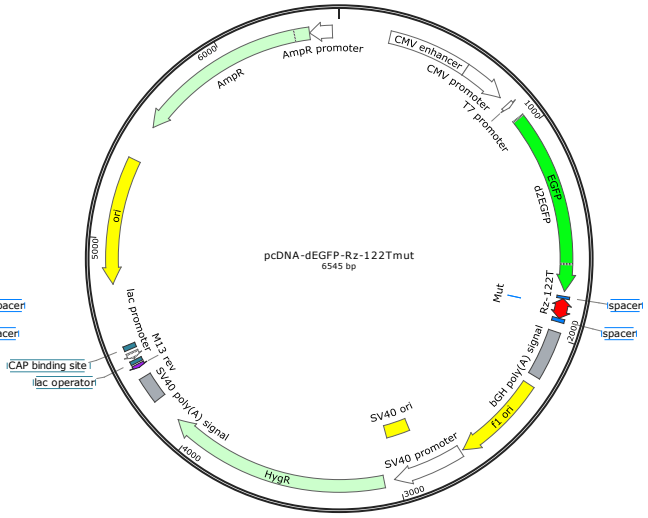

pcDNA-Rz-122T-dEGFP
7902 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
U6 promoter
cPPT/CTS
KS primer
Factor Xa site
WPRE
loxP
d2EGFP
EGFP
Rz-122T
spacer
CMV promoter
KS primer
loxP
CMV enhancer
ori
5' LTR (truncated)
spacer

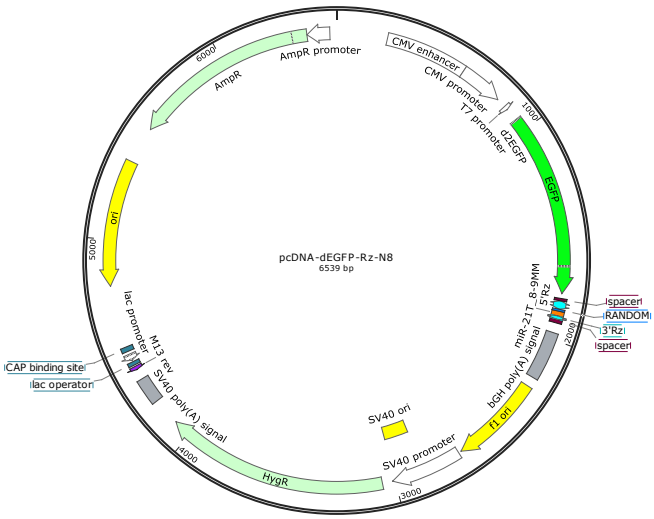

pcDNA-Rz-122Tmut-dEGFP
7902 bp
CMV enhancer
AmpR promoter
AmpR
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
U6 promoter
cPPT/CTS
KS primer
Factor Xa site
WPRE
loxP
MUT
d2EGFP
EGFP
Rz-122T
spacer
CMV promoter
KS primer
loxP
CMV enhancer
ori
5' LTR (truncated)
spacer

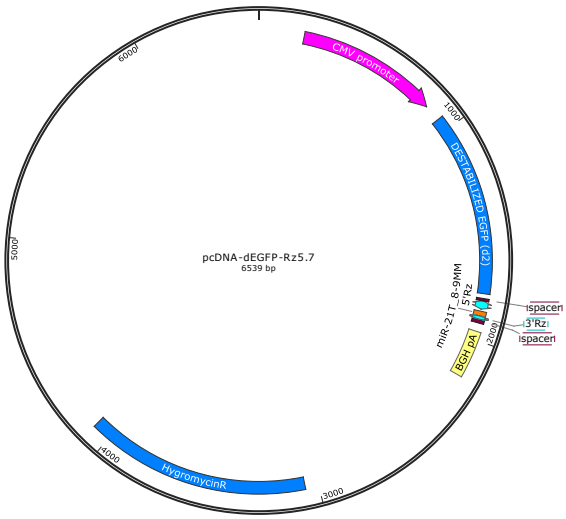

pcDNA-2xRz-122T-dEGFP
6638 bp
AmpR promoter
AmpR
CMV enhancer
CMV promoter
T7 promoter
spacer
Rz-122T
spacer
spacer
Rz-122T
spacer
EGFP
d2EGFP
bGH poly(A) signal
f1 ori
SV40 promoter
SV40 ori
HygR
SV40 poly(A) signal
M13 rev
lac promoter
CAP binding site
lac operator
ori
6000
1000
2000
3000
4000
5000

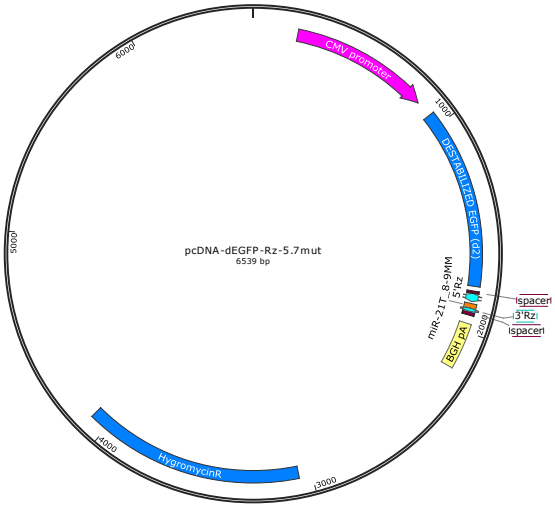

pcDNA-2xRz-122Tmut-dEGFP
6638 bp
AmpR promoter
AmpR
CMV enhancer
CMV promoter
T7 promoter
MUT
MUT
spacer
Rz-122T
spacer
spacer
Rz-122T
spacer
EGFP
d2EGFP
bGH poly(A) signal
f1 ori
SV40 promoter
SV40 ori
HygR
SV40 poly(A) signal
M13 rev
lac promoter
CAP binding site
lac operator
ori
6000
1000
2000
3000
4000
5000

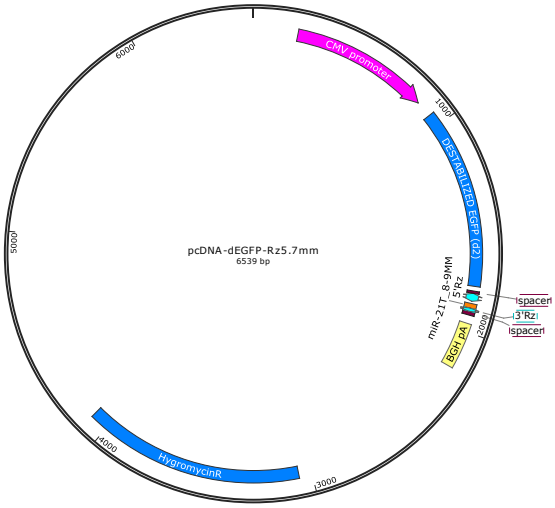

pcDNA-dEGFP-Rz-122T
6545 bp
AmpR promoter
AmpR
CMV enhancer
CMV promoter
T7 promoter
EGFP
d2EGFP
Rz-122T
spacer
spacer
bGH poly(A) signal
f1 ori
SV40 promoter
SV40 ori
HygR
SV40 poly(A) signal
M13 rev
lac promoter
CAP binding site
lac operator
ori
6000
1000
2000
3000
4000
5000

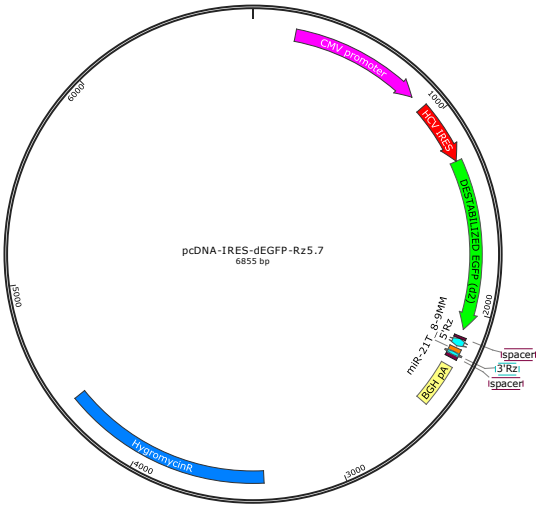

pcDNA-dEGFP-Rz-122Tmut
6545 bp
AmpR promoter
AmpR
CMV enhancer
CMV promoter
T7 promoter
EGFP
d2EGFP
Mut
Rz-122T
spacer
spacer
bGH poly(A) signal
f1 ori
SV40 promoter
SV40 ori
HygR
SV40 poly(A) signal
M13 rev
lac promoter
CAP binding site
lac operator
ori
6000
1000
2000
3000
4000
5000

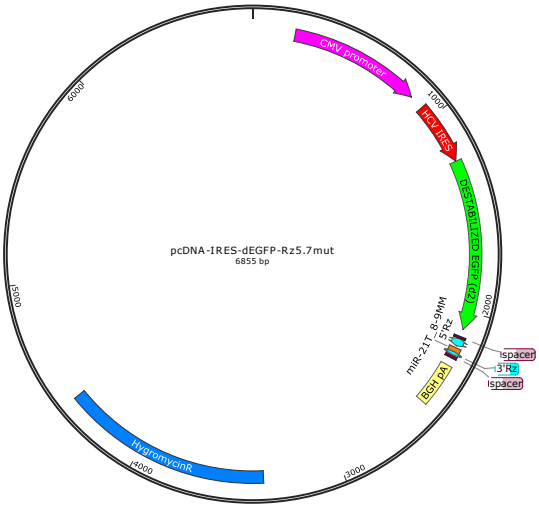

pcDNA-dEGFP-Rz-N8
6539 bp
AmpR promoter
AmpR
CMV enhancer
CMV promoter
T7 promoter
d2EGFP
EGFP
miR-21T_8-9MM
5'Rz
spacer
RANDOM
3'Rz
spacer
bGH poly(A) signal
f1 ori
SV40 promoter
SV40 ori
HygR
SV40 poly(A) signal
M13 rev
lac promoter
CAP binding site
lac operator
ori
6000
1000
2000
3000
4000
5000

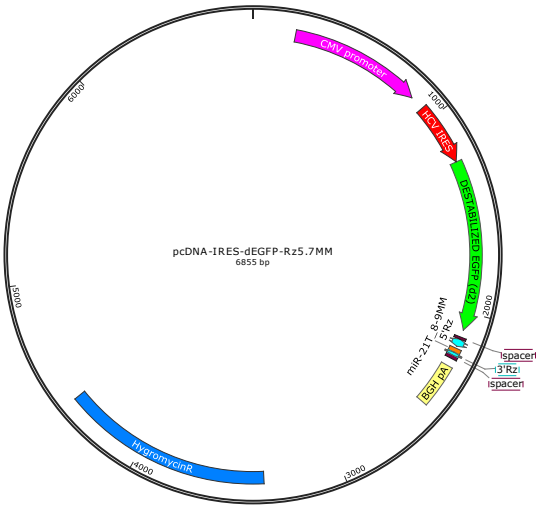

pcDNA-dEGFP-Rz5.7
6539 bp
CMV promoter
DES(NB)(2xRz) EGFP (d2)
miR-21T_8-9MM
5'Rz
spacer
3'Rz
spacer
BGH pA
HygromycinR
6000
1000
2000
3000
4000
5000

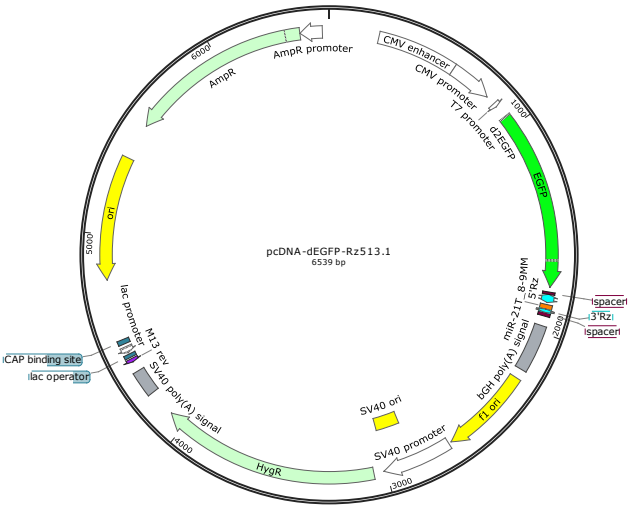

pcDNA-dEGFP-Rz-5.7mut
6539 bp
CMV promoter
d2EGFP
miR-21T-8-9MM
5'Rz
BGH pA
spacer
3'Rz
spacer
HygromycinR

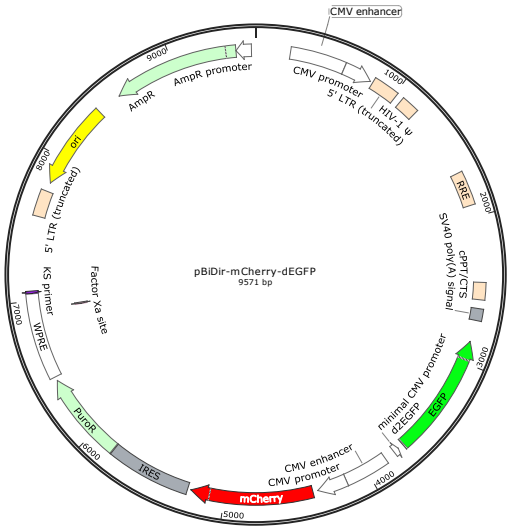

pcDNA-dEGFP-Rz5.7mm
6539 bp
CMV promoter
d2EGFP
miR-21T-8-9MM
5'Rz
BGH pA
spacer
3'Rz
spacer
HygromycinR

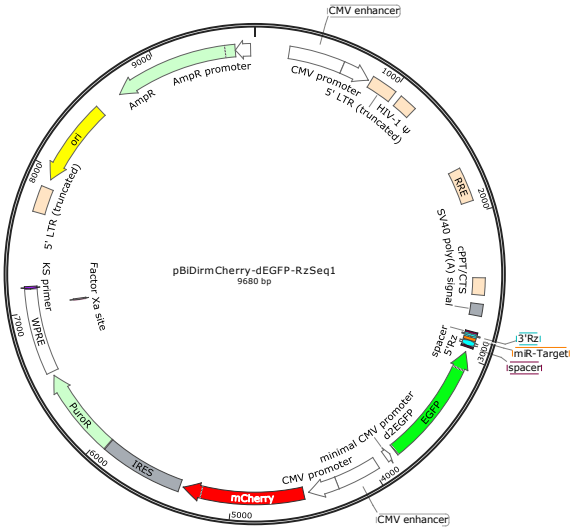

pcDNA-IRES-dEGFP-Rz5.7
6855 bp
CMV promoter
HCV IRES
d2EGFP
miR-21T-8-9MM
5'Rz
BGH pA
spacer
3'Rz
spacer
HygromycinR

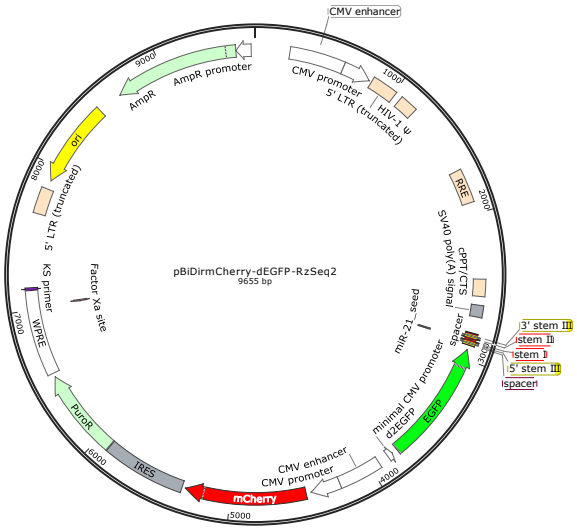

pcDNA-IRES-dEGFP-Rz5.7mut
6855 bp
CMV promoter
HCV IRES
d2EGFP
miR-21T-8-9MM
5'Rz
BGH pA
spacer
3'Rz
spacer
HygromycinR

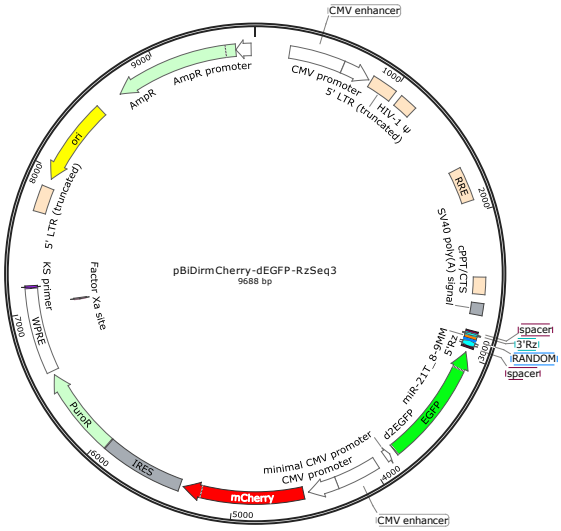

pcDNA-IRES-dEGFP-Rz5.7MM
6855 bp
CMV promoter
HCV IRES
d2EGFP
miR-21T-8-9MM
5'Rz
BGH pA
spacer
3'Rz
spacer
HygromycinR

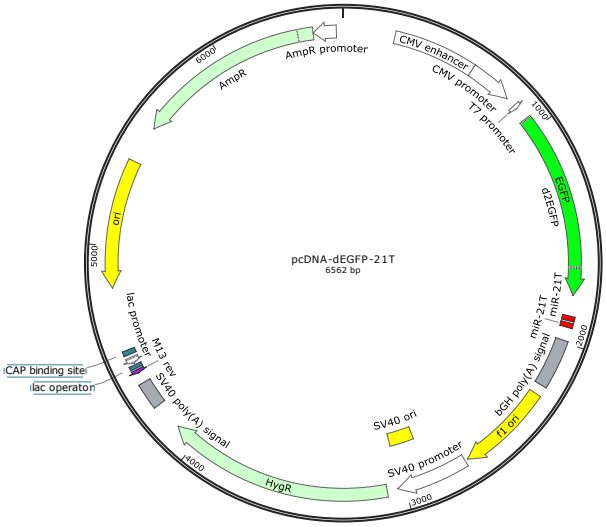

pcDNA-dEGFP-Rz513.1
6539 bp
AmpR promoter
AmpR
CMV enhancer
CMV promoter
T7 promoter
d2EGFP
EGFP
ori
lac promoter
M13 rev
CAP binding site
lac operator
SV40 poly(A) signal
miR-21T-8-9MM
5'Rz
spacer
3'Rz
spacer
bGH poly(A) signal
f1 ori
SV40 ori
SV40 promoter
HygR


pBiDir-mCherry-dEGFP
9571 bp
CMV enhancer
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
SV40 poly(A) signal
cPPT/CTS
minimal CMV promoter
d2EGFP
EGFP
CMV enhancer
CMV promoter
mCherry
IRES
PuroR
WPRE
KS primer
Factor Xa site
5' LTR (truncated)
ori
AmpR
AmpR promoter


pBiDirmCherry-dEGFP-RzSeq1
9680 bp
CMV enhancer
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
SV40 poly(A) signal
cPPT/CTS
spacer
5'Rz
3'Rz
miR-Target
spacer
minimal CMV promoter
CMV promoter
d2EGFP
EGFP
CMV enhancer
mCherry
IRES
PuroR
WPRE
KS primer
Factor Xa site
5' LTR (truncated)
ori
AmpR
AmpR promoter


pBiDirmCherry-dEGFP-RzSeq2
9655 bp
CMV enhancer
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
SV40 poly(A) signal
cPPT/CTS
spacer
miR-21 seed
3' stem II
stem II
stem I
5' stem II
spacer
minimal CMV promoter
d2EGFP
EGFP
CMV enhancer
CMV promoter
mCherry
IRES
PuroR
WPRE
KS primer
Factor Xa site
5' LTR (truncated)
ori
AmpR
AmpR promoter


pBiDirmCherry-dEGFP-RzSeq3
9688 bp
CMV enhancer
CMV promoter
5' LTR (truncated)
HIV-1 ψ
RRE
SV40 poly(A) signal
cPPT/CTS
spacer
miR-21T_8-9MM
5'Rz
spacer
3'Rz
RANDOM
spacer
minimal CMV promoter
CMV promoter
d2EGFP
EGFP
CMV enhancer
mCherry
IRES
PuroR
WPRE
KS primer
Factor Xa site
5' LTR (truncated)
ori
AmpR
AmpR promoter


pcDNA-dEGFP-21T
6562 bp
AmpR promoter
AmpR
ori
lac promoter
M13 rev
CAP binding site
lac operator
SV40 poly(A) signal
HygR
SV40 promoter
SV40 ori
f1 ori
bGH poly(A) signal
miR-21T
miR-21T
CMV enhancer
CMV promoter
T7 promoter
EGFP
d2EGFP