

Deciphering *cis/trans* RNA regulatory interactions *in vivo*



Bruno R Steinkraus
St Edmund Hall
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2016

Deciphering *cis/trans* RNA regulatory interactions *in vivo*

Bruno R Steinkraus, St Edmund Hall, DPhil Clinical Laboratory Sciences,
Michaelmas 2016

Abstract

Post-transcriptional regulation of the genetic programme is achieved through the sophisticated interplay between *cis* regulatory elements and *trans* interacting factors. Deciphering these interactions, especially those between microRNAs (miRNAs) and destabilising miRNA response elements (MREs), requires technologies that permit their unbiased interrogation under native, physiological conditions. Here I describe the development of miR-TRACE, a novel and versatile platform for miRNA-target identification that analyses *in vivo* miRNA-MRE binding. Following iterative optimisation I demonstrate the reliability of this technology in identifying *bona fide* miRNA targets on a candidate scale. Subsequently, I implement several experimental and computational innovations to develop miR-TRACEseq, a multiplex platform enabling unbiased high-throughput analysis of miRNA-MRE interactions across the transcriptome. As a proof of concept, I applied this technology to a highly conserved candidate miRNA in S2R+ cells (miR-184) and implemented a logistic regression model to map all miR-184 MREs. This analysis confirmed known miR-184 MREs but also discovered novel targets.

In parallel, during my thesis I have applied the core principles underlying miR-TRACE to study the translational quiescence of *bicoid* (*bcd*) mRNA during *Drosophila* oogenesis. Several lines of evidence, foremost *bcd*'s association to mRNA processing bodies (P-bodies), suggest the possibility of a miRNA-mediated mechanism underlying *bcd* silencing. To explore this scenario I have developed a multipronged experimental approach employing miR-TRACE, unbiased small RNA interactome capture, and computational analysis. Together, these analyses uncovered several miRNAs specifically associated with *bcd* RNA in ovary tissues, of which the two most notable candidates were miR-184 and miR-989. Interestingly, the levels of these miRNAs are significantly reduced in a *bcd* pulldown from embryos, a finding consistent with relief of *bcd* silencing in this tissue. This study provides the first evidence for a potential role of miRNAs in regulating the translation of *bcd* RNA. This opens up a completely new line of investigation in the quest for elucidating what remains one of the most enigmatic accounts of gene expression regulation during the oocyte-to-embryo developmental transition.

Declaration

I hereby declare that the work presented in this Thesis is my own, and the experiments detailed throughout this Thesis were performed by myself, unless stated otherwise. The experiments were conducted in the Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DS, UK. This Thesis has not been submitted for any other degree or professional qualification.

The following parts of this Thesis were completed by others:

- 1.) The miR-184 KO cell line was a gift from Andrew Bassett, Cambridge.
- 2.) Cloning of the miR-184 genomic rescue fragment and overexpression construct was performed by Yale Michaels, Fulga lab, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DS, UK. This relates to the following figures 3.5 and 3.7.
- 3.) The miR-TRACE training data (CRISPR-MRE and paired miR-184/transcriptome bins) was generated by Qianxin Wu and Quentin Ferry and only modified by myself. QW and QF are both in the Fulga Lab, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DS, UK. This relates to figure 3.8.

The following parts of this Thesis were carried out in collaboration with the following people:

- 1.) The miR-TRACE ASO group selection algorithm in chapter 4 has been coded by Quentin Ferry, Fulga lab. This relates to figure 4.3.

- 2.) All sequencing and DASH relating to miR-TRACE have been performed with the help of Qianxin Wu, Fulga lab. This relates to figures 4.12-4.16.
- 3.) The logistic regression model (miR-TRACEseq) was written by David Knapp, Fulga lab. This relates to figures 2.1, 4.17-4.19.
- 4.) Mapping of the small RNA data was performed by Stephen Taylor, CBRG, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DS, UK. This relates to figure 5.1, 5.12-5.13.

Bruno R Steinkraus

Acknowledgements

I would like to thank my supervisor, Tudor Fulga, who took me aboard his lab, funded me and taught me the scientific approach to problem solving. I am beyond grateful to Qianxin Wu for her relentless intellectual input and advice, and owe a debt of gratitude to other members of the Fulga lab: Yale Michaels for scientific and life wisdom; Markus Toegel and Aron Szabo for experimental support and continuous discussions of failed experiments; Quentin Ferry for his coding skills and generating the grouping algorithm; Toni Bäumlér for cloning and FloC analysis and driving the OneStart team; Tim Rajakumar for getting me started in the lab; and finally, David Knapp for help with the data analysis of the miR-TRACEseq project.

My thanks also go to Tatjana Sauker-Spengler for teaching me the importance of being scrupulous but also sometimes 'just giving it a shot', and Tom Milne for sharp feedback at all times. The entire TSS and Milne labs were extremely helpful and supportive throughout my whole DPhil. I am obliged to Stephen Taylor, who performed the small RNA sequencing analyses, and other members of the CBRG. Amy Trebes and Paolo Piazza of the Wellcome Trust were so kind to let me use their Pippin size selector. I would also like to thank Timothy Weil and Richard Parton for help with fly work and advice on *bicoid*. I am grateful to Alfredo Castello and Marko Noerenberg for support with setting up interactome capture in the lab. Andrew Bassett provided tremendously useful experimental feedback and details to the mapping procedure and generated the miR-184 KO cell line. Diana Chin helped with the cloning of the reporter constructs. Furthermore, I would like to thank Ilan Davis and Vincenzo Cerundolo for being on my thesis committee. Finally, I am deeply indebted to and grateful for the unconditional support of my family and friends.

Publications

Steinkraus BR[#], Toegel M[#], Fulga TA. *Tiny giants of gene regulation: methods and tools for studying miRNA functions*. **WIREs Developmental Biology** 2016 May;5(3):311-62

Fulga TA, McNeill EM, Binari R, Yelick J, Blanche A, Booker M, **Steinkraus BR**, Schnall-Levin M, Zhao Y, DeLuca T, Bejarano F, Han Z, Lai EC, Wall DP, Perrimon N, Van Vactor D. *A transgenic resource for conditional competitive inhibition of conserved Drosophila microRNAs*. **Nature Communications**. 2015 Jun 17;6:7279

Tan JY, Vance KW, Varela MA, Sirey T, Watson LM, Curtis HJ, Marinello M, Alves S, **Steinkraus BR**, Cooper S, Nesterova T, Brockdorff N, Fulga TA, Brice A, Sittler A, Oliver PL, Wood MJ, Ponting CP, Marques AC. *Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7*. **Nature Structural Molecular Biology**. 2014 Nov;21(11):955-61

Under consideration

Wu Q, Ferry Q, Baeumler TA, Michaels Y, Vitsios D, Habib O, Arnold R, Jiang X, Tapia M, **Steinkraus BR**, Maio S, Piazza P, Xu N, Holländer GA, Milne TA, Kim J, Enright AJ, Bassett AR, Fulga TA. *In situ functional dissection of RNA cis-regulatory elements by multiplex genome engineering*. Under consideration.

Phillips TJ, Scott H, Sood A, Menassa DA, Bignell AL, Morton JS, Akagi T, Azuma K, Rogers MF, Inman GJ, Holmans P, **Steinkraus BR**, Collett GP, Kemp H, Murphy MP, Fulga TA, Akashi M, Davidge ST, Case CP. *Treating the placenta with nanoparticles to prevent altered placental secretion and fetal development in response to hypoxia*. Under consideration.

In preparation

Steinkraus BR, Knapp D, Wu Q, Fulga TA. *Transcriptome-wide mapping of RNA regulatory elements by TRACE*. In preparation.

List of Abbreviations

A	Adenosine
AGO	Argonaute
ASO	Antisense oligonucleotide
<i>bcd</i>	<i>bicoid</i>
BLE	<i>bicoid</i> localisation element
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
CDS	Coding sequence
CLIP	Crosslinking Immuno-Precipitation
DV	Dorsoventral
DASH	Depletion of abundant sequences by hybridisation
DMD	Duchenne Muscular Dystrophy
DNA	Deoxyribonucleic acid
EGFP	enhanced green fluorescent protein
Exu	Exuperantia
FDR	False discovery rate
FPKM	Fragments Per Kilobase Million
G	Guanosine
gDNA	genomic DNA
	genome-engineering based evaluation of RNA regulatory
genERA	elements activity
GLM	general linear model
<i>grk</i>	<i>gurken</i>
GSC	Germline stem cell
GSP	Gene specific primer

h	hour
hnRNP	Heterogenous nuclear ribonucleoprotein
	Individual-nucleotide resolution Crosslinking and Immuno-
iCLIP	Precipitation
IRES	Internal ribosome entry site
IVT	In Vitro Transcribed
KO	Knockout
lncRNA	long non-coding RNA
mCherry	Monomeric Cherry
MCP	MS2 coat protein
Me31B	Maternal expression at 31B
min	minute(s)
miR-TRACE	<i>in vivo</i> detection of miRNA-target interaction events
miRISC	miRNA Induced RNA Silencing Complex
miRNA	microRNA
ml	millilitre
mm	millimetre
mM	milliMol
MRE	miRNA Response Element
mRNA	Messenger Ribonucleic Acid
MZT	Maternal-to-zygotic transition
<i>nos</i>	<i>nanos</i>
OET	Oocyte-to-embryo transition
<i>orb</i>	<i>oo18 RNA binding protein</i>
ORF	Open reading frame
<i>osk</i>	<i>oskar</i>

P-bodies	Processing bodies
PABP	Poly(A) binding protein
PAP	Poly(A) polymerase
PBS	Phosphate Buffered Saline
PBST	PBS with Tween 20 (0.1%)
PCR	Polymerase Chain Reaction
piRNA	PIWI interacting RNA
Pum	Pumilio
RBP	RNA binding protein
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein
RPS	Relative Protection Score
rRNA	ribosomal RNA
RSS	RNA secondary structure
RT-PCR	Reverse Transcription quantitative PCR
s	second(s)
sgRNA	single guide RNA
siRNA	Short interfering RNA
Sqd	Squid
SRE	Smaug recognition element
Stau	Staufen
T	Thymidine
Tris	Tris(hydroxymethyl)aminomethane
Triton-X 100	Octylphenoxypolyethoxyehtanol
UTR	Untranslated region

UV	Ultraviolet
WT	Wildtype
μl	Microlitre
μm	Micrometre
ρ	Spearman's correlation
χ	chi square test

Table of Contents

Abstract	ii
Declaration	iii
Acknowledgments	v
Publications	vi
List of Abbreviations	vii
List of Figures	xix
List of Tables	xxi

Chapter 1: Introduction..... 1

1.1 Technology development for miRNA target identification 2

1.1.1 The non-coding genome as a key player in post-transcriptional regulation	2
--	---

1.1.2 miRNAs	4
--------------------	---

1.1.2.1 Biogenesis of miRNAs	6
------------------------------------	---

1.1.2.2 Mechanism of action: Effector steps of silencing.....	12
---	----

1.1.2.2.1 Translational inhibition post-initiation	15
--	----

1.1.2.2.2 Translational inhibition pre-initiation.....	15
--	----

1.1.2.2.3 Target degradation.....	18
-----------------------------------	----

1.1.2.2.4 Relative contribution of decay vs. translational inhibition and order of events	19
---	----

1.1.2.3 miRNA target identification strategies	20
--	----

1.1.2.3.1 Profiling Approaches.....	26
-------------------------------------	----

1.1.2.3.2 RIP and CLIP	31
------------------------------	----

1.1.2.3.3 RNA Capture.....	40
----------------------------	----

1.1.2.3.3.1 mRNA pull down.....	41
1.1.2.3.3.2 miRNA pull down.....	44
1.2 Elucidating the regulatory mechanisms underlying <i>bicoid</i> RNA	
translational quiescent state <i>in vivo</i>.....	47
1.2.1 Synopsis of <i>Drosophila</i> oogenesis and early embryogenesis.....	48
1.2.2 Maternal mRNA Localisation and Translational Repression.....	52
1.2.2.1 The localisation of <i>bicoid</i> mRNA.....	54
1.2.2.2 Translational regulation of <i>bicoid</i> mRNA	57
1.3 Thesis Aims.....	63
Chapter 2: Materials and Methods	66
2.1 Buffers & Solutions	66
2.2 <i>In vivo</i> detection of <u>miRNA-target interaction events</u> (TRACE)	67
2.2.1 Material (Lysate) Preparation.....	67
2.2.1.1 S2R+ cells	67
2.2.1.2 Ovary	68
2.2.1.3 Embryo	68
2.2.2 RNase H Mapping.....	69
2.2.2.1 solo ‘one-ASO:one-MRE’	69
2.2.2.2 grouped ‘one-ASO:one-MRE’	70
2.2.2.3 multiplex ‘one-ASO:many-MREs’	70
2.2.2.4 Formulas for RPS and TRACE Score	71
2.2.3 TRACEseq - Library preparation and MiSeq.....	71
2.2.4 DASH and NextSeq	72
2.2.4.1 sgRNA generation	72
2.2.4.2 DASH reaction, qPCR and NEXTseq.....	73
2.2.5 TRACEseq bioinformatics and GLM	74

2.2.5.1 NGS analysis.....	74
2.2.5.2 Logistic regression model.....	74
2.2.5.2.1 True positives.....	74
2.2.5.2.2 Logistic regression training	76
2.2.5.2.3 Assessment of GLM.....	77
2.3 <i>Bicoid</i> translational quiescence methods.....	77
2.3.1 Nanostring analysis.....	77
2.3.2 miRNA target prediction	77
2.3.3 piRNA analysis.....	78
2.3.4 Pulldown probe generation	78
2.3.5 Probe - Bead coupling.....	80
2.3.6 Material Generation.....	80
2.3.6.1 Ovary	80
2.3.6.2 Embryo	81
2.3.7 RNA Capture.....	81
2.3.7.1 Capture bead preparation.....	81
2.3.7.2 Lysate preparation.....	82
2.3.7.3 Pull Down	82
2.3.8 Library Preparation and NGS Analysis	83
2.4 General Molecular Methods.....	83
2.4.1 RNA Work	83
2.4.2 cDNA synthesis.....	85
2.4.3 qPCR.....	85
2.4.4 miRNA cDNA synthesis	85
2.4.5 Taqman miRNA qPCR.....	86
2.4.6 Digital droplet PCR (ddPCR).....	86

2.4.7 <i>In vitro</i> Transcription.....	87
2.4.8 Flow Cytometry	88
2.5 Cell Culture.....	89
2.5.1 Cell maintenance and passage.....	89
2.5.2 Transfections.....	89
2.6 Drosophila Methods	90
2.6.1 Fly husbandry.....	90
2.6.2 Ovary dissections.....	90
2.6.3 Embryo collections	90
 Chapter 3: <i>In vivo</i> Detection of miRNA-target Interaction Events	
(miR-TRACE) Decodes Functional miRNA Response Elements	91
3.1 Introduction.....	91
3.2 Chapter Aims	95
3.3 Results.....	95
3.3.1 RNase H protocol overview.....	95
3.3.2 Overview of miR-TRACE assay	97
3.3.3 Optimisation of RNase H miR-TRACE Mapping Protocol.....	99
3.3.4 Stepwise Analysis Pipeline Used To Calculate TRACE Score	101
3.3.5 Proof Of Concept – miR-TRACE reflects miRISC binding at <i>crok</i> MRE	103
3.3.6 Stepwise analysis pipeline for TRACE using <i>rp49</i> as normalisation factor	105
3.3.7 Generation of a training dataset.....	108
3.3.8 miR-TRACE analysis closely reflects MRE score	109
3.4 Discussion.....	113
3.4.1 Optimisation of RNase H cleavage conditions	113

3.4.2 TRACE reflects previously calculated MRE Scores.....	114
3.4.3 TRACE Score as a new metric for miRNA::target interactions	114
3.4.4 Technical considerations for miR-TRACE.....	116
3.4.5 Comparison of miR-TRACE to other miRNA target MRE identification and validation techniques.....	118
3.4.6 Final Conclusions.....	119

Chapter 4: Design and Implementation of high-throughput <i>in vivo</i> detection of miRNA-Target Interaction Events by Massively Parallel Sequencing (miR-TRACEseq).....	120
4.1 Introduction.....	120
4.2 Chapter Aims	125
4.3 Results.....	126
4.3.1 Creating ASO groups streamlines TRACE	126
4.3.2 TRACE on ASO groups maintains TRACE Scores observed in individual assays	128
4.3.3 Short ASOs sustain efficient RNase H-mediated cleavage at MRE sites.....	132
4.3.4 Short ASOs effectively uncover differences in RPS between WT and KO	134
4.3.5 Analysis of 12 target MREs by multiplex TRACE.....	135
4.3.6 Development of a high-throughput readout for multiplex TRACE analysis	140
4.3.7 Normalisation by an exogenously delivered mRNA.....	143
4.3.8 The miR-184 6mer seed motif is significantly enriched in TRACE libraries	145
4.3.9 Depth of sequencing limits the ability to calculate meaningful	

RPS values.	147
4.3.10 DASH removes contaminating RNA	151
4.3.11 Post-DASH sequencing on NextSeq delivers high number of target reads	153
4.3.12 Development of quantitative TRACEseq computational pipeline (logistic regression model)	155
4.3.13 TRACEseq reveals enrichment of 8mer seed match in true targets and abundant MREs in UTRs	158
4.3.14 TRACEseq uncovers previously validated and novel MREs.....	160
4.4 Discussion.....	162
4.4.1 Quantification of MRE accessibility by multiplex TRACE assay with grouped ASOs.....	162
4.4.2 Quantification of MRE accessibility by multiplex TRACE with single ASO	164
4.4.3 Seed-targeted RT initiation enables high-throughput NGS-based TRACE analysis	166
4.4.4 Development of quantitative TRACEseq computational pipeline (logistic regression model)	169
4.4.5 Comparison to other high-throughput <i>cis/trans</i> target identification techniques.....	171
4.4.6 Final Conclusions.....	175

Chapter 5: Towards Characterisation of the *bicoid* short RNA

Interactome.....	176
5.1 Introduction.....	176
5.2 Chapter Aims	183
5.3 Results.....	184

5.3.1 <i>Drosophila</i> ovaries express miRNAs and piRNAs predicted to target <i>bcd</i>	184
5.3.2 Unbiased mapping across the <i>bcd</i> 3'UTR using TRACE.....	188
5.3.3 RNA accessibility reflects known functional <i>cis</i> regulatory elements in the <i>bcd</i> 3'UTR	191
5.3.4 The <i>bcd</i> 3'UTR harbours several differentially bound MREs across the OET.....	194
5.3.5 Design and implementation of <i>bcd</i> RNA pulldown strategy	195
5.3.6 Generation of antisense probes	196
5.3.7 <i>Bcd</i> RNA pulldown optimisation	199
5.3.8 Confirmation of <i>bcd</i> capture efficiency by ddPCR.....	201
5.3.9 Implementation of a small RNA sequencing protocol	203
5.3.10 Interactome capture isolates abundant miRNAs associated to <i>bcd</i>	206
5.3.11 miRNA- <i>bcd</i> interactions are decreased in early embryos	210
5.3.12 <i>bcd</i> 3'UTR confers post-transcriptional repression in S2R+ cells	211
5.4 Discussion.....	214
5.4.1 <i>Drosophila</i> ovaries contains miRNAs and piRNAs predicted to target <i>bcd</i>	214
5.4.2 Development and implementation of a <i>bcd</i> small RNA interactome capture protocol	218
5.4.3 <i>bcd</i> repression – small RNA or protein?	224
5.4.4 Final conclusions.....	228
Chapter 6: General Discussion and Future Perspectives	229
6.1 <i>In vivo</i> detection of miRNA-target interaction events (miR-TRACE) as a novel miRNA target identification platform.....	230

6.2 Elucidating the molecular basis for <i>bcd</i> translational quiescence during oogenesis	234
6.3 Final conclusions	236
Chapter 7: References	237
Appendix: Primers and ASOs	268

List of Figures

Figure 1.1 Biogenesis of miRNAs	7
Figure 1.2 miRNA mechanism of action in animals.....	13
Figure 1.3 Different MRE types typically found in animals	23
Figure 1.4 RNAseq for profiling miRNA targets.....	27
Figure 1.5 Example of SILAC approach for miRNA target discovery.....	28
Figure 1.6 Ribosome profiling for miRNA target discovery	30
Figure 1.7 AGO-HITS-CLIP is a powerful tool for miRNA target discovery.....	32
Figure 1.8 PAR-CLIP	35
Figure 1.9 iCLIP	37
Figure 1.10 iPAR-CLIP and CLASH.....	39
Figure 1.11 Artificial RNA bait approaches	41
Figure 1.12 miR-CATCH captures native mRNAs	43
Figure 1.13 mRNA target isolation by miR-CLIP	45
Figure 1.14 <i>Drosophila</i> oogenesis.....	48
Figure 1.15 Maternal mRNAs are localised to opposite poles in the oocyte	52
Figure 1.16 <i>bcd</i> mRNA secondary structure	56
Figure 2.1 Derivation of true positives for model training	76
Figure 3.1 RNase H involvement in miR-TRACE.....	96
Figure 3.2 miR-TRACE assay	98
Figure 3.3 Dilution of lysis buffer improves dynamic range.....	100
Figure 3.4 Stepwise calculation of RPS and miR-TRACE score.....	102
Figure 3.5 miR-TRACE detects miRISC occupancy at <i>crok</i> MRE	103
Figure 3.6 Calculation of TRACE score using <i>rp49</i>	106

Figure 3.7 Normalisation using <i>rp49</i> displays superior sensitivity	107
Figure 3.8 Selection of miR-TRACE training data	109
Figure 3.9 miR-TRACE scores of twelve MREs	110
Figure 3.10 TRACE score in comparison to MRE score and fold change	111
Figure 4.1 Multiplexing strategies to improve TRACE throughput	121
Figure 4.2 Increasing the assay throughput <i>via</i> read out multiplexing	123
Figure 4.3 Overview of group selection algorithm	127
Figure 4.4 Grouped TRACE scores of different MRE classes	129
Figure 4.5 Grouped TRACE assay reflects performance of solo TRACE	131
Figure 4.6 Shorter ASOs can still cleave MREs	133
Figure 4.7 Shortened ASOs still detect difference in <i>crok</i> MRE	134
Figure 4.8 Single ASO, multiplexed TRACE scores of MRE classes	136
Figure 4.9 Single ASO TRACE assay reflects performance of solo TRACE	138
Figure 4.10 Seed specific priming engenders effective reverse transcription ...	141
Figure 4.11 Introduction of TRACEseq controls	144
Figure 4.12 6mer RT primer motif is enriched in library preparation	146
Figure 4.13 MiSeq sequencing provided insufficient depth	148
Figure 4.14 Vast majority of reads map to ribosomal RNA regions	149
Figure 4.15 DASH removes contaminating rRNA	151
Figure 4.16 Post-DASH sequencing generates high MRE coverage	154
Figure 4.17 ROC of TRACEseq logistic model	157
Figure 4.18 Extended 6mer motif analysis and gene location	159
Figure 4.19 Logistic regression predicts 48 MREs	161
 Figure 5.1 <i>Bicoid</i> carries putative target sites for piRNAs	 187
Figure 5.2 Exploratory TRACE mapping	188

Figure 5.3 TRACE map of the <i>bicoid</i> 3' UTR	190
Figure 5.4 RNase H Sensitivity of <i>bicoid</i> reflects <i>cis</i> elements	192
Figure 5.5 Predicted MREs increase accessibility across OET	194
Figure 5.6 Schematic <i>bicoid</i> pulldown overview.....	196
Figure 5.7 Generation of antisense probes	198
Figure 5.8 Development of pulldown protocol	200
Figure 5.9 <i>bicoid</i> can be specifically captured from ovaries.....	202
Figure 5.10 Bioanalyzer traces of small RNA library preparation.....	204
Figure 5.11 Pippin size selection cleans up library trace	206
Figure 5.12 Top candidates of the <i>bicoid</i> pulldown	208
Figure 5.13 <i>bicoid</i> contains MREs for all top miRNA pulldown candiates.....	209
Figure 5.14 qRT-PCR validation of pulldown	210
Figure 5.15 <i>bicoid</i> 3' UTR represses reporter but not <i>via</i> miR-184	212

List of Tables

Table 2.1 Buffers & Solutions.....	66
Table 5.1 miRNAs expressed in the ovary	185

Chapter 1: Introduction

Unlike DNA, which is organised in a regular formation around histone proteins, mRNA exists as complex ribonucleoprotein particles (RNP) that form as a consequence of multiple *cis* and *trans* interactions (Gebauer et al., 2012). *Trans*-acting RNA binding proteins (RBPs) and short regulatory RNAs such as microRNAs (miRNAs), recognise *cis* elements located in the mRNA sequence and thereby exert control over a sophisticated gene expression program. In particular, interactions between miRNAs and *cis* elements known as miRNA response elements (MREs) can have profound consequences on the stability of a target RNA and constitute a major mechanism to control and fine-tune gene expression (Bartel, 2009). By and large, the *cis/trans* recognition code grants RNA access to the cell's enormous repertoire of post-transcriptional regulators, and thereby allows the cell to direct thousands of different RNAs into distinct fates. Therefore, decoding RNA-RBP and RNA-RNA (in particular RNA-miRNA) networks lies at the heart of understanding post-transcriptional gene regulation.

Since genetic analysis uncovered the first *cis/trans* axis regulating the translation of ferritin mRNA in 1987, systems approaches have established post-transcriptional regulation as a powerful widespread mode of gene expression regulation (Hentze et al., 1987; Schwanhäusser et al., 2011). However, the functional diversity of *cis/trans* interactions also offers ample opportunity for errors, which is reflected in myriad diseases caused either by mutations in *cis* elements or perturbations in *trans* interacting factors (e.g. RBPs) (Castello et al., 2013a; Cooper et al., 2009). Specifically, through loss-of-function (LOF) deletions of *cis* elements mRNAs may escape translational repression, thereby creating potentially pathogenic levels of protein. Conversely, gain-of-function (GOF) mutations may subject RNAs to ectopic regulation by miRNAs or RBPs.

Although the general principles governing RNA-RNA recognition have already been plentifully exploited in RNA antisense technology and experimental therapy, our knowledge of miRNA target interactions *in vivo* remains fragmentary (Janssen et al., 2013; Kole et al., 2012). At the heart of this quest is the elucidation of the molecular logic by which miRNAs recognise their respective targets amidst thousands of potential base pairing partners. Despite the existence of powerful computational tools to predict miRNA target sites, the most fundamental question pertaining to miRNA targeting often remains unresolved: is a particular miRNA target site physiologically active in a specific tissue at a specific time? Addressing this question largely depends on the technologies available to interrogate miRNA-target interactions *in vivo*. **In this thesis, I have developed *in vivo* detection of miRNA-target interaction events (miR-TRACE) to identify novel and validate known MREs in their native context. Furthermore, I have specifically adapted the core principles of this technology to shed new light on the long-standing phenomenon of *bicoid* mRNA silencing during *Drosophila* oogenesis.**

1.1 Technology development for miRNA target identification

1.1.1 The non-coding genome as a key player in post-transcriptional regulation

After a decade-long concerted effort the Human Genome Project (HGP) reported just 24,000 protein coding genes encoded in as little as 1.5% of human DNA (Lander et al., 2001; Venter et al., 2001). This observation immediately reiterated two paradoxical questions: First, why does the number of genes in an organism not increase with its complexity? Humans possess only marginally more genes than nematodes, which in turn have more genes than chicken (Taft et al., 2007). On the other end, rice plants contain more than twice the number of genes as humans. Second, how and why do species afford to propagate 98.5% apparently inert DNA?

The two questions are innately linked in their answer: Over the past 15 years it has become apparent that complexity is generated through the elaborate execution of finely orchestrated regulatory programmes that have, interestingly, been attributed mostly to the 98.5% of non-coding DNA. For example, 80.4% of the human genome displays biochemical activity (transcription factor binding, transcription, chromatin structure organisation and histone modification), underpinning the existence of an intricate landscape of promoters and enhancers that interact in 3D over large distances with an enormous regulatory potential (ENCODE Project Consortium, 2012).

Although only 1.5% of the human genome is coding for proteins, as much as 74.7% is transcribed into RNA (Djebali et al., 2012). Even when adding together the well-studied examples of housekeeping non-coding RNAs such as ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear RNAs (snRNA) constituting the spliceosome, this value would not suffice to account for the extent of the non-coding genome. Interestingly, in 1965 the French biochemists Jacob and Monod speculated on the existence of “repressor RNAs” as a regulatory control element (Jacob and Monod, 1961). Indeed since the early 1990s, *regulatory* non-coding RNAs (ncRNA) have been reported on a case-by-case basis until their widespread abundance was uncovered by high-throughput sequencing in the early 2000s (Lagos-Quintana et al., 2001; 2003; Lai et al., 2003; Lau et al., 2001; Reinhart et al., 2002).

One of the first accounts of a regulatory non-coding RNA dates back to 1991, describing the *Xist* RNA which displayed no apparent open reading frames (ORF) nor association with ribosomes (Brockdorff et al., 1992; Brown et al., 1991). Henceforth, the remarkable function of non-protein based macromolecules to conduct X chromosome inactivation has led to the inception of an entire field studying long non-coding RNAs (lncRNAs) operating both at the DNA (Engreitz et al., 2013; Simon et

al., 2013; Ulitsky et al., 2011) level as well as on RNAs (Tan et al., 2014). The most recent class of ncRNAs to be discovered are so-called circular RNAs (cRNA) which are generated through back-splicing of introns (Hansen et al., 2013; Memczak et al., 2013). Arguably, the best-studied clade of ncRNAs is those of short (<30 nucleotides) non-coding regulatory RNAs which comprises microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and endogenous short-interfering RNA (endo-siRNAs). Out of this ensemble, miRNAs have emerged as a dominant class of vital orchestrators of post-transcriptional regulation of gene expression.

1.1.2 miRNAs

MicroRNAs (miRNAs) are short non-coding RNAs, ~22 nucleotides (nt) long, that have emerged as a rich and intricate layer of post-transcriptional regulation. First discovered in 1993 as a developmental switch regulating *lin-14* expression in *C. elegans* (Lee et al., 1993; Wightman et al., 1993), miRNAs have henceforth been uncovered in all clades of life, including plants, unicellular organisms and animals, but also viruses. It is estimated that 1-3% of the mammalian genome encodes miRNAs (Bartel, 2004). Their target sites, also known as miRNA response elements (MREs), have been located in almost all coding RNAs (Friedman et al., 2009). miRNAs mediate a host of major biological processes, including developmental transitions (Lee et al., 1993; Wightman et al., 1993; Wu et al., 2012), stem cell maintenance (Shcherbata et al., 2007), as well as cell fate determination (Cayirlioglu et al., 2008).

While essential orchestrators of cellular physiology under homeostasis, the aberrant expression and function of miRNAs has been, not surprisingly, related to a wide spectrum of clinical pathologies. For instance, miRNAs (oncomiRs) have been found to play causative roles in neoplastic transformation (He et al., 2005). Derailed miRNA expression has further been implicated in a host of cardiovascular conditions (Ikeda

et al., 2007; Reilly et al., 2016) as well as in cardiac failure (Thum et al., 2008). Furthermore, miRNAs have emerged as important factors in the control of inflammation, for example as regulators of Toll-like receptors (O'Neill et al., 2011). These properties make them promising therapeutic targets. For example, miRavirsen, a first-in-class locked nucleic acid (LNA) antisense inhibitor of hepatic *miR-122*, has recently demonstrated efficacy in reducing hepatitis C virus (HCV) propagation in Phase II clinical trials (Janssen et al., 2013).

The analysis of multiple patient tumour samples revealed distinct miRNA signatures in different cancer types, which outperformed mRNA profiling in terms of cancer classification (Lu et al., 2005). The same study also demonstrated the potential of miRNAs to highlight the developmental lineage of which a tumour is derived from as well as its differentiation state. These findings, together with the observation that miRNAs are frequently secreted into the extracellular space and stable in blood serum and other body fluids, make them specific, robust and above all non-invasive biomarkers (Etheridge et al., 2011).

Since miRNAs obey standard Watson-Crick base pairing, they have been hailed, together with their closely related short interfering RNAs (siRNAs), as therapeutic 'magic bullets' that can target virtually any RNA sequence (Kole et al., 2012). This is of particular interest for diseases where it has been impractical to target pathogenic proteins with small molecule inhibitors.

All these advances rely on a deep characterisation of the biology of miRNAs as well as their *modus operandi*. To address this, the miRNA field has loosely branched into three areas of active investigation: *i*) miRNA biogenesis and regulation, *ii*) effector mechanisms of miRNA silencing, *iii*) miRNA target identification. While the origin and biochemistry of miRNAs are prerequisite for our understanding of their activity, their gross biological consequences eventuate from the target pathways they regulate.

Hence, in the next few sections I will briefly introduce *miRNA biogenesis* and *mechanisms of miRNA silencing*, and then provide an up-to-date comprehensive overview of the target identification field, which directly relates to the main subject of this thesis.

1.1.2.1 Biogenesis of miRNAs

The production of miRNAs is complex and provides various checkpoints for controlling mature miRNA levels in the cell (Ha and Kim, 2014; Winter et al., 2009). Although most miRNA genes reside in introns of coding or non-coding transcripts, they can also be found in exons (Kim et al., 2009b; Ozsolak et al., 2008). Therefore, their expression can be controlled either by the same promoter as the host gene, albeit often with a different transcription start site, or by their own, independent promoter (Monteys et al., 2010). The majority of miRNAs are transcribed by RNA Polymerase II (Pol II) (Figure 1.1) and are as a consequence processed from capped and polyadenylated transcripts (Cai et al., 2004; Lee et al., 2004).

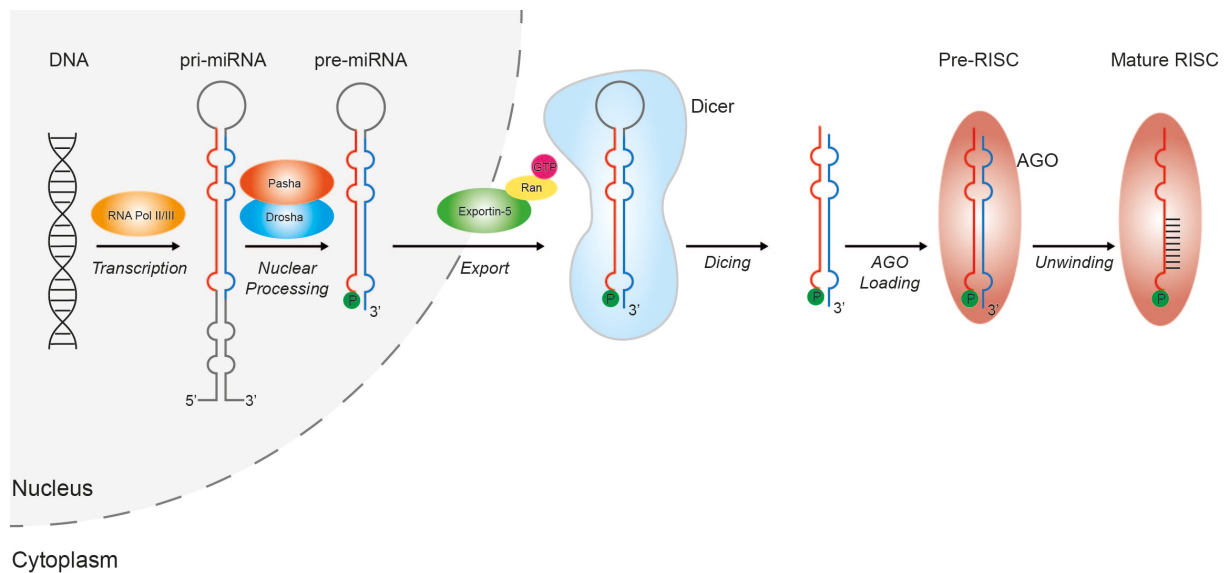


Figure 1.1 | Biogenesis of miRNAs. (Adapted from Winter et al., 2009 & Ha et al., 2014). miRNAs genes are most frequently transcribed by RNA Pol II into primary-miRNA transcripts (pri-miRNA). Pri-miRNAs are processed by the nuclear RNase III enzyme Drosha and its co-factor Pasha to pre-miRNAs. Pre-miRNAs are shuttled to the cytoplasm by an Exportin-5 transporter complex. In the cytoplasm, pre-miRNAs are further processed by the RNase III Dicer. Following this maturation, miRNA duplexes are loaded into AGO, assembling the miRNA-induced-silencing-complex (miRISC). The duplex is then unwound and the passenger strand ejected, generating a mature miRISC complex with a single stranded guide miRNA ready to screen targets.

Less frequently, miRNAs are produced by RNA Polymerase III (Pol III), for example in the case of some miRNAs of the herpes virus family (Pfeffer et al., 2005). Pol III transcribed transfer RNA (tRNA) transcripts can also harbour miRNAs (Babiarz et al., 2008). Transcription by Pol II subjects miRNA genes to positive and negative regulation by common transcription factors such as ZEB1 and ZEB2, MYC, p53 and MYOD1 (Kim et al., 2009b; Krol et al., 2010). Naturally, miRNAs are also regulated by epigenetic mechanisms, such as DNA methylation and histone modification (Saito et al., 2006; Scott et al., 2006).

Following transcription, the earliest precursor of the mature miRNA, the primary miRNA (pri-miRNA), undergoes several steps of nuclear processing (Lee et al., 2002) (summarised in Fig. 1.1). Pri-miRNAs can be several hundred nucleotides

long, form a stem of 33-35 bp and a hairpin loop, and have free 5' end 3' ends. The microprocessor, an enzyme complex of RNases and co-factors, recognises pri-miRNAs and processes them. Integral to this processor is the 160kDa RNase III endonuclease Drosha, which acts on the dsRNA stem of pri-miRNA to generate ~65nt long hairpins, called precursor miRNAs (pre-miRNA) (Lee et al., 2003). Drosha further requires the presence of Pasha (known as DGCR8 in humans) (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). Towards the carboxyl terminus (C-terminus) Drosha possesses a dsRNA-binding domain as well as two RNase III domains (RIIID) that dimerise intramolecularly and comprise the catalytic centre of the microprocessor (Han et al., 2004). The first domain cleaves the 3' end of the pri-miRNA stem, and the second domain processes the 5' end resulting in a 3' overhang of two nucleotides (Auyeung et al., 2013; Blaszczyk et al., 2001). This process is facilitated by Pasha which has two RNA binding domains (RBD) to position the pri-miRNA substrate (Han et al., 2006). Drosha cuts the pri-miRNA hairpin ~11 bp upstream of the basal intersection of ssRNA and dsRNA and ~22 bp downstream from the terminal loop junction (Han et al., 2006; Zeng et al., 2005).

The junction near the basal end serves to locate the desired cleavage target site (Auyeung et al., 2013; Han et al., 2006). The apical junction facilitates this process (Ma et al., 2013; Zeng et al., 2005). Furthermore, it has been shown that sequence intrinsic motifs can regulate miRNA processing, for example the CNNC motif in the basal region acts as a binding site for DEAD-box RNA helicase p72 (DDX17) which augments the activity of Drosha (Auyeung et al., 2013).

Processing of pri-miRNAs into pre-miRNAs is likely to take place co-transcriptionally, prior to splicing, as pri-miRNAs are abundant in chromatin-associated nuclear fractions (Morlando et al., 2008; Pawlicki and Steitz, 2008). Interestingly, it has also

been proposed that Drosha regulates general mRNA stability through cleaving pri-miRNA substrates that reside in longer mRNAs (Ha and Kim, 2014).

Since the microprocessor constitutes the nexus of pre-miRNA production, several mechanisms have evolved to regulate its activity. Interestingly, Drosha and DGCR8 are capable of cross-regulating each other (Han et al., 2009). *Via* protein-protein interactions DGCR8 stabilises Drosha which in turn cleaves DGCR8 mRNA in exon 2 (Han et al., 2009; Kadener et al., 2009; Yeom et al., 2006). Further regulatory control is mediated through post-translational modifications that affect protein stability (Herbert et al., 2013; Wada et al., 2012). Nuclear localisation of Drosha requires phosphorylation by glycogen synthase 3 β (GSK3 β) (Tang et al., 2011; 2010). The microprocessor is further controlled by additional RBPs that associate both with Drosha and pri-miRNA, for example p68 (DDX5) and p72 who are essential for the production of certain miRNAs (Fukuda et al., 2007).

Pre-miRNAs associate with GTP-binding nuclear protein RAN-GTP and form a transport complex with exportin 5 (Figure 1.1) (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003). Once in the cytoplasm, pre-miRNAs enter the final stages of their maturation. The key enzyme mediating this process is Dicer, another RNase type III endonuclease of approximately 200kDa (Bernstein et al., 2001; Hutvagner et al., 2001).

Somewhat similar to Drosha, Dicer bears two RNase III domains at the C-terminus that dimerise intramolecularly to form a catalytic centre (Zhang et al., 2004). The amino-terminal helicase domain interacts with the terminal loop of pre-miRNAs (Tsutsumi et al., 2011; Gu et al., 2012). Dicer's PIWI-AGO-ZWILLE (PAZ) domain recognises the terminus of the pre-miRNA (Macrae et al., 2006; Park et al., 2011; Tian et al., 2014). The distance between the PAZ domain and the RNase III catalytic centre roughly corresponds to ~22 nt and has been dubbed a 'molecular ruler'

(Macrae et al., 2006; Tian et al., 2014). The so-called '3' counting rule' stemmed from observations that Dicer tends to cleave its pre-miRNA at a fixed distance of 21-25nt from the 3' end (Macrae et al., 2006; Tian et al., 2014). However, the opposite (5' counting rule) is also possible, where Dicer cleaves 22 nt away from the 5' end, yet this has only been observed in mammals and flies (Park et al., 2011).

Dicer itself can be regulated by various co-factors as well as by its own products. Interestingly, RNase III proteins have been found to frequently bind other dsRBD proteins, for example *Drosophila* Dicer interacts with Loquacious, which has two isoforms, each comprising three dsRBDs (Förstemann et al., 2005). Both isoforms are required for the production of most miRNAs (Fukunaga et al., 2012). Further regulation is mediated *via* feedback loops: for instance, human *DICER1* mRNA carries target sites for let-7, exerting negative feedback onto *dicer* if miRNAs become overabundant in the cell (Forman et al., 2008; Tokumaru et al., 2008).

The final stages of miRNA maturation include miRISC loading of the RNA duplex, unwinding of the double strand and ejecting the non-targeting ('passenger' strand). The key component of the miRISC complex is the argonaute (AGO) (Meister, 2013). AGO proteins exist in three clades: AGO, PIWI and worm-specific AGO (WAGO). The ubiquitously expressed AGO proteins are the effectors of miRNAs and siRNAs, whilst PIWI proteins are restricted to the germ line. The structure of AGO proteins is of a bilobal conformation: the N-terminal lobe contains an N-terminal domain as well as a PAZ domain, the C-terminal lobe is composed of a middle (MID) domain and a PIWI pocket. The 5' monophosphate of the guide RNA interacts with a 5' phosphate binding pocket between the MID and PIWI domains (Elkayam et al., 2012; Ma et al., 2005; Nakanishi et al., 2012; Schirle and Macrae, 2012). The guide strand of the RNA rests in the basic channel formed by the N-terminal lobe and extends to the PAZ domain where it is held in place. Importantly, nucleotides 2-7, the so-called

'seed' (see below), are forced into an A-form helix which thermodynamically favours scanning of putative targets. The PIWI domain, which is related to the RNase H domain is capable of cleaving ('slicing') perfectly complementary targets between position 10 and 11 from the 5' end of the miRNA (Parker et al., 2005; Song et al., 2004). Importantly, only AGO2 has slicer activity in mammals and flies (Ha and Kim, 2014; Liu et al., 2004).

Several mechanisms exist to sort distinct small RNA species into their respective RISCs. In flies, miRNAs always load into AGO1 whilst siRNAs load into AGO2. It is a central mismatch at position 9-10 that prevents miRNA duplexes from entering AGO2, which only accepts perfectly paired dsRNA stems (Czech et al., 2009; Förstemann et al., 2007). *Drosophila melanogaster* (*Dm*) Dcr-2 with help of dsRBD protein R2D2 also regulates AGO2 access entry to the miRISC as it preferentially binds fully complementary short RNA duplexes. Furthermore, AGO1 binds in general to a 5' uracil (U) which is commonly found in miRNAs, in contrast to AGO2 which prefers siRNAs with a 5' cytosine (C) (Kawamata et al., 2009). To prepare the miRISC complex for target scanning, the duplex needs to be unwound and the passenger strand ejected. In perfectly matched duplexes, AGO2 can simply cleave the passenger strand which is then removed by C3PO (Liu et al., 2009; Matranga et al., 2005). In case mismatches are present in the guide strand, mostly at position 2-8 as well as 12-15, cleavage does not occur, and instead the duplex is unwound (Kawamata et al., 2009). Which strand is propagated is determined primarily based on thermodynamic stability at the 5' ends of the RNA duplexes: the less stable 5' terminus generally becomes the guide strand (Khvorova et al., 2003; Schwarz et al., 2003). A second qualifier is the presence of a U at position 1, which is preferentially loaded into AGO (Hu et al., 2009). Ejected passenger strands are typically degraded rapidly.

1.1.2.2 Mechanism of action: Effector steps of silencing

Before discussing the downstream effector steps I am briefly outlining the process of translation initiation as one of the key processes which are antagonised by miRNA action. RNA Pol II transcribed mRNAs are fitted with a 5'-cap structure (m⁷GpppN) and a poly(A) tail, both of which are required for translation. They are essential to recruit the 43S pre-initiation complex consisting of the 40S ribosomal subunit, eIF-GTP-MET-tRNA^{Met}, eukaryotic translation initiation factor 3 (eIF3), eIF1, eIF1A and eIF5. Once in the cytoplasm, the 5' cap is bound by the eIF4F complex, which comprises the cap-binding protein eIF4E, the scaffold eIF4G as well as the helicase eIF4A (Fig. 1.2a). At the 3' end, the poly(A) tail is recognised by the poly(A)-binding protein PABP. Once PABP and eIF4G interact with each other, the entire mRNA is looped, eliciting 43S pre-initiation complex (small subunit) loading (Jackson et al., 2010). This complex moves in a 5' to 3' direction and once it encounters the first AUG start codon, the 63S large subunit is recruited to complete the formation of the functional 80S eukaryotic ribosome.

The very basic steps of miRNA-mediated repression are the following (Fig. 1.2) (Huntzinger and Izaurralde, 2011): 1) miRNA guided AGO recognises a target RNA *via* an MRE (Fig. 1.2a); 2) The miRISC complex recruits GW182, which in turn breaks open the closed loop mRNA structure required for translation (Fig. 1.2b); 3) GW182 recruits deadenylases that remove the poly(A) tail (Fig. 1.2c); 4) miRISC recruits DCP2 which decaps the mRNA with the help of other decapping factors (Fig. 1.2e); 5) The exposed RNA is degraded by XRN1 *via* 5'-to-3' exonuclear digestion (Fig. 1.2f). However, not all RNAs are degraded; there is a second mode of repression, translational inhibition, which does not entail target destabilisation (Fig. 1.2d). Although the overt consequences of miRNA targeting appear to be RNA destabilisation, the molecular intricacies and the order of events is still under active investigation (Guo et al., 2010; Iwakawa and Tomari, 2015). I will now briefly

summarise evidence for either mode and then evaluate their relative contributions to the overall effect of miRNA-mediated silencing.

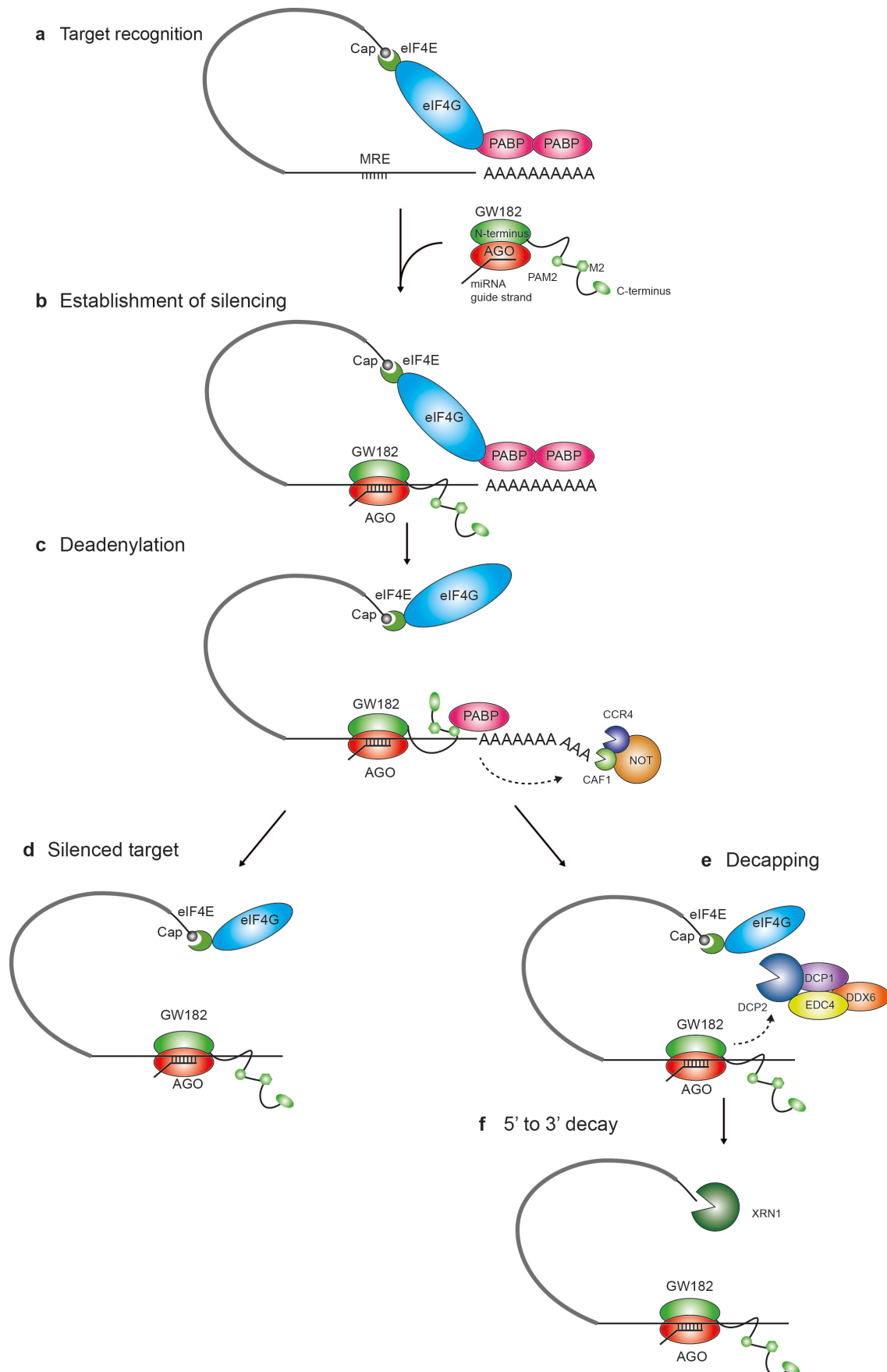


Figure 1.2 | miRNA mechanism of action in animals. (Adapted from Huntzinger et al. 2011) **a**, the target RNA is shown in closed loop formation, which is mediated through an interaction between the poly(A) binding protein (PABP) bound to the poly-A tail and eukaryotic translation initiation factor 4G (eIF4G) which recognises the 5' cap. This closed loop is the hallmark of translating mRNAs. **b**, the miRNA response element (MRE) located in the 3' UTR of the mRNA and upstream of its poly-A tail is recognised by a miRNA guiding the RISC complex. AGO further interacts with and thus recruits GW182 to the target RNA. **c**, via its PAM2 domain GW182 interacts with PABP and thereby breaks open the closed loop conformation. GW182 also recruits the CCR4-NOT deadenylase complex which now has access to and digests the poly-A tail. **d**, mRNA in open conformation is inaccessible to translation and can remain silenced or **e**, GW182 further recruits a decapping complex spearheaded by DCP2, but also bound by co-factors DCP1, EDC4 and DDX6, which removes the 5' cap. **f**, with its 5' cap removed, the mRNA becomes a substrate for the exonuclease XRN1 which eliminates the RNA via 5' to 3' digest.

In the early 2000s it was believed that animal miRNAs predominantly repress translation whilst plant miRNAs were considered to cleave their targets due to perfect complementarity (Ameres and Zamore, 2013; Olsen and Ambros, 1999; Seggerson et al., 2002). This notion has changed and it is now appreciated that animal miRNAs are capable of target degradation and plant miRNAs can elicit translational inhibition (Iwakawa and Tomari, 2015). If the miRNA bears extensive complementarity to the target RNA, AGO can cleave the mRNA between nucleotides 10 and 11 opposite the small RNA. In plants this is the dominant *modus operandi* (Rogers and Chen, 2013). However, in the animal kingdom, miRNAs primarily recognise their targets via incomplete base pairing, mostly determined by the miRNA seed (nucleotides 2-7) (Bartel, 2009). This generally engenders either translational inhibition or mRNA decay, or both. While the minimal miRISC complex (miRNA-AGO) provides the specificity code for targeting, the nexus of downstream effector silencing in animals are GW182 proteins. Three paralogs exist in vertebrates (TNRC6A (GW182), TNRC6B and TNRC6C) with GW182 being the sole protein in flies. In their amino terminal domain, GW182 proteins harbour glycine-tryptophan repeats (GW-repeats)

which are bound by the tryptophan binding pockets of AGO's PIWI domain, thus providing the link to the miRISC. At their carboxyl termini, GW182 proteins have a bipartite silencing domain (SD).

1.1.2.2.1 Translational inhibition post-initiation

The concept that miRNAs predominantly repress their targets *via* translational inhibition stemmed from observations in *C. elegans*, which showed that *lin-4* repression of *lin-14* and *lin-28* had no effect on their mRNA levels (Olsen and Ambros, 1999; Seggerson et al., 2002). The same studies found the repressed mRNAs associated with polysome fractions, suggesting that repression occurs subsequent to translation initiation. Sucrose gradient studies revealed that miRNA targets are indeed present in translated fractions. The so-called 'ribosome drop off' hypothesis (ribosome loading occurs but is not productive) attempted to provide an explanation for this observation (Petersen et al., 2006). Further lending support to this notion, it was claimed that cap-independent translation, for example *via* internal ribosome entry sites (IRES), can also be silenced by miRNAs, but this remains controversial (see below) (Petersen et al., 2006).

1.1.2.2.2 Translational inhibition pre-initiation

Conversely, it has also been reported that miRNA targets do not associate with ribosomes in sucrose gradient experiments, suggesting that translational inhibition is exercised before the initiation step (Pillai et al., 2005). Interestingly, this study claimed that the poly(A) tail is not a requirement for miRNA mediated silencing. This is in contrast to work by Humphrey *et al.* as well as Wang *et al.* who showed that a poly(A) tail is necessary for repression (Humphreys et al., 2005; Wang et al., 2006). Evidence for three different but most likely overlapping mechanisms has been presented: i) PABP displacement, ii) GW182 mediated recruitment of translational repressors, iii) abrogation of eIF4A from cap-binding complex eIF4F.

PABP displacement

It has been noted that the PAM2 motif present in the SDs of GW182 can interact with the C-terminal MLLE domain of PABP (Zekri et al., 2009), and this process is conserved across species (Kuzuoglu-Öztürk et al., 2012). If the interaction between GW182 and PABP is abrogated, silencing is impaired (Huntzinger et al., 2010). In flies, the binding between GW182 and eIF4G with PABP is mutually exclusive and overexpressing PABP relieves target repression (Zekri et al., 2009). This suggests that GW182 inhibits the interaction between eIF4G and PABP and thereby obstructs the initiation of translation by breaking up the loop structure required for ribosome loading.

In contrast, evidence exists that mRNAs without a poly(A) tail are still repressed by miRNAs (Eulalio et al., 2009), or GW182 tethering (Chekulaeva et al., 2009), suggesting a PABP-independent mechanism. Supporting this notion, depleting PABP or inhibiting it with PABP-binding protein 2 (Paip2) which abrogates eIF4G-PABP binding as well as PABP-poly(A) interactions, still yielded efficient silencing in S2 cell lysates (Fukaya and Tomari, 2011). Similar experiments in zebrafish showed the same result, hinting at other modes of repression (Mishima et al., 2012).

Recruitment of Translational Inhibitors through GW182

Non-polyadenylated transcripts can be repressed by GW182-mediated recruitment of the CCR4-NOT complex, suggesting deadenylases are capable of inducing translational repression beyond poly(A) reduction (Braun et al., 2011). This was proposed following tethering experiments of CAF1 (one of the factors of the CCR4-NOT deadenylase complex) which showed repression of mRNAs with and without poly(A) tails (Cooke et al., 2010). Conversely, knocking down CAF1 in *Drosophila* S2 cells had no effect on target repression (Fukaya and Tomari, 2012). Tethering a

mutated version of CAF1 that can no longer bind CNOT1 and CCR4, did not confer target repression in the presence or absence of the poly(A) tail. However, a CAF1 mutant with an intact CNOT1 binding pocket is still capable of target repression, suggesting that it can act through recruiting additional inhibitors *via* CNOT1, as it can no longer interact with CCR4 directly (Mathys et al., 2014). Taken together, these studies suggest GW182 can elicit target repression independently of deadenylation.

miRNA-mediated dissociation of eIF4A

Several studies revealed that replacing the 5' m⁷G-cap with an artificial A-cap analogue renders RNA refractory to translational repression *via* miRNAs (Fukaya and Tomari, 2012; Humphreys et al., 2005; Pillai et al., 2005). Further lending support to the notion that silencing takes place prior to translation initiation, it was shown in three independent studies that miRNA targets are refractory to silencing if translation initiation is IRES-mediated (Humphreys et al., 2005; Mathonnet et al., 2007; Wakiyama et al., 2007). Increasing the concentration of the initiation factor eIF4F components (cap-binding protein eIF4E, the scaffold protein eIF4G and the DEAD box RNA helicase eIF4A) *in vitro* attenuated silencing, suggesting both the binding factor eIF4F and miRISC are competing for mRNA. However, it remained uncertain which exact factor of the eIF4F complex was targeted by the miRISC. An elegant pull-down study employing UV-crosslinking to stabilise eIF4E and eIF4A on cap-labeled RNA revealed that in the presence of cognate miRNAs, eIF4A no longer co-precipitates (Fukaya et al., 2014). The levels of eIF4E however, were unaffected by the addition of the miRNA. A similar study in mammalian cell lysates uncovered that eIF4A_I and eIF4A_{II} are both removed from the cap-binding complex eIF4F upon addition of a cognate miRNA (Fukao et al., 2014). Hence, on the basis of these two studies it is plausible that miRNAs abrogate ribosome loading, with the exact order of

molecular events still under active investigation. Collectively, these results suggest that mRNA dissociation of eIF4A is one mechanism by which miRISC represses its targets.

1.1.2.2.3 Target degradation

There is ample evidence that miRNA overexpression leads to downregulation of predicted target mRNAs (Baek et al., 2008; Giraldez et al., 2006; Guo et al., 2010; Lim et al., 2005; Selbach et al., 2008). Conversely, depleting a miRNA leads to target upregulation. Taken together, the consensus appears to be that the predominant mode for target repression in animals is mRNA destabilisation.

GW182 is capable of recruiting two deadenylase complexes to the mRNA: CCR4-NOT and PAN2-PAN3. Perturbing PAN2 or PAN3 expression in cells led to minor effects on deadenylation (Chen et al., 2009). On the contrary, depleting CCR4-NOT modules resulted in marked reductions of mRNA deadenylation with subsequent destabilisation (Braun et al., 2011). Depleting cells for components of CAF1-CCR4-NOT results in upregulation of miRNA targets (Eulalio et al., 2009). Consistently, expressing dominant-negative variants also leads to target overexpression (Piao et al., 2010).

One key mechanism of GW182 entails exposing the poly(A) tail to deadenylases as well as recruiting these enzymes. Underpinning this concept, GW182 tethering to an mRNA removed PABP before the onset of deadenylation (Moretti et al., 2012). PABP contains several RNA-recognition motifs (RRM1-4) at its N-terminus, an unstructured proline-linker, and the so-called PABC or MLLE domain at its C-terminus (Kozlov et al., 2010). PABP binds PABP-interacting motifs (PAM) which are abundant in RNA translation and decay machinery (Kozlov et al., 2010). Interestingly, GW182 harbours two PAMs in the SD, thereby allowing it to interact with PABP (Jinek et al., 2010).

Post-deadenylation, mRNAs are predominantly degraded *via* the 5'-to-3' mRNA decay pathway. The miRISC complex recruits the decapping enzyme DCP2 (Rehwinkel et al., 2005). This process requires several co-factors such as DCP1, EDC4 (GE1), PAT as well as DEAD-box protein RCK (Me31B), which have been shown to immunoprecipitate with AGO (Chu and Rana, 2006; Eulalio et al., 2007b). Interestingly, while deadenylation generally seems to precede decapping, a poly(A) tail is not a requirement for decapping, suggesting that decapping can be performed independently (Béthune et al., 2012; Subtelny et al., 2014). With their poly(A) tail and 5'-cap removed, mRNAs become substrates for the 5'-to-3' exonuclease XRN1.

1.1.2.2.4 Relative contribution of decay vs. translational inhibition and order of events

mRNA destabilisation can occur independently of translation. Introducing secondary structures into the 5' UTR of reporter RNAs that renders them incompetent for translation do not prevent degradation *via* the miRNA pathway (Eulalio et al., 2009; Wu et al., 2006). In line with this observation, injection into zebrafish embryos of mRNA targets with a defective cap structure which abrogates translation, does not impede miRNA silencing (Mishima et al., 2006). Inhibiting translation initiation with hippuristanol or preventing elongation with cycloheximide had no effect on deadenylation of miRNA targets, further suggesting that degradation is not dependent on translation (Fabian et al., 2009). It has also been noted that deadenylation is not essential to miRNA action as reporters in which poly(A) tails have been replaced with self-cleavable ribozymes or histone stem loops are still subject to repression, albeit not to the same degree as polyadenylated equivalents (Eulalio et al., 2008; Wu et al., 2006).

Selbach *et al.* observed that 8h after transfection of a miRNA mimic, targets were mainly regulated at the protein level (Selbach et al., 2008). However, after 32 h changes in the mRNA levels were detected. Transfecting *miR-124* into HEK293T

cells revealed that 75% of cases of protein reductions could be attributed to mRNA degradation (Hendrickson et al., 2009).

Direct molecular evidence for mRNA decay as a dominant mechanism also stemmed from RNAseq experiments performed in tandem with ribosome profiling which revealed that 84% of cases of protein reduction can be attributed to a decrease in mRNA levels (Guo et al., 2010). Few cases of reduced protein output (as measured by ribosome profiling) were noted without concomitant mRNA degradation. In these cases, the study made a point for translational inhibition at initiation rather than during elongation ('ribosome drop off') as they observed the ribosome density remained constant on target RNA that were not subject to degradation (Guo et al., 2010). However, through monitoring ribosome occupancy, Bazzini and colleagues found that loss of ribosome occupancy of target RNAs precedes deadenylation and mRNA decay *via miR-430*, a major regulator of the maternal-to-zygotic transition (Bazzini et al., 2012). Similar observations were also made in *Drosophila* S2 cells using artificial sensors, which reported that translational inhibition is manifested before RNA degradation (Djuranovic et al., 2012). Taken together, although translational inhibition may precede mRNA destabilisation, mRNA decay is the predominant contributor to protein repression.

1.1.2.3 miRNA target identification strategies

While miRNA origin and biogenesis as well as the biochemistry of their silencing have generated immense excitement and grown into *bona fide* scientific fields, deciphering the biological functions of miRNAs relies on defining their direct endogenous mRNA targets. As such, perhaps the most central question concerning miRNA biology is how do miRNAs choose their targets? In plants, this is relatively straightforward, as miRNAs act similar to siRNAs in the sense that targeting is manifested by extensive complementarity (>10bp), suggesting that these interactions

can be uncovered by sequence alignment algorithms (Rhoades et al., 2002). However, in animals extended complementarity is rare (Rhoades et al., 2002) and mRNA cleavage is the exception rather than the rule (Davis et al., 2005; Yekta et al., 2004). Instead, animal miRNAs recognise their targets primarily through the 'seed' region (nucleotides 2-7 counting from the 5' end of the miRNA) (Lai, 2002; Lewis et al., 2003; Stark et al., 2003). This concept emerged when Lai noticed that 5' ends of specific miRNAs exhibited perfect (8/8 bp) complementarity to 3' UTR motifs such as the *k-box* required for negative RNA regulation (Lai, 2002). Subsequent studies in *C. elegans* revealed the 5' seed region as the most conserved feature of miRNAs (Lim et al., 2003). This was consistent with one of the first systematic efforts to predict animal miRNAs which revealed the highest signal:noise ratio if obligate 5' end miRNA pairing was chosen as a parameter (Lewis et al., 2003). The importance of the seed was later underpinned in translational repression experiments, where miRNA 5' ends were particularly sensitive to mutagenesis, whilst the 3' region could tolerate mismatches (Brennecke et al., 2005; Doench and Sharp, 2004; Kloosterman et al., 2004). Variation of the seed influences the potency of target repression (Bartel, 2009) (Figure 1.3). For instance, 8mer seeds are reported to confer strongest repression, followed by 7mer-m8 (seed paired at miRNA position 8), in turn followed by 7mer-A1 (position 1 is adenosine), whilst 6mer (only paired 2-7) are least repressive (Brennecke et al., 2005; Lewis et al., 2005). Importantly, the 'seed-centred' view of miRNA-target interactions has recently been challenged as high-throughput target identification platforms uncovered a host of non-canonical miRNA-target configurations (Helwak et al., 2013). However, the efficacy of non-canonical target sites remains controversial (Agarwal et al., 2015).

Perhaps the greatest difficulty in identifying *bona fide* targets stems from the incomplete and imperfect (e.g. G:U wobble) base pairing between animal miRNAs

and mRNAs. As such, miRNAs can potentially interact with hundreds of target RNAs and *vice versa* every target RNA can interact with a large number of miRNAs. Therefore, a key consideration in miRNA studies is to recognise 'real' miRNA-target interactions with biological functions, and isolate these from the slew of random binding events likely to occur co-incidentally. Conversely, co-expression of miRNAs and target transcripts in the same cell is necessary but by no means sufficient for regulation. Underpinning this observation, it has been abundantly noted that the primary MRE sequence is only one determinant for targeting. Adding further complexity to this matter, miRNA-target interactions are under the governance of MRE secondary structure (Kertesz et al., 2007) and are regulated by an interplay with RNA binding proteins (Kedde and Agami, 2008; Meisner and Filipowicz, 2011; van Kouwenhove et al., 2011). Association with RBPs can result in both positive and negative regulation. For instance, if transcripts encoding E2F2 and p27 associate with proteins of the *Pumilio* family, the effect of miRNA mediated translational repression is augmented (Kedde et al., 2010).

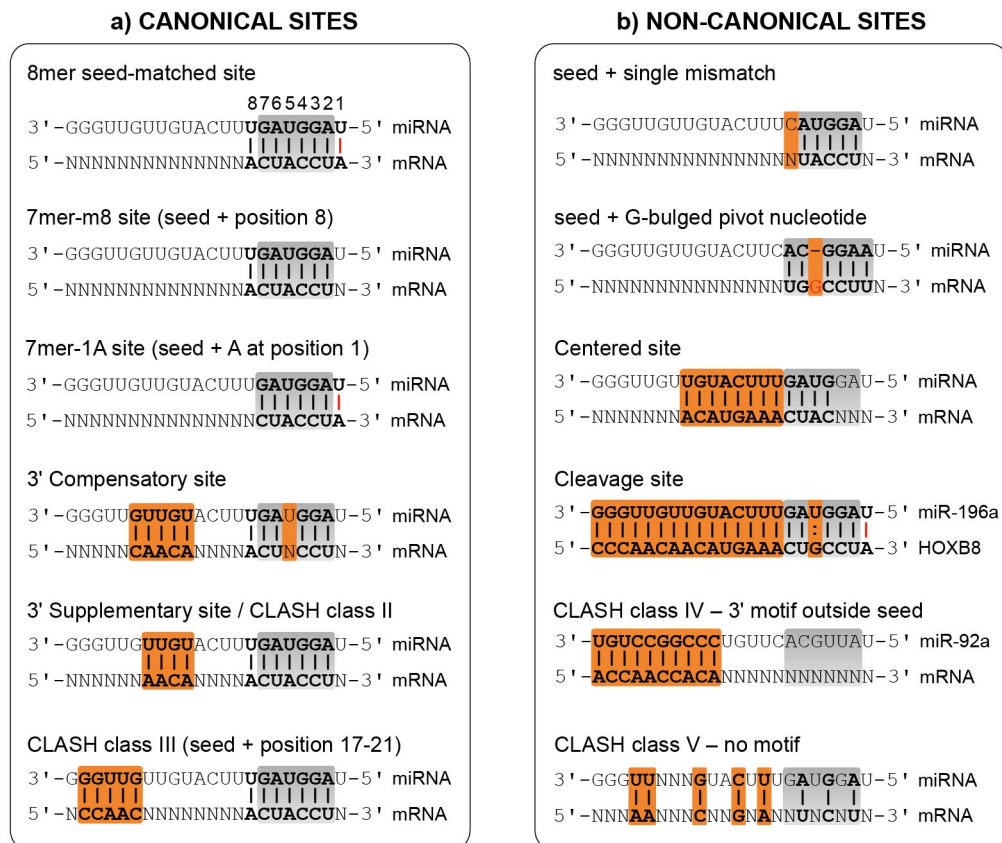


Figure 1.3 | Different MRE types typically found in animals. (Adapted from Steinkraus *et al.* 2016) **a**, Canonical MREs display perfect complementary across the miRNA seed sequence. The strongest repression is mediated by 8mer which are matched from 1-8. 7mer-m8 are matched in position 2-7 but also at position 8. 7mer-1A have complementarity at position 2-7 but are also matched at position 1 (A-U). 3' compensatory sites reveal a mismatch or G:U wobble in the seed region but compensate for this with complementarity outside the seed region. CLASH revealed two site types that were matched at the seed but revealed complementarity at position 13-16 (CLASH class II) and 17-21 (CLASH class III). **b**, Noncanonical sites exhibit mismatches in their seed region. AGO-CLIP data revealed a frequent bulge nucleotide (so-called 'pivot') between position 5 and 6 of the miRNA. Centered sites exhibit incomplete seed pairing but include complementarity immediately adjacent to it. Cleavage sites are those with extensive complementarity, however, they are rare in animals. CLASH class IV MREs display a minimum of 9 consecutive base pairs outside the seed. CLASH class V lack any dominant motif. Grey box = miRNA seed region (nucleotides 2-7); orange box = typical motif for respective class; bold = base paired; : = G:U wobble; red bar = complementary at miRNA position 1 (improbable to engage in base pairing as buried inside AGO)

The opposite has been observed in zebrafish primordial germline cells, where dead end 1 protein occludes *miR-430* binding sites in *tdrd7* and *nanos*, thereby relieving target repression (Kedde et al., 2007). Furthermore, HuR (ELAV) AU-rich element (ARE) binding proteins are well known to affect miRNA-mediated repression (Bhattacharyya et al., 2006b; Kim et al., 2009a). For instance, stress induction in Huh7 hepatoma cells counteracts *miR-122* repression of *CAT-1* via HuR association and subsequent translocation from processing bodies (P-bodies) to ribosomes (Bhattacharyya et al., 2006a). A global effect on miRNA regulation by RBDs was attested through transcriptome-wide HuR-RNA binding maps which uncovered that HuR binding frequently occurs close to miRNA target sites, albeit typically not overlapping the predicted MRE sequence (Lebedeva et al., 2011; Mukherjee et al., 2011).

Since it has become clear that discovering *bona fide* targets is key to understanding miRNA functions, a host of methodologies has been developed for the purpose of identifying and validating such interactions. These include a large number of bioinformatics platforms as well as experimental approaches (Steinkraus et al., 2016). Albeit often imperfect, the fact that miRNAs obey stereotypical rules of Watson-Crick base pairing has been abundantly exploited in computational algorithms designed to predict miRNA-target binding interactions. The first target identification algorithm was developed in 2003 (Enright et al., 2003) and has since been followed by a profusion of *in silico* tools. These can generally be divided into two branches: filtering and machine learning (ML) algorithms. Filtering methods rely on experimentally defined criteria (e.g. seed) against which putative miRNA-target interactions are screened. In contrast, ML approaches are based on larger datasets where a classifier is trained with true (positive) and false targets (negative), and then

applied to an unknown dataset to divide the entries into likely positive and negative interactions according to the parameters it has been trained on.

Most filtering approaches are based on target site intrinsic factors, such as configuration of the MRE, thermodynamics of miRNA-target pairing, as well as evolutionary conservation of both the miRNA and the target. Since most miRNA-target interactions appear to involve seed pairing, the first algorithms employed seed pairing as their primary feature. For example, TargetScan, developed in the Bartel and Burge laboratories, scans mRNAs for complementarity to miRNA position 2-8 and calculates the net free energy of duplex formation between miRNA and target using the RNAfold code (Hofacker, 2003; Lewis et al., 2003). An updated version now permits the search of weakly conserved targets and miRNAs and also allows mismatches in the seed if the site has 3' compensatory binding. To produce a score for each interaction, TargetScan considers the architecture of the MRE (e.g. seed type and 3' pairing) and the MRE context (e.g. the position in the 3' UTR, the number of target sites as well as the AU content) (Garcia et al., 2011).

The miRanda algorithm developed by Enright *et al.* has also been a commonly used miRNA target prediction tool (Enright et al., 2003). The algorithm employs slightly more permissive parameters than TargetScan and thus typically predicts more putative target sites. miRanda scans 3' UTRs for complementarity with a seed bias, calculates the free energy of duplex formation and also considers conservation through a hidden Markov model from the Phast-Cons algorithm. The software has since been improved to calculate the mirSVR, a regression model which weighs both sequence and context parameters of the duplex (Betel et al., 2010).

Despite ever increasing sophistication of computational prediction platforms, an exponentially growing amount of experimental training data, as well as the materialisation of artificial intelligence (AI) models, most *in silico* approaches are still

limited by high false positive (46-63%) and false negative rates (44-82%) (Lee et al., 2015; Rajewsky, 2006). Consequently, experimental approaches remain the gold standard for the identification of physiological miRNA targets. The earliest miRNA-target axes were extrapolated from reverse genetic screens, notably in worms. Putative target genes were inferred as those whose mutagenesis opposed or restored the phenotypes caused by certain miRNA loss-of-function. However, reverse genetic screens have lost traction over the years, as they are labour-intensive and difficult to perform in certain model organisms. Further complicating this approach, miRNA null mutants (esp. in *C. elegans* and mice) rarely cause dramatic detectable phenotypes, which is paramount for identifying causal genes (Miska et al., 2007; Park et al., 2012). As a consequence, a spectrum of alternative miRNA target identification methodologies has been developed, including miRNA loss-of-function or gain-of-function profiling, as well as advanced techniques assessing physical binding between miRNAs and existing targets.

1.1.2.3.1 Profiling Approaches

The observation that the level of miRNAs and putative mRNA targets frequently anti-correlate has been abundantly exploited to screen for physiological miRNA targets. If a set of candidates is already defined, qRT-PCR can rapidly inform on target levels. However, for the purpose of unbiased genome wide target identification, high-throughput techniques are required. The earliest profiling experiments were performed using microarrays which revealed that a single miRNA can regulate hundreds of mRNAs (Lim et al., 2005). However, the advent of massively parallel sequencing technologies, foremost RNA sequencing (RNAseq), has revolutionised high-throughput readouts (Fig. 1.4). Indeed, comparing data generated by microarray and RNAseq for the *miR-155* targetome revealed that RNAseq is capable of uncovering a significantly larger number of putative targets (Xu et al., 2010).

Furthermore, RNAseq provides greater sensitivity (in terms of discrete counts), which is particularly important for miRNA studies given their relatively modest effects on the majority of their targets.

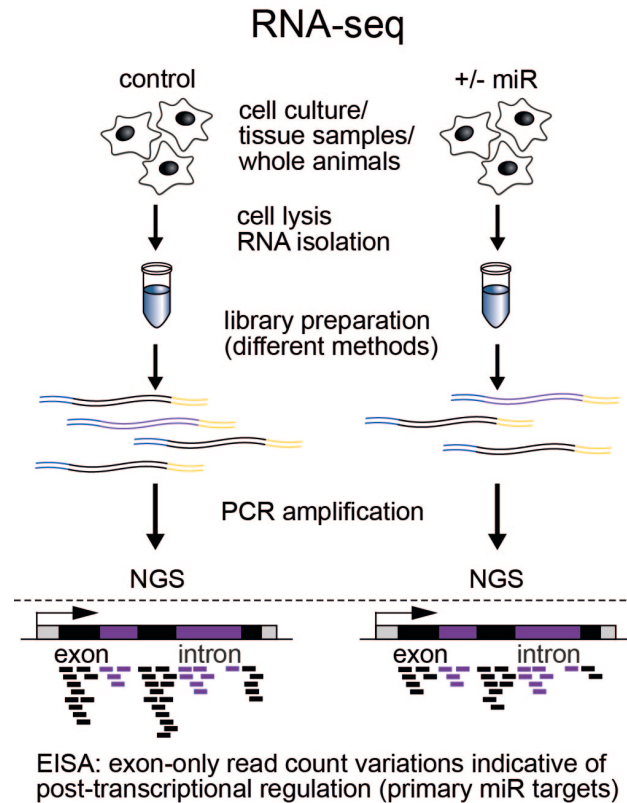


Figure 1.4 | RNAseq for profiling miRNA targets. (Adapted from Steinkraus *et al.* 2016). RNAseq experiments provide short reads across the entire transcriptome mapping to various gene structures (5' UTR, 3' UTR, exon and introns). RNAseq is routinely performed in WT and miRNA loss-of-function (LOF) or gain-of-function (GOF) backgrounds. Putative miRNA targets are expected to anti-correlate with the miRNA of interest. Exon-intron-split analysis (EISA) is capable to differentiate primary from secondary targets through considering intron read counts which are not expected to change in primary targets.

Although the consensus appears to suggest that miRNAs regulate their targets through mRNA destabilisation as opposed to translational inhibition, physiological effects are mediated on the protein level. Furthermore, as mRNA and protein levels frequently do not correlate, it has been of interest to assess the cell's protein fraction for putative miRNA targets. A host of technologies has been developed, ranging from the relatively straightforward two-dimensional gel electrophoresis (2D-DIGE), to chemically labelled proteomics (iTRAQ and SILAC) to label-free approaches (Fig. 1.5).

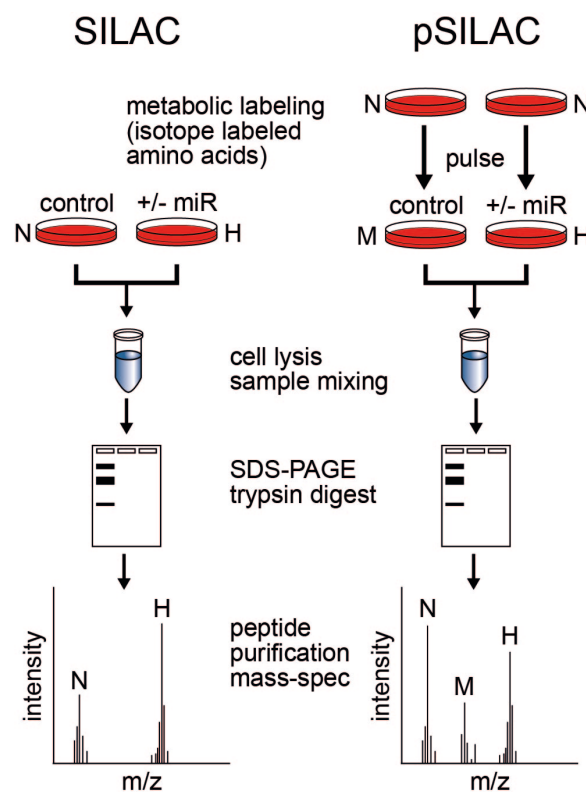


Figure 1.5 | Example SILAC approach for miRNA target discovery. (Adapted from Steinkraus *et al.* 2016). The experimental condition is incubated with a heavy isotope (H) while the controls contain the normal isotope. Cells are lysed and samples are mixed before separation on SDS-PAGE gel followed by trypsin digest. Differences in protein expression is read out on the basis of the ratio between isotopes. In pulsed (p)SILAC, both a medium (m) and a heavy (h) isotope are added to both experimental and control conditions and only newly synthesized proteins are analysed.

While tremendous progress is underway in this field, alternative modes of assessing physical RNA association with the translation machinery have been developed. As it has been particularly desirable to distinguish between mRNA destabilisation and translational inhibition, advances in ribosome profiling have been exploited in the miRNA field. A pioneering approach was that of polysome profiling by the Bartel lab. This method counts the ribosome occupancy (number of RNAs associated to at least one ribosome) and the ribosome density (how many ribosomes are on average bound per 100 bp RNA). In these experiments cycloheximide is used to arrest translation, followed by cell lysis and sucrose gradient centrifugation, to then separate and compare the ribosome bound and unbound RNA fraction *via* microarray (Hendrickson et al., 2009).

An improvement to polysome profiling is ribosome profiling (Fig. 1.6). In this case, RNAseq replaces the microarray readout, which enables higher resolution analysis of the binding data (nucleotide grade). A critical improvement to the protocol is the addition of an RNase I digestion step that degrades all RNA not protected by ribosomes. cDNA is reverse transcribed from protected RNA fragments and then sequenced. The readout generates two informative metrics: the number of ribosomes bound to individual RNA species as well as the precise base pair location that is bound by the ribosome. This technique has fuelled the miRNA mechanism of action debate, as profiling ribosomes in developing zebrafish embryos revealed that maternal target RNAs are frequently repressed on the translational level before they are degraded as part of the maternal-to-zygotic transition (Bazzini et al., 2012).

Ribosome profiling

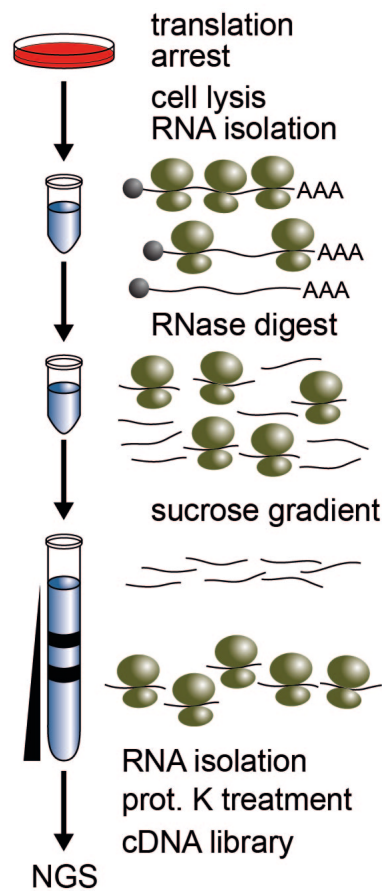


Figure 1.6 | Ribosome profiling for miRNA target discovery. (Adapted from Steinkraus *et al.* 2016). Translation is arrested before cells are lysed. RNase digests any RNA not protected by ribosomes. A sucrose gradient ultracentrifugation separates ribosome-bound from unbound RNAs. RNA is isolated and released from ribosome complexes *via* proteinase K treatment. cDNA is generated and readout includes the number of ribosomes per RNA as well as nucleotide level resolution of ribosome footprint of RNA species.

While profiling approaches are relatively amenable and powerful in the earliest stage of target identification, their major shortcoming is their incapacity to differentiate primary from secondary targets. As a consequence, it is desirable to obtain physical binding data, unequivocally confirming direct miRNA-target association.

1.1.2.3.2 RIP and CLIP

The hallmark of miRNA-targeting is the miRISC assembly on target transcripts, which forms a ternary complex that includes miRNA, mRNA as well as AGO proteins. Hence, each of these three components offers a potential pulldown target. The majority of these implementations rely on RNA immunoprecipitation (RIP) of miRISC protein components, foremost AGO which is in direct contact with miRNA and their target RNAs. Both miRNA and longer mRNA can be co-immunoprecipitated and analyzed *via* qRT-PCR, microarray (RIP-Chip) and RNAseq (RIP-seq) (Beitzinger et al., 2007; Easow et al., 2007; Fasanaro et al., 2009; Jovanovic et al., 2012). Additional components of the RISC complex have also been utilised. For instance, pulldowns of GFP-tagged AIN-1 and AIN-2, the GW182 orthologs in *C. elegans*, have uncovered miRNA-target interactions in specific tissues or whole animals (Kudlow et al., 2012; Zhang et al., 2007; 2009). Although RIP strategies provided valuable insights into miRNA targeting, it has become apparent that these approaches suffer from a high number of false positive hits. This caveat is intrinsic to the experimental framework underlying RIP approaches: in order to maintain protein-RNA interactions throughout the purification process it is essential to perform the assay under low-stringency conditions. While non-denaturing conditions preserve physiological protein-RNA interactions, they also permit negatively charged RNA and positively charged proteins to randomly associate post-lysis. This occurrence has been explicitly demonstrated in at least two studies and warrants the exclusive use of high-stringency protocol conditions (Mili and Steitz, 2004; Riley et al., 2012).

A fundamental prerequisite for the use of denaturing conditions is the prior stabilisation of *bona fide* physiological interactions *in vivo*. Based on this consideration, the Darnell lab has pioneered a high-throughput crosslinking based approach termed *Argonaute high-throughput sequencing of RNAs isolated by crosslinking and immunoprecipitation* (Ago-HITS-CLIP, Fig. 1.7), which relies on

tissue/cell UV irradiation prior to cell lysis (Chi et al., 2009).

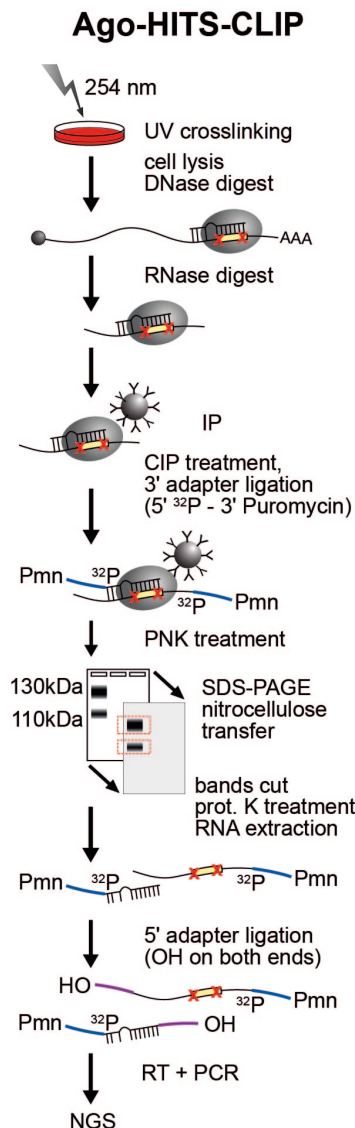


Figure 1.7 | AGO-HITS-CLIP is a powerful tool for miRNA target discovery. (Adapted from Steinkraus *et al.* 2016) Cells are crosslinked *in vivo* with UV irradiation at 254 nm. Cells are lysed and DNA is digested. This is followed by an RNase digest to remove unprotected RNA fragments. Immunoprecipitation (IP) of ternary complexes using an anti-AGO antibody is performed next. RNA is treated with calf intestinal phosphatase (CIP). Radiolabelled 5' adaptors that include a puromycin moiety are ligated. Samples are treated with polynucleotide kinase (PNK) and libraries are separated into small RNA and longer mRNA using SDS-PAGE. The bands are transferred to a nitrocellulose membrane, excised, proteinase K treated and purified. 5' adaptors are ligated leaving OH groups on both ends. Reverse transcription (RT) generates cDNA which is amplified and sequenced.

This irradiation step crosslinks RNA bases to aromatic amino acids of proteins if they are in direct contact (zero distance). In brief, following cell lysis, a limited RNase treatment digests unprotected (non-crosslinked) RNA fragments and miRNA::target:AGO complexes are immunoprecipitated with an anti-AGO antibody. The isolated small and mRNA fractions are radiolabelled and separated through SDS-PAGE and transferred to a nitrocellulose membrane. AGO proteins are then digested by a proteinase K step. Small RNAs and longer mRNAs are isolated

separately from the membrane, reverse transcribed into cDNA and sequenced as two separate libraries. The libraries are computationally deconvoluted and integrated into complementary miRNA::target matches. The incorporation of crosslink induced mutation sites (CIMS) aids defining the AGO footprint on target RNAs.

Since its first implementation in 2009, AGO-HITS-CLIP has been used to elucidate a number of miRNA targetomes across multiple tissues and organisms, including mouse brain (Chi et al., 2009), *C. elegans* larvae (Zisoulis et al., 2010) as well as multiple human and mouse cell lines (Leung et al., 2011). Besides confirming previously defined miRNA-target axes, the AGO-HITS-CLIP data raised two interesting points. First, miRNA binding sites are not confined to 3' UTRs of mRNAs but can also reside in the ORF as well as the 5' UTR. However, the efficacy of these sites has been controversial, as in worms deficient for ALG-1 (Argonaute-like protein 1), mRNAs with ORF target sites were not found to be upregulated (Zisoulis et al., 2010). Conversely, in murine stem cells ORF MREs appear to confer similar repression with 3' UTR sites (Leung et al., 2011). An integrative meta-analysis has recently concluded that MREs residing in the ORF display a propensity for mediating silencing *via* translational repression, in contrast to ordinary sites in the 3' UTR which preferentially cause mRNA destabilisation (Hausser et al., 2013). Second, HITS-CLIP data challenged the concept of canonical, seed-obligate target pairing as the studies revealed a host of imperfect miRNA-target interaction. At first, the analysis was restricted to seed-paired interactions and ignored so-called 'orphan clusters'. However, revisiting these clusters for *miR-124* binding revealed a striking pre-dominant motif: a G insertion at position 6 (Chi et al., 2012). If miRNA nucleotides 2-6 are perfectly paired, a G-bulge can form in the target site and generate a cognate target site for *miR-124*. Nucleotide position 6 of the miRNA has since been called the 'pivot' nucleotide. At least 15% of AGO-HITS-CLIP identified sites display this

alternative mode of targeting, suggesting that it is widespread and also evolutionary conserved (Chi et al., 2012). A study investigating the *miR-155* targetome in *C. elegans* revealed that ~40% of its target sites comprised a single nucleotide mismatch in the seed region (Loeb et al., 2012). Albeit widespread, the repressive efficacy of these sites have been shown to be lower than canonical sites with one study denominating them as completely ineffective (Agarwal et al., 2015). A further occurring mode of target recognition was noted in *C. elegans*, where seed paring happens despite a G:U wobble base pair (Zisoulis et al., 2010).

To enhance the nucleotide resolution detection of miRNA::target interactions, a modified HITS-CLIP protocol, *Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation* (PAR-CLIP), was developed (Fig. 1.8) (Hafner et al., 2010).

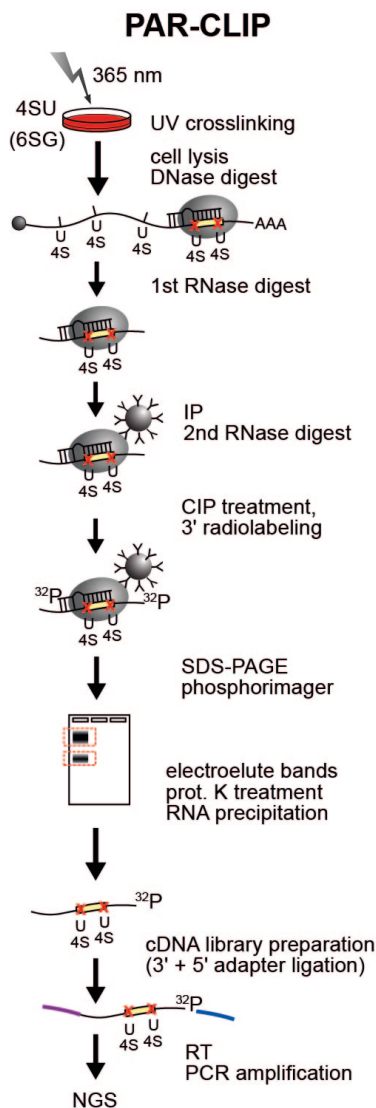


Figure 1.8 | PAR-CLIP. (Adapted from Steinkraus et al. 2016). Modified ribonucleotides (4SU or 6SG) are administered to cells that incorporate them into RNAs. *In vivo* crosslinking is performed with UV irradiation at 365 nm. Cells are lysed and DNA is digested. This is followed by a first RNase digest to remove longer unprotected RNA fragments. Immunoprecipitation (IP) of ternary complexes using an anti-AGO antibody is performed next, followed by a second RNase treatment. RNA is incubated with calf intestinal phosphatase (CIP) and 3' radiolabeled. Samples are separated into small RNA and longer mRNA using SDS-PAGE. The bands are electroeluted, proteinase K treated and RNA precipitated. 5' and 3' adaptors are ligated and reverse transcription (RT) generates cDNA which is amplified and sequenced.

The hallmark of PAR-CLIP is the administration of modified ribonucleosides 4-thiouridine (4SU) and 6-thioguanosine (6SG) to cells in culture, which are readily incorporated into nascent RNAs. 4SU and 6SG display improved crosslinking efficiency when irradiated at 365 nm. Importantly, during cDNA synthesis, reverse transcriptase makes an error opposite nucleotide 4SU; instead of dA the RT

incorporates dG. This phenomenon allows the tracing of these positions by the presence of conversion 'hot spots'. A study performed in HEK293 cells used this approach to assess miRNA::target interactions and uncovered that only 50% of miRNA-mRNA binding appears to occur in the 3' UTR whilst the other 50% takes place in the ORF and 5' UTR. Confirming previous reports, it was noted that the 3' UTR sites mediate mRNA destabilisation whilst the ORF and 5' UTR sites were only minimally effective in doing so. This observation was surprising as all site types displayed a comparable architecture in terms of sequence and structure. The study further investigated non-canonical target sites and found a relatively low incidence with as little as 6% incorporating a mismatch or G:U or U:U wobble in their seed. Analogous to the concept of 3' compensatory sites, this investigation noted that complementarity at positions 13-15 of the miRNA can moderately reinforce miRNA-mRNA binding (Hafner et al., 2010).

Although powerful in identifying exact locations of RBP-RNA interactions, PAR-CLIP is limited by its application to only a limited sample range. Although modified ribonucleosides can be readily adopted by cell cultures as well as consumed by *C. elegans*, they cannot be delivered to most tissues and animals. Further complicating the use of such chemistries, it has been noted that 4SU can cause toxicity and stress responses (Burger et al., 2013; Kemény-Beke et al., 2006).

One technical phenomenon noted for both HITS-CLIP and PAR-CLIP is the incomplete removal of protein 'stumps' from crosslinked RNAs despite extensive proteinase K treatments. Residual amino acids can remain associated to nucleic acids and obstruct the passage of reverse transcriptase. This can either result in truncations, deletions or mutations (crosslinked induced mutation sites = CIMS) of the final cDNA product (Zhang and Darnell, 2011). While CIMS have been exploited for pinpointing RBP motifs, they have also complicated cDNA library preparations. To

overcome this technical limitation, a modified technology was developed and called *individual nucleotide resolution CLIP* (iCLIP) (Fig. 1.9) (Huppertz et al., 2014; König et al., 2010; 2011)

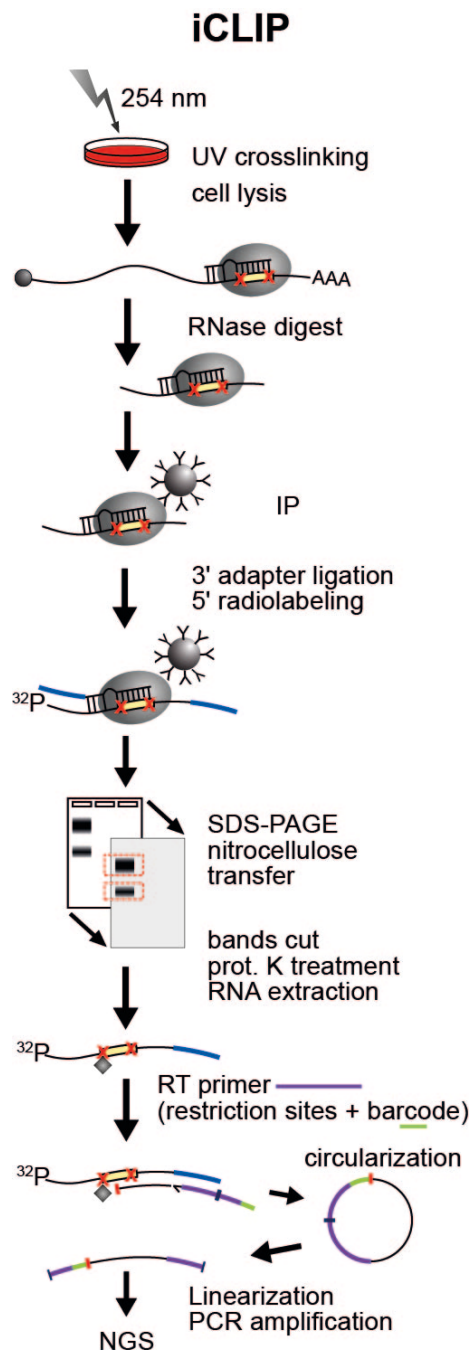


Figure 1.9 | iCLIP. (Adapted from Steinkraus *et al.* 2016). Cells are crosslinked *in vivo* with UV irradiation at 254 nm. Cells are lysed followed by an RNase digest to remove unprotected RNA fragments. Immunoprecipitation (IP) of ternary complexes using an anti-AGO antibody is performed next. 3' adaptors are ligated and 5' ends are radiolabeled. Libraries are separated into small RNA and longer mRNA using SDS-PAGE. The bands are transferred to a nitrocellulose membrane, excised, proteinase K treated and purified. A barcoded reverse transcription (RT) generates cDNA which are circularised to include fragments of all sizes. Circular cDNA is linearised and PCR amplified before sequencing.

To capture incomplete, truncated cDNA fragments, iCLIP includes ligation of an adaptor that enables cDNA circularisation. Circularised fragments are then linearised by endonuclease treatment of the adapter, generating a pool of all cDNA fragments, which include any trimmed, crosslinked sites. The protocol also uses barcoded primers in its library preparation to remove any PCR bias. iCLIP has successfully been applied to map miRNA::target interactions in whole worms with a higher resolution and better sensitivity than previous approaches (Broughton and Pasquinelli, 2013).

To capture RNA-RNA interactions directly and match them with higher fidelity, the Tollervey lab developed an approach called *crosslinking, ligation and sequencing of hybrids* (CLASH) (Fig. 1.10) (Helwak et al., 2013; Kudla et al., 2011). The novelty of this strategy is an intermolecular RNA-RNA ligation step that, when applied to miRNAs, generated miRNA-mRNA chimeras due to their close proximity in the engaged miRISC complex (Helwak et al., 2013). However, only 2% of the sequencing reads appeared to constitute *bona fide* hybrids. Interestingly, non-canonical sites comprising a G:U wobble or a single nucleotide mismatch in the seed were found to be 70% more abundant than classical seed matched sites. Moreover, five stratified MRE classes emerged from the bioinformatics analysis of all sequenced hybrids. Intriguingly, obligate seed pairing was only reported for one class and as much as 18% of hybrids displayed binding entirely outside the seed region.

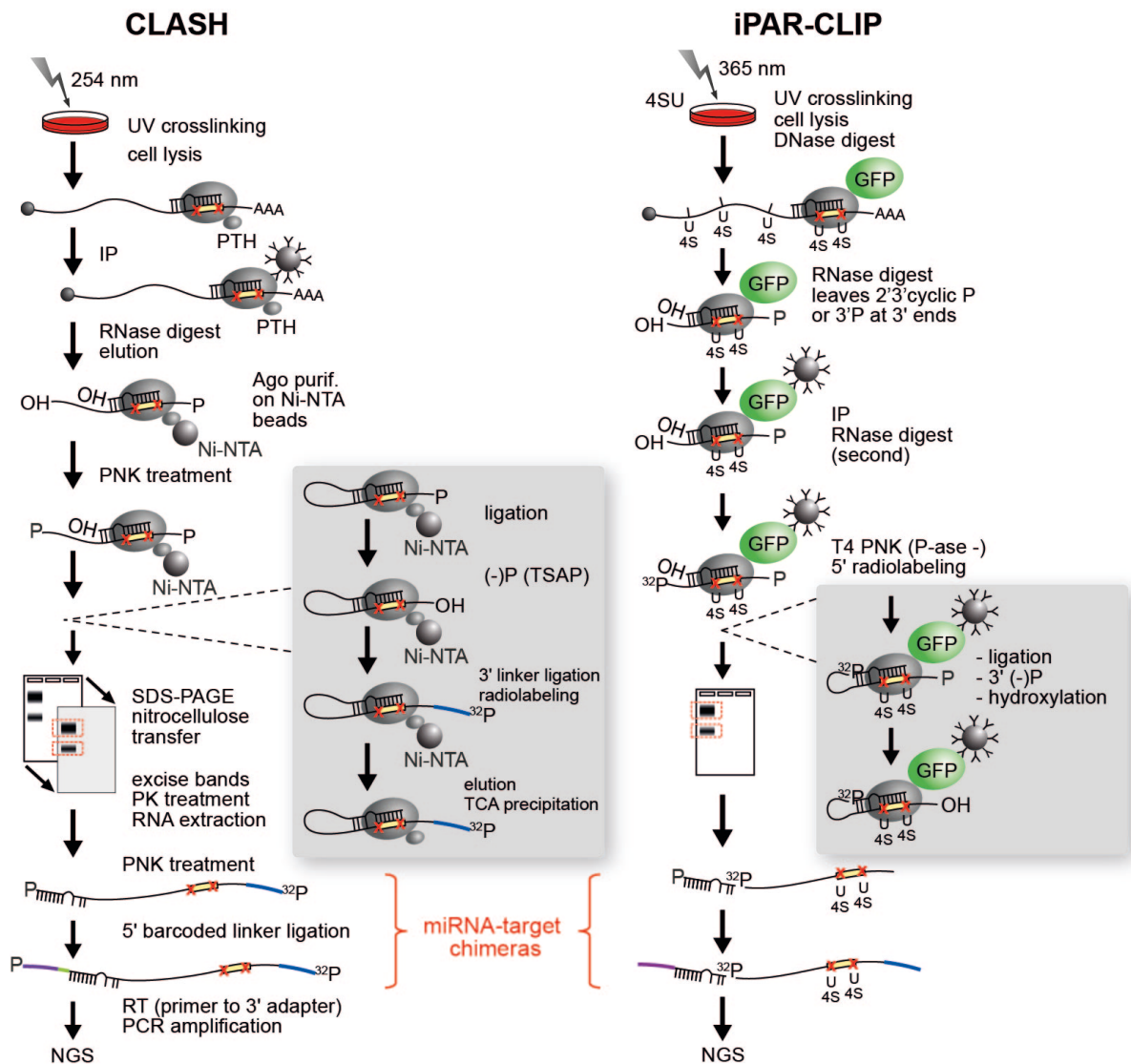


Figure 1.10 | CLASH and iPAR-CLIP. (Adapted from Steinkraus *et al.* 2016). **CLASH:** Cells stably expressing an AGO-PTG fusion are UV crosslinked *in vivo* with 254 nm UV irradiation. Cells are lysed and immunoprecipitated with an anti-AGO antibody. A limited RNase digest is performed and the ternary complex is eluted off the antibody. PTH-AGO is captured by a nickel-charged affinity resin (nitrilotriacetic acid; Ni-NTA) and mRNA is phosphorylated by polynucleotide kinase (PNK). An intramolecular ligation step is performed next which creates a chimera between the miRNA and mRNA (grey box). The chimera is dephosphorylated by thermosensitive alkaline phosphatase (TSAP). 3' linkers are ligated and the phosphate is radiolabeled. The AGO-bound chimera is eluted off the resin and precipitated using trichloroacetic acid (TCA). The complex is purified through SDS-PAGE, bands are excised and proteinase K treated, before RNA is extracted. RNA is treated with polynucleotide kinase (PNK) before 5' barcoded linker is ligated. RT is performed with a primer against the 3' adaptor, prior to PCR amplification and sequencing. **iPAR-CLIP:** This protocol includes 4SU modified ribonucleotides for enhanced crosslinking efficiency at 365 nm UV irradiation as well as a GFP-AGO fusion protein. Post-crosslinking, cells are lysed and DNA is digested.

The protocol next employs an RNase digest that renders 3' ends with 2'3'-cyclic or 3' phosphates. Ternary complexes are immunoprecipitated and an additional RNase digest is performed. The mRNA is radiolabeled through 5' phosphorylation. Intramolecular ligation is performed next, followed by a hydroxylation step (grey box). The 5' adaptor is ligated to the target chimeras and the library is PCR amplified and sequenced.

A fusion between PAR-CLIP and CLASH has recently been developed and termed iPAR-CLIP (Fig. 1.10) (Grosswendt et al., 2014). A ligation step has been added following immunoprecipitation and the subsequent analysis incorporated both T>C mutations at crosslinking sites (PAR-CLIP) but also chimeric reads (CLASH). Both AGO-binding sites and chimeric reads converged on 3' UTR MREs in the majority of target sites. Although 80% of the mapped binding sites were seed matched, these included up to two mismatches and bulges. Positions 2 and 7 were particularly prone to mismatches or bulges. One surprising outcome of the analysis was that chimeras were also found in control samples where no ligase enzyme had been added, indicating that endogenous ligases are capable of forming RNA chimeras. This observation prompted the reanalysis of previous data generated by alternative approaches, which led to the discovery of similar binding patterns in those datasets (Grosswendt et al., 2014).

1.1.2.3.3 RNA Capture

The aforementioned approaches rely on targeting the protein component (AGO) of the miRISC complex to co-purify mRNA and miRNAs. Targeting AGO, the universal agent of every miRNA::target interaction, bears the advantage of capturing the global miRNA interactome as opposed to the interaction profile of just a single target or a single miRNA. However, this also confounds the downstream analysis (see above) and requires a miRNA mutant to validate individual mRNA-miRNA target axes.

Targeting the relevant RNA counterpart, either the miRNA or the mRNA, presents an

attractive alternative for isolating either mRNA or miRNA specific target complexes. Different methodologies have been developed and evaluated to either pull down RISC complexes *via* the miRNA or *via* the target RNA component, and identifying the corresponding binding partners. The prerequisite for both strategies is the same: preservation of the miRNA::target duplex throughout purification steps.

1.1.2.3.3.1 mRNA pull down

A commonly used pulldown method targeting the mRNA fraction, is the incorporation of tags into the primary mRNA sequence that can facilitate the isolation process (Fig. 1.11).

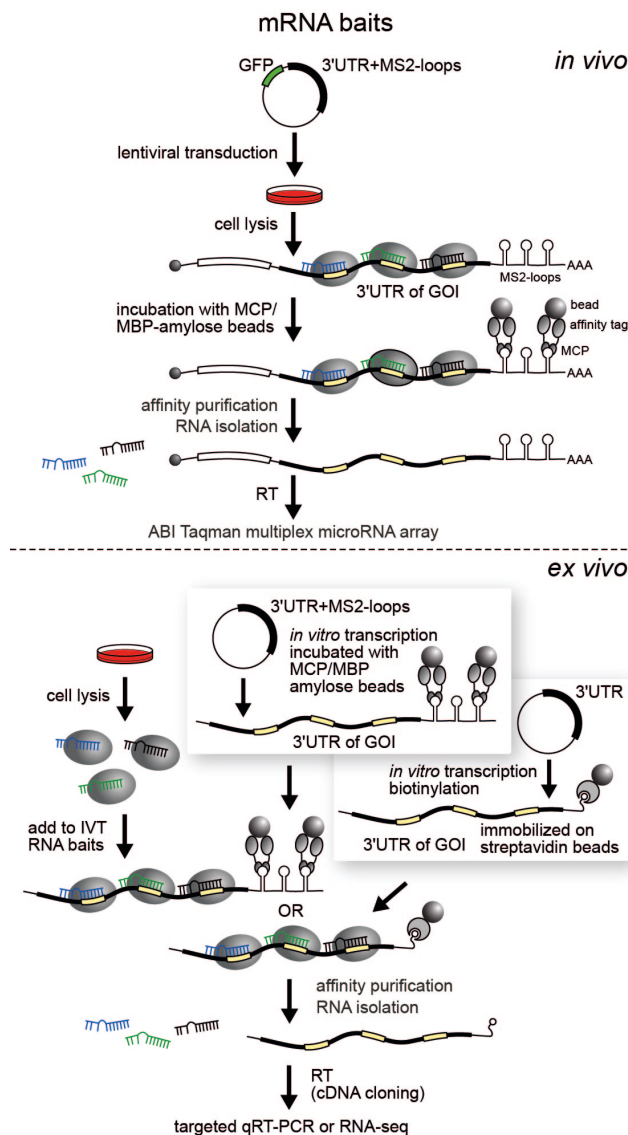


Figure 1.11 | Artificial RNA bait approaches. (Adapted from Steinkraus *et al.* 2016). *In vivo* RNA capture relies on transfecting an artificial 3' UTR-MS2 loop chimera into cells. After transient expression, cells are lysed and the lysate is incubated with MCP-amylose beads which specifically capture the MS2 containing RNA construct. The small RNAs are eluted and identified *via* Taqman miRNA qPCR or NGS. The *ex vivo* counterpart works through incubating native cell lysate with IVT RNA constructs including RNA segments of interest that either contain a MS2 loop tag, or are biotinylated. The former requires isolation through MCP, the latter through streptavidin beads.

One commonly employed tag is the MS2 stem-loop sequence originating from MS2 bacteriophage (Bertrand et al., 1998). The MS2 loop is recognised with high affinity by the MS2 coat protein (MCP). This framework has been applied to discover miRNAs interacting with *Hand2* in rat cardiomyocytes (Vo et al., 2010). In this study, three MS2 loops were inserted into the 3' UTR of *Hand2*, which was then captured using an immobilised MCP-maltose binding protein (MBP) on an amylose column. miRNAs associated to *Hand2* were analysed by a multiplex Taqman miRNA array which validated a previously reported regulatory function for *miR-1*, but also uncovered *miR-133a* as a new regulator of *Hand2*.

An analogous approach, called MS2-TRAP, has been applied to miRNAs interacting with *lincRNA-p21* in murine embryonic fibroblast (MEF) cells (Yoon et al., 2012). In this protocol, 24 repeats of the MS2 sequence were fused to the 3' end of *lincRNA-p21*, and a MCP-GST column was used to capture the RNA from lysates. Four of five previously predicted miRNAs were detected in the eluate fraction through RT-qPCR assay.

A complementary technique that resembles RNA chromatography was developed and termed *miRNA mapping by in vitro RNA affinity purification* (miTRAP) (Fig. 1.11) (Braun et al., 2014). In this assay, four MS2 repeats were fused to *in vitro* transcribed 3' UTRs and attached to an amylose column *via* MCP-MBP fusion proteins. U2OS and HEK293 cell lysates were applied *ex vivo* to columns fitted with the 3' UTRs of either *ZEB2* or *MYC*. This study identified 18 miRNAs bound to *MYC*, eight of which had been previously predicted. This relatively high number of regulatory miRNAs prompted the question whether *MYC* is subject to a combinatorial regulation by multiple miRNAs.

Obviating the need for MS2 tags, IVT RNAs can be made amenable for pulldown through incorporation of biotinylated nucleotides (Shi et al., 2013). Biotinylated IVT 3'

UTRs of *TCF8* and *Smad4* have thereby been immobilised on streptavidin-coated magnetic beads. Incubating purified miRNA pools from mouse liver tissue or human bronchial epithelial cells with the coated beads revealed multiple miRNAs binding both *TCF8* and *Smad4*. Although the physiological relevance of these interactions remains doubtful, the assay is readily applicable to a wide array of samples (including archived patient samples).

The goal of RNA bait approaches is to capture RNAs in their native state, with minimal external modifications. To address this experimentally, Hassan *et al.* developed a technique termed miR-CATCH which relies on biotinylated oligonucleotide antisense purification of native mRNA (Fig. 1.12) (Hassan *et al.*, 2013).

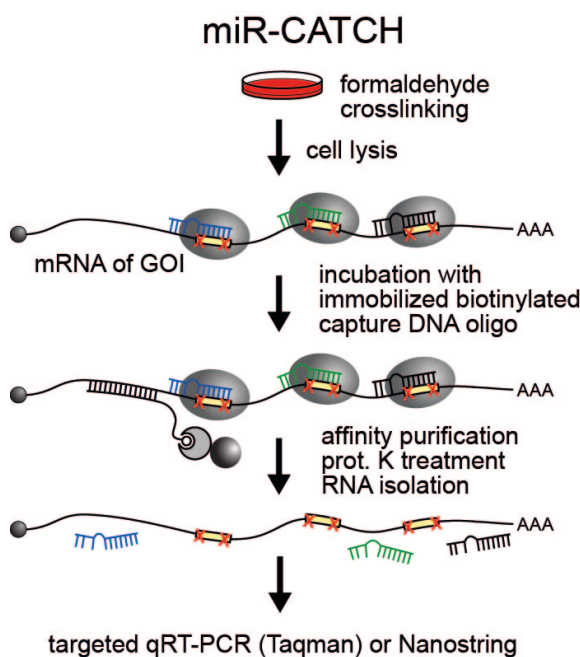


Figure 1.12 | miR-CATCH captures native mRNAs. (Adapted from Steinkraus *et al.* 2016). This technique employs antisense mediated affinity purification of previously stabilised (formaldehyde crosslinking) mRNA-miRNA-AGO ternary complexes. Biotinylated antisense oligonucleotides are designed against a target mRNA of interest and incubated with crosslinked cell lysate under stringent conditions. Streptavidin-coated beads are then employed to extract the antisense oligonucleotides with their intended target. Post-crosslink reversal, co-purifying smallRNAs are analysed by Nanostring and Taqman qRT-PCR.

Using this approach, the authors isolated three mRNAs (*alpha-1 antitrypsin (AAT)*, *interleukin-8*, and *leucoprotease*) from crosslinked cell lysate and identified co-purified short RNAs by Nanostring profiling and qPCR analysis.

1.1.2.3.3.2 miRNA pull down

The aforementioned techniques all target the mRNA component of the miRISC complex with the goal of identifying associated miRNAs. The complementary approach, isolating miRISC *via* the miRNA, has also been proposed and employed with moderate success. For instance, biotinylated miRNAs have been employed in cell culture studies (Lal et al., 2011, Li et al., 2012). Following the same principle, *labelled miRNA pull-down* (LAMP) employs digoxigenin (DIG) for the purpose of labelling pre-miRNAs. Labeled pre-miRNAs are incubated with cell lysate and processed *ex vivo*. Isolating mature miRNAs on anti-DIG beads and subsequent profiling of target RNAs revealed *Hand2* as a target of *miR-1* in zebrafish (Hsu et al., 2009).

An update of these approaches is termed *tandem affinity purification of miRNA target mRNAs* (TAP-Tar) and relies on a two-step pulldown (Nonne et al., 2010). First, FLAG-tagged Ago complexes are affinity purified. Second, biotinylated miRNAs are isolated through streptavidin coated magnetic beads. Isolating biotinylated miR-20a co-purified *E2F2* mRNA.

An alternative technology called miR-TRAP makes use of biotinylated miRNAs that have been appended with photoreactive psoralen (Pso) moieties (Baigude et al., 2012). Irradiation at 360 nm activates Pso to become a RNA-RNA crosslinker which can form a covalent bond to uridine present in target mRNAs. Importantly, crosslinking can be reversed by 254 nm light irradiation. Fusing nucleotide 9 of the *miRNA-29a* to a Pso group, the authors found two predicted targets to be enriched in the eluate.

An update to miR-TRAP has been developed under the name miR-CLIP, which re- includes an immunoprecipitation step and employs optimised biotin and Pso configurations (Fig. 1.13) (Imig et al., 2015). Employing this setup, 600 putative mRNA targets, including lincRNA H19 were captured by a biotinylated *Pso-miR-106a*.

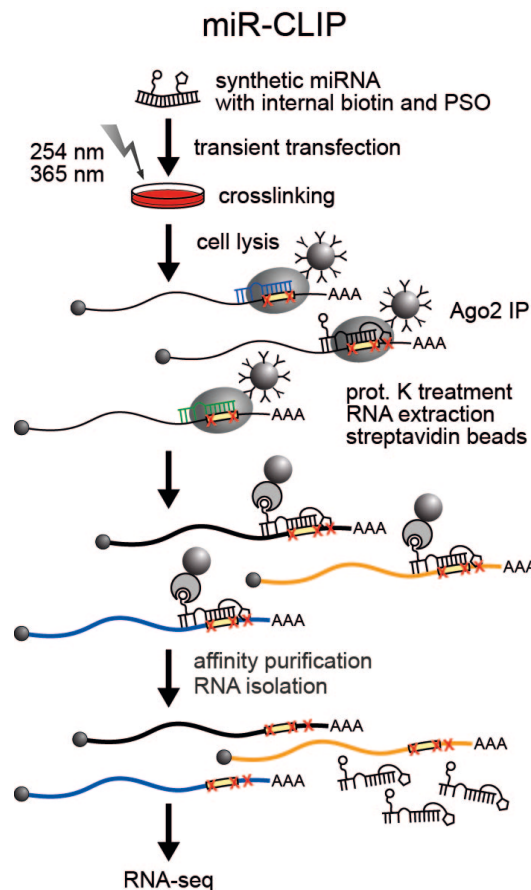


Figure 1.13 | mRNA target isolation by miR-CLIP. (Adapted from Steinkraus *et al.* 2016). The miRNAs of interest is modified to include biotinylated nucleotides as well as psoralen groups, a photoinducible, reversible RNA-RNA crosslinker. This artificial miRNA is subsequently transfected into cells and UV crosslinked to its targets. After AGO immunoprecipitation and proteinase K digest, RNA is extracted. Small RNAs and their binding partners are affinity purified, followed by photoinduced release of the long RNA fraction. Any co-purifying mRNAs are identified by RNAseq.

Strikingly, breakthroughs in our knowledge of miRNAs have almost exclusively relied on novel technologies (Steinkraus et al., 2016). Examples include: NGS studies which enabled the development of target identification algorithms, AGO-HITS-CLIP uncovering non-canonical sites, and most recently, genome engineering based approaches to specifically delete MREs *in situ* and assess the consequences on target abundance (Bassett et al., 2014).

A state-of-the-art miRNA-target identification pipeline should hence rely on the following technologies: If a miRNA of interest is known, the first step should be a bioinformatic target prediction using both miRanda and TargetScan, providing a first overview of potential targets at no cost. To refine the target search, unbiased transcriptomics using high-resolution RNAseq should be employed, however, this requires a miRNA loss-of-function mutant. To distinguish primary (direct miRNA binding) from secondary targets (upstream regulator of target RNA is bound by miRNA) AGO-HITS-CLIP can generate *in vivo* binding maps of AGO-miRNA-mRNA ternary complexes, although this protocol remains technically challenging.

Perhaps the greatest technical challenge in the miRNA target identification field is to study miRNA::target interactions under physiological conditions, under cellular homeostasis and physiological miRNA-mRNA stoichiometry. Before presenting a novel technique that meets this requirement, I will introduce next the background to my second project: elucidating the regulatory mechanisms underlying *bicoid* RNA translational quiescent state *in vivo*.

1.2 Elucidating the regulatory mechanisms underlying *bicoid* RNA translational quiescent state *in vivo*

The polarisation of oocytes in species ranging from flies to mice precedes the onset of development and is required for establishing the embryonic axes through precise localisation and post-transcriptional regulation of maternally transcribed mRNAs. In *Drosophila*, establishment of anterior-posterior axial polarity is dependent on *bicoid* RNA, which is produced by nurse cells and deposited in the oocyte (Berleth et al., 1988; Driever and Nüsslein-Volhard, 1988). In the oocyte, the mRNA is anchored at the anterior boundary where it remains translationally quiescent until early embryogenesis. Following egg activation Bicoid protein is expressed and fulfils its role as a transcription factor. Although an extensively studied molecule, the molecular mechanisms underlying *bcd* translation quiescent state during oogenesis remain largely elusive. Interestingly, super-resolution microscopy and ultra-structural analysis revealed that *bcd* mRNA is enriched in P-bodies (Weil et al., 2012), which function as sites of mRNA metabolism, and have been intimately linked to the miRNA pathway (Eulalio et al., 2007a; Liu et al., 2005). Furthermore, the onset of *bicoid* mRNA translation is marked by an elongation of its poly-A tail at egg activation (Sallés et al., 1994; Eichhorn et al., 2016). Regulation of poly-A length is also one of the mechanisms underlying miRNA mediated repression and has more recently been attributed to piRNAs (abundant short non-coding RNAs in the germline) (Eulalio et al., 2009; Rouget et al., 2010). Taking these two lines of evidence into consideration, we hypothesise that short non-coding RNAs regulate *bicoid* translational quiescence during oogenesis.

1.2.1 Synopsis of *Drosophila* oogenesis and early embryogenesis

D. melanogaster is part of the *Drosophila* genus which belongs to the Diptera order of higher flies. *Drosophila* are holometabolous insects that undertake complete metamorphosis through four key developmental stages – embryo, larva, pupa, and imago (adult) – over the short time course of 10 days. Their oogenesis is of a polytrophic meroistic nature, which includes a body of supporting nurse cells growing alongside the developing oocyte forming a cyst called the ‘egg chamber’.

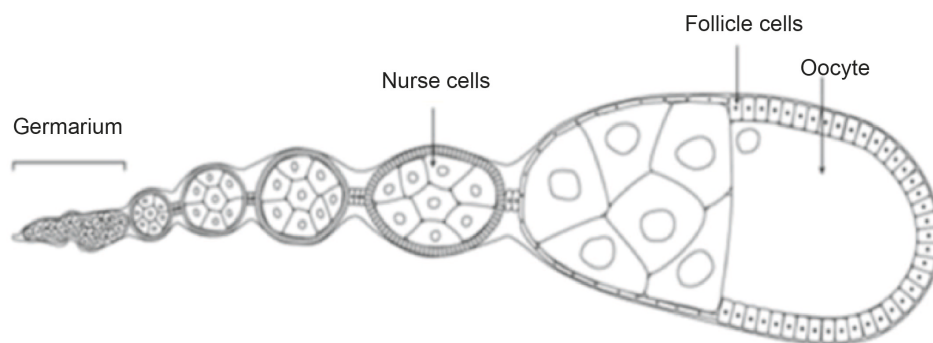


Figure 1.14 | *Drosophila* oogenesis. (Adapted from McLaughlin et al. 2015). Every ovary houses 16-20 ovarioles which are tubes of connected egg chambers. Starting in the germarium, a 16 cell egg chamber undergoes 14 morphologically distinct stages of development, culminating in the production of a mature egg (not shown).

The reproductive organs of the female fly comprise a pair of ovaries each consisting of 16-20 independent ovarioles, which in turn consist of tubular strings of 6-8 egg chambers connected by somatic stalk cells (Fig. 1.14). Morphologically, each ovariole can be subdivided into three zones: the terminal filament (TF), the germarium and the vitellarium. The TF is made up of 8-9 somatic cells that link the germarium to the encompassing ovariole sheath and dictate the direction in which

the ovariole will develop (Godt and Laski, 1995). The germarium can be further subdivided into four specialised compartments (1, 2a, 2b and 3) that coordinate the initial steps of oogenesis, including the division and differentiation of germline and somatic stem cells, the assembly of the egg chamber and the transit into the vitellarium. The vitellarium, finally, represents the largest part of the ovariole in which the oocyte goes through 14 distinct morphological stages resulting in maturation, oocyte polarisation, vitellogenesis and choriogenesis (McLaughlin and Bratu, 2015; Spradling, 1993).

Within region 1, the very anterior end of the germarium, resides a stem cell that divides asymmetrically to give rise to a renewed germline stem cell and a differentiated cystoblast. In region 2a and 2b the cystoblast carries out four additional cell divisions with only partial cytokinesis, generating a cyst of 16 interconnected sister cells called cystocytes (King et al., 1968). Out of these 16 cells, 2 proto-oocytes are selected, one of which becomes the oocyte as a result of the asymmetric accumulation of the fusomes, cytoplasmic structures containing endoplasmic reticulum, cytoskeletal proteins and microtubules (Huynh and St Johnston, 2004). The remaining 15 cells become nurse cells serving essential nutrients, RNA and protein, as well as organelles to the developing oocyte through cytoplasmic bridges termed 'ring canals'. The germarium also houses somatic stem cells which produce an array of differentiated follicle epithelial cells that form a monolayer around the 16 germ cells. Following complete follicle cell ensheathment in region 3, the cyst buds out of the germarium and enters the vitellarium as a *stage 1* egg chamber.

During early oogenesis, nurse cells undergo 10-12 endocycles (endoreplication) in which DNA synthesis (S) occurs with gap phases (G) but without mitosis and cytokinesis (Dej and Spradling, 1999). By this means, the resultant polyploid cells 'supercharge' their RNA and protein synthesis capacity by generating additional DNA

template for transcription, in aid of optimal nutrient supply for the oocyte (Nordman and Orr-Weaver, 2012). Follicle cells also perform 6-8 endocycles. However, from stage 10b onwards they switch to gene amplification cycles (E/A cycles), leading to the multiplication of four key genomic loci producing the chorion (egg shell) as well as the vitelline membrane required to protect the mature egg (Claycomb and Orr-Weaver, 2005; Klusza and Deng, 2011). Transcription in the oocyte itself is largely quiescent as its nucleus condenses into a compact structure called the karyosome (Huynh and St Johnston, 2004).

While the polarity of the egg chamber is already established in the germarium, the oocyte itself becomes polarised through a series of events occurring during mid-oogenesis (*stage 6-7*). Key to this process is the posterior localisation of *gurken* (*grk*) RNA and the local expression of its transforming growth factor alpha like (TGF α) protein product, which signals a 'posterior' fate to the Torpedo receptor in the proximally overlying follicle cells (Neuman-Silberberg and Schupbach, 1993). Once defined 'posterior', these somatic cells signal back to the oocyte causing a repolarisation of the cytoskeleton (González-Reyes et al., 1995; Ruohola et al., 1991). The microtubule-organizing centers (MTOC) at the posterior end of the oocyte break down and reform again at the anterior and lateral margins of the oocyte, so that the 'minus' end lies at the anterior edge and the 'plus' end projects towards the posterior (Januschke et al., 2006; Theurkauf et al., 1992). This newly arranged microtubule network also attracts maternal RNAs *bicoid* (*bcd*) and *oskar* (*osk*) to the anterior cortex and posterior ends of the oocyte, respectively. Shuttling *bcd* and *osk* to opposite poles manifests the hallmark of anterioposterior (AP) polarisation, as the mRNAs remain localised throughout the remainder of oogenesis and initial embryogenesis. The reorganisation of the microtubule networks also results in the migration of the nucleus to the dorsoanterior (DA) corner of the oocyte, which

coincides with *grk* RNA appearing as a cap around the nucleus. The 'dorsal' fate is reinforced through Grk protein signalling to the overlying follicle cells (González-Reyes et al., 1995; Januschke et al., 2002; Zhao et al., 2012). Hereupon, the follicle cells begin to secrete yolk into the oocyte leading to an increase in oocyte size. Various reformations of the follicle cell population follow, culminating in a stage 10B oocyte where the oocyte now assumes 50% volume of the entire egg chamber (Fig. 1.15). At this stage, the oocyte is completely surrounded by follicle cells which produce a protective vitelline and chorion seal. Concomitantly, the cytoskeleton is reassembled engendering ooplasmic 'streaming' (Gutzeit, 1986). In the nurse cells, filamentous actin (F-actin) is rearranged into bundles that project from the plasma towards the nuclear membrane. At stage 11 the polyploid nurse cell nuclei become permeabilised and a myosin-mediated contraction 'dumps' the entire nurse cell contents into the oocyte (Buszczak and Cooley, 2000). The nurse cells undergo apoptosis and are removed from the egg chamber leaving behind a stage 14 oocyte, representing the mature egg.

The oocyte enters meiosis, however, already arrests at prophase I while still in the germarium. At stage 13 the oocyte advances to a second checkpoint at metaphase I (Stetina and Orr-Weaver, 2011). The passage of the egg into the oviduct triggers the completion of meiosis eliciting egg activation, which immediately precedes fertilisation and oviposition (Mahowald et al., 1983). Following fertilisation, the embryonic nucleus undergoes 13 rapid synchronised mitotic divisions, without cytokinesis, giving rise to ~8000 nuclei in a single cell (syncytial blastoderm). About 1.5 h after fertilisation these nuclei migrate to the periphery and membrane invaginations cellularise the embryo (cellular blastoderm). Protein morphogen gradients such as Bicoid and Nanos establish embryo segmentation coincidental with

the maternal-to-zygotic transition of genetic control from the maternal system to the new organism.

1.2.2 Maternal mRNA Localisation and Translational Repression

In the 1980s Christiane Nüsslein-Vollhard and Eric Wieschaus uncovered the genetic basis underlying *Drosophila* embryonic development through mutagenesis screens (Nüsslein-Vollhard and Wieschaus, 1980). One particular phenomenon they noticed was that patterning of the future embryo was determined in particular by the localisation of maternal effector transcripts to distinct locations in the developing oocyte (Fig. 1.15) (Frohnhofer and Nüsslein-Vollhard, 1986).

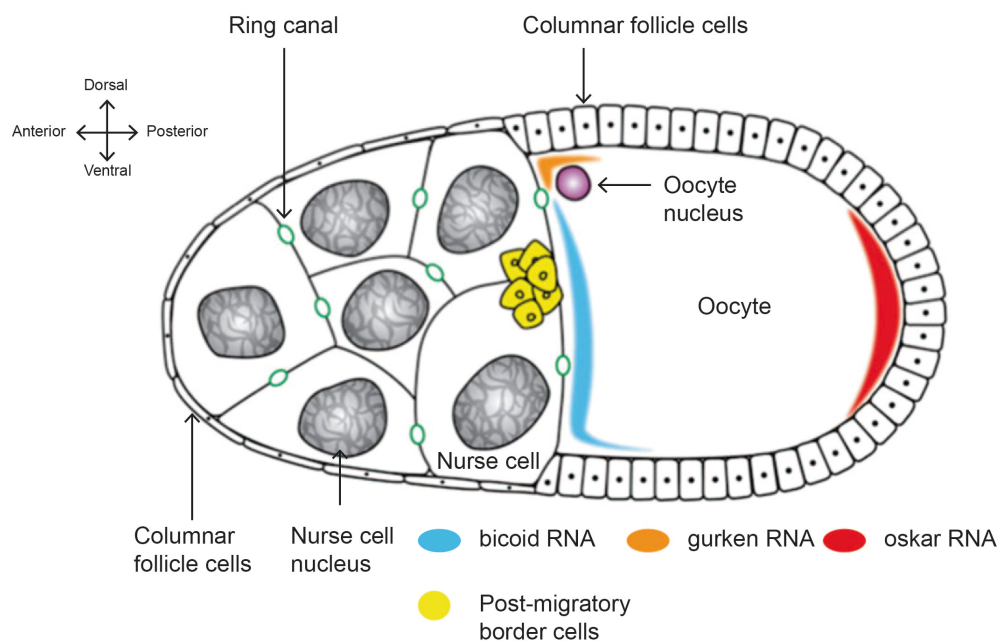


Figure 1.15 | Maternal mRNAs are localised to opposite poles in the developing oocyte. (Adapted from McLaughlin *et al.* 2015). A stage 10B oocyte is already polarised in terms of anteriorposterior (localisation of anterior *bicoid* and posterior *oskar* mRNA) and dorsoventral (localisation of dorsal *gurken* mRNA) body axes.

For example, *bcd* mRNA specifies anterior head and thorax structures through accumulating at the anterior of the oocyte, where after fertilisation it forms a

morphogenic transcription factor sitting atop a gene hierarchy that in turn subdivides the embryo into smaller functional segments (Driever and Nüsslein-Volhard, 1988; St Johnston et al., 1989). To the contrary, *osk* mRNA localises to the future posterior end of the oocyte where its protein product specifies the development of germ plasm. Conversely, around *stage 7 gurken (grk)* mRNA assembles at the dorsoanterior periphery of the egg chamber, forming a cap around the nucleus, and specifying the dorsal fate at this location (González-Reyes and St Johnston, 1998).

The localisation of *nanos (nos)* during oogenesis is somewhat less obvious, however, the RNA visibly accumulates at the posterior after *stage 10* (Forrest and Gavis, 2003). Unlike for *bcd*, this process is not mediated by active transport but rather by trapping of *nos* that reaches the posterior and the degradation of randomly diffused RNA. Strikingly, Nos is expressed already in the nurse cells and transported as a protein into the oocyte, where it appears uniformly up to stage 12. Interestingly, at stage 13 Nos protein is only detectable at the very posterior tip, suggesting translational repression underlies the localised expression of Nos protein (Bergsten and Gavis, 1999). This is mediated *via* a bipartite 90 nt translational control element (TCE) located in the 3' UTR (Gavis et al., 1996). Stem loop III of the TCE is recognised by Glorund, a *Drosophila* hnRNP F/H homologue, which represses unlocalised *nos* during oogenesis (Kalifa et al., 2006). During early embryogenesis, unlocalised *nos* is repressed by Smaug which binds stem loop II of the TCE (Smibert et al., 1996). The predominant mechanism for this was considered Smaug mediated recruitment of Cup culminating in cap-dependent translational repression (Nelson et al., 2004). An unexpected addition to this model was the discovery that piwi-interacting RNAs (piRNAs), guide the deadenylation complex CCR4-NOT specifically to *nos* mRNA *via* base pairing (Rouget et al., 2010). This marked the first report of piRNA involvement in translational regulation, as they had previously been

considered defenders of the germ line against transposons (Brennecke et al., 2007; 2008; Malone et al., 2009). Furthermore, it added a function for small RNAs in the dynamic regulation of maternal RNAs which had previously only been linked to miRNA mediated decay as part of the MZT (Bushati et al., 2008).

All four maternal mRNAs, *bcd*, *grk*, *osk* and *nos*, are precisely regulated in time and space with regard to their localisation and translational control (Fig. 1.15) (Kugler and Lasko, 2012). While the translational control of *osk*, *grk* and *nos* has been shown to rely on repressor proteins such as Bruno, Squid and Smaug, the translational silencing of *bcd* has remained, at least in large parts, elusive (Cáceres and Nilson, 2009; Kelley, 1993; Kim-Ha et al., 1995; Snee et al., 2014; Webster et al., 1997). This is surprising since *bcd* reveals one of the most striking translational profiles: the mRNA is abundantly produced during oogenesis yet completely repressed right until egg activation (oviposition). It is only then that the mRNA becomes competent for translation.

1.2.2.1 The localisation of *bicoid* mRNA

Like other maternal mRNAs, transcription of *bcd* is initiated in nurse cells during *stage 4 and 5* of oogenesis by the transcription factor Serendipity delta (Payre et al., 1994). The first *bcd* RNA is detected in the oocyte at stage 4-6 (Macdonald and Kerr, 1997; Payre et al., 1994; St Johnston et al., 1989). *Bcd* RNPs are shuttled from the nurse cells into the oocyte by means of the MT/dynein/Bic-D/Egl pathway and become localised at the anterior margin (Berleth et al., 1988; Bullock and Ish-Horowicz, 2001). The molecular motor freighting *bcd* through active transport is Dynein, which tracks along microtubules towards the 'minus' end (Schnorrer et al., 2000; Weil et al., 2006). Exuperantia (Exu), Swallow (Swa) and Staufén (Stau) are

three essential proteins for *bcd* localisation, acting at different time points (St Johnston et al., 1989).

Exu is required only in the earliest stages of *bcd* transport, when *bcd* is shuttled from the nurse cells into the oocyte (Cha et al., 2001). Swa was considered the direct adaptor between *bcd* and its molecular motor Dynein (Arn et al., 2003; Schnorrer et al., 2000). However, super-resolution light and immunoelectron microscopy revealed Swa is transported independently of *bcd*, does not co-localise, and instead associates with the membrane and prepares the local actin cytoskeleton for *bcd* anchoring (Weil et al., 2006; 2010). The actin regulator Hu li tai shao (*Hts*) has also been implicated in facilitating this process (Pokrywka et al., 2014), underpinning the concept that localizing an RNA can certainly rely on indirect factors (Weil, 2014). Another essential factor is Stau, a double stranded RNA binding protein, which assists *bcd* localisation in late oogenesis and early embryogenesis (Ramos et al., 2000; St Johnston et al., 1991).

The aforementioned *trans* factors interact with various *cis* elements in the *bcd* 3' UTR. The *bcd* gene has five isoforms, and a 3' UTR length of ~800 nt. The earliest study to establish a functional link between the 3' UTR and *bcd* transport deleted a 625 nt segment that would completely abrogate anterior localisation (Macdonald and Struhl, 1988). A more refined analysis of this large deletion in 50 bp increments uncovered the so-called *bcd* localisation element 1 (BLE1) (Fig. 1.16) which is both necessary and sufficient for transit into the developing oocyte and initial assembly at the anterior border (Macdonald et al., 1993).

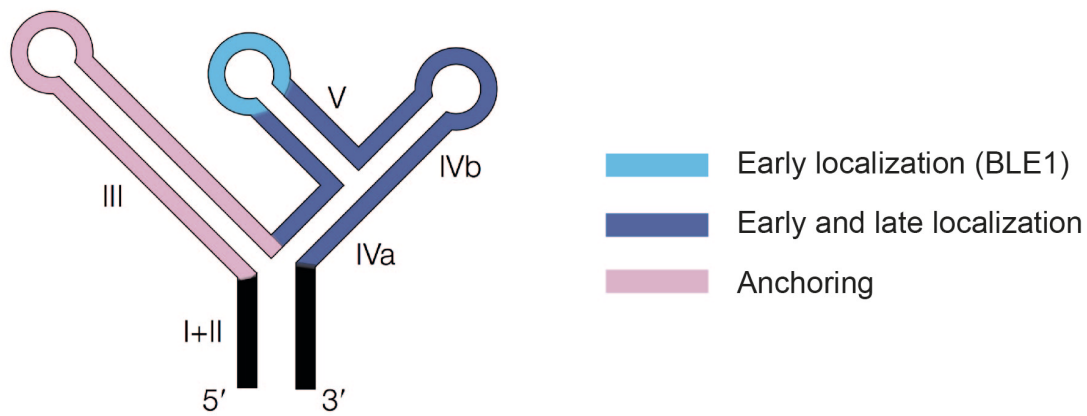


Figure 1.16 | *bcd* mRNA secondary structure. (Adapted from Jansen *et al.* 2001). The *bcd* 3' UTR exhibits a predicted structure including five domains, notably stem loop V which houses the *bcd* localisation element 1 (BLE1) at the very terminal end of the loop. Stem loop IV is redundant in early localisation and functions also in later transport. Stem loop III dimers interact during the anchoring process.

Remarkably, a single point mutation in this region, G>U4496 abrogates earliest localisation around stage 4-5 (Macdonald and Kerr, 1998). BLE1 resides in the terminal region of stem loop V, one of several putative secondary structure elements of the 3' UTR (Fig. 1.16). Localisation during later stages of oogenesis is also mediated through a redundant interaction between the transport machinery and stem loop IV, and is not dependent on the BLE (Macdonald and Kerr, 1997). *Bcd* anchoring during late oogenesis requires stem loop III, which forms a dimer *via* intermolecular base pairing in a 'hand-by-arm' fashion (as opposed to 'kissing loops') (Brunel and Ehresmann, 2004; Ferrandon *et al.*, 1994; Wagner *et al.*, 2001).

While a significant effort has uncovered many of the *cis* elements and *trans* interacting factors underlying the precise localisation of *bcd* mRNA, our knowledge regarding its remarkable translational profile is still limited. Although *bcd* RNA is present in the egg chamber from stage 6 onwards, its protein product is first detected in early embryos, indicating the mRNA is robustly silenced throughout most of oogenesis, corresponding to a time period of several days (Driever and Nüsslein-

Volhard, 1988). While mRNAs are generally silenced during transit, the extensive, absolute quiescence post-localisation of *bcd* is unique in the oocyte.

1.2.2.2 Translational regulation of *bicoid* mRNA

Bicoid protein production is only licensed after egg activation, a process that includes vitelline membrane crosslinking, and the resumption and completion of meiosis (Endow and Komma, 1997). Interestingly in insects, egg activation occurs prior to and independently of fertilisation, when the mature egg ovulates (ovary exit) and enters the significantly narrower lateral oviduct (Doane, 1960; Heifetz et al., 2001; Mahowald et al., 1983). Changes in the egg environment, including ion concentration as well as mechanical stimuli, have been proposed to elicit egg activation. Interestingly, mechanical force alone, exerted *via* straining of the dorsal appendages of the egg, was reported sufficient to engender activation (Endow and Komma, 1997). However, this observation has recently been questioned (York-Andersen et al., 2015). Furthermore, submerging eggs into hypotonic buffer led to swelling of the hitherto desiccative eggs and subsequent activation (Mahowald et al., 1983; Page and Orr-Weaver, 1997). It was thence shown that osmotic and hydrostatic pressure induced egg activation in a manner that required external calcium influx, an important second messenger that had already been implicated in other species including vertebrates (Horner and Wolfner, 2008). This result dovetailed with the observation that eggs cannot resume meiosis and translate *bcd* in mutants of the calypsin encoding gene *sarah* (*sra*) (Horner et al., 2006; Takeo et al., 2006). A clearer picture of this process was recently provided by York-Andersen *et al.* who employed injected and genetically encoded calcium sensors to uncover a single transient calcium wave that propagates rapidly from the posterior through the entire egg to mark the onset of activation (York-Andersen et al., 2015).

Of particular interest is how egg activation terminates *bcd*'s translational quiescence. One mechanism employed by various organisms to post-transcriptionally regulate maternal effector RNAs is the level of cytoplasmic adenylation (Gebauer et al., 1994; Sheets et al., 1995) (Eichhorn et al., 2016; Subtelny et al., 2014). Consistently, it was revealed that the effector step of *bcd* translational activation is manifested by an elongation of its poly(A) tail (Sallés et al., 1994). Silent ovarian *bcd* has a poly(A) tail of ~70 adenosines (A), which becomes elongated to ~140 A's competent for translation in 0.5 h to 2 h old embryos, and then shortened again after 2 h demarcating the degradation of maternal mRNAs as part of the maternal-to-zygotic (MZT) transition. Injection of a truncated *bcd* transcript lacking 537 nt of its 3' UTR was only able to rescue the *bcd*^{E1} anterior-posterior defect mutant if the construct had an *in vitro* added poly(A) tail of 150-200 nt. Poly(A) tails of ~60 nt were insufficient to rescue the morphogenetic defects, corroborating that an extended poly(A) tail is indispensable for *bcd* translation.

A separate line of experiments involving *nanos* (*nos*) mRNA, a maternal effector for posterior patterning, further underpinned the importance of cytoplasmic polyadenylation for *bcd* translation. The *bcd* 3' UTR contains a *nanos response element* (NRE) which prevents embryonic Bicoid translation in the posterior zone of the embryo (Wharton and Struhl, 1991). In *Bicaudal-D* (*Bic-D*) mutants *nos* localisation is impaired and Nanos protein is scattered throughout the embryo (Wharton and Struhl, 1989). Consequently, Bicoid protein is repressed in *Bic-D* mutants and an analysis of the *bcd* poly(A) tail length from wild type and *Bic-D* mutant flies revealed a shortened poly(A) tail in the latter (Wharton and Struhl, 1991). In 2008, Benoit and colleagues identified the responsible effector protein, a polyadenylase GLD-2 homologue, encoded by *wispy* (*wisp*) (Benoit et al., 2008). Wisp activity is required for *bcd* poly(A) tail elongation from ~70 to ~150 A's.

Interestingly, although other mRNAs like *osk*, *nos*, and *cyclin B* also undergo cytoplasmic polyadenylation, they display a major poly(A) elongation already during oogenesis, particularly between *stage 10 and 14* (c.f. *bcd* polyadenylation which only starts after egg activation) (Benoit et al., 2008). These events are abolished in Wisp mutants suggesting that Wisp is active during oogenesis and responsible for the polyadenylation of multiple maternal RNAs. It remains unclear why *bcd* is not a substrate for Wisp during this time. This observations suggests the involvement of an additional, *bcd* specific, silencing mechanism.

The study also concurred with a previously proposed link between meiotic arrest and translational control. It had been noticed that *bcd* polyadenylation is impeded in females mutant for *cortex* (*cort*), an activator of the anaphase-promoting complex (APC), and *grauzone* (*grau*), the transcription factor required for *cort* expression (Chen et al., 2000; Lieberfarb et al., 1996; Page and Orr-Weaver, 1996). Neither of these mutants was able to complete meiosis as the transition from metaphase to anaphase was impeded. The link to polyadenylation, and thus *bcd* translation, was provided by a model which suggested that in *cort* mutants APC-mediated proteolysis of cyclin components of maturation/mitotic promoting factor (MPF) is impaired (Chu et al., 2001). In wild type embryos, MPF cyclically represses poly(A) polymerase (PAP) by hyperphosphorylation. However, in *cort* mutants MPF is constitutively active, hence repressing PAP continuously and preventing polyadenylation. Like *bcd*, *cort* is a direct target of Wisp which explains why *wispy* mutant embryos phenocopy *cort* mutants with regard to their arrested development (Benoit et al., 2008).

A recent study employing integrated RNAseq, ribosome profiling and poly(A) tail length profiling by sequencing (PAL-seq) underscored a global correlation between poly(A) tail length and translational efficiency in *Drosophila* oogenesis and early embryogenesis (Eichhorn et al., 2016). The authors measured the correlation of

thousands of mRNA poly(A) tails *versus* their translational efficiency across the oocyte-to-embryo transition. Interestingly, they noticed that *bcd* undergoes an increase in translational efficiency with regards to its poly(A) tail elongation which exceeds the ratios of the vast majority of other RNAs. This suggested additional regulation, either through a *bcd* specific repressor preventing translation during oogenesis, or a translational activator. Due to the comparable poly(A) tail length of *bcd* to other efficiently translated mRNAs at stage 14, the authors speculated that a *bcd* specific repressor scenario is more likely.

In a second line of experiments conducted in a *wispy* mutant background, when cytoplasmic polyadenylation is impaired, the authors noted that the relative poly(A) tail lengths and the corresponding TEs of many RNAs were preserved (although absolute tail-length was lower). This suggested that a large fraction of translational efficiency during the oocyte-to-embryo transition is orchestrated by selective deadenylation, a majority of which could be attributed to the action of PAN GU kinase (PNG) and Smaug (Eichhorn et al., 2016).

While cytoplasmic polyadenylation of *bcd* is necessary for translation, it was shown that it is not sufficient. Verotti and colleagues uncoupled cytoplasmic polyadenylation from translational activation through insertion of a 63 nt *bcd* ORF fragment into its 5' UTR, generating a backfold termed *bcd*-AS (Verrotti et al., 1999). Upon injection into embryos, although *bcd*-AS was polyadenylated in an ordinary manner, it remained incompetent for translation *in vivo*. Presumably this structural modification perturbed the interaction between the 5' and 3' mRNA ends and thus disrupted the closed loop formation required for translation. This conclusion was later supported by an *in vivo* experiment where overexpression of poly(A) polymerase (PAP) resulted in an extended *bcd* poly(A) tail (~250 A) which did not lead to protein production during

oogenesis, although levels of embryonic Bicoid appeared increased (Juge et al., 2002).

While the signal for egg activation, and by extension *bcd* translation, has been revealed to be a calcium wave, it remains unclear how *bcd* is so robustly silenced throughout oogenesis. The first mechanistic cue to this longstanding biological conundrum stemmed from a super-resolution and immunoelectron microscopy study (Weil et al., 2012). The authors detected *bcd* particles in the center of processing bodies (P-bodies), which are mRNA storage and degradation hubs void of ribosomes. Zooming in on the dorsoanterior margin, where *bcd* partially overlaps with actively translated *grk*, Weil and colleagues noticed *grk* particles accumulated at the periphery of P-bodies, which led them to propose a paradigm for translational control based on the differential localisation of these maternal RNAs within P-bodies. Consistently, following egg activation, when *bcd* is released from its focal anterior anchor (Weil et al., 2008), it is no longer detected in P-bodies, which imminently afterwards disperse (York-Andersen et al., 2015). Since P-bodies and the miRNA core machinery are intimately linked, and P-bodies frequently form in response to miRNA-dependent silencing (Eulalio et al., 2007a), it is likely that miRNAs constitute one possible mechanism through which *bcd* translation is arrested throughout oogenesis.

Furthermore, Gamberi and colleagues noticed failure of *bcd* clearance and protracted Bicoid protein expression (beyond 4h after egg lay) in early embryos derived from *Pumilio* (*Pum*) null mutant female flies (Gamberi et al., 2002). Interestingly, Pum1 and Pum2 have both been intimately linked to the miRNA pathway through inducing a conformational change in the RNA that exposes important MREs in the oncogenes *p27* and *e2f3* mRNA for miRNA binding (Kedde et al., 2010; Miles et al., 2012). This suggests that *bcd* may also be regulated by miRNAs during the MZT, in line with

many other maternal RNAs which are cleared by the zygotically transcribed miR-309 cluster (Bushati et al., 2008).

To summarise, the following lines of evidence suggest miRNAs are mediating *bcd* silencing: *i*) several conserved miRNAs are predicted to target *bcd* (see chapter 5) *ii*) *bcd* mRNA, while silent, is associated with P-bodies (which contain miRNA effector proteins), *iii*) *bcd* requires polyadenylation before translation (antagonising poly(A) tails is one of the main mechanisms miRNAs employ to silence their targets) *iv*) *nos*, another maternal mRNA, is deadenylated by the piRNA pathway (suggesting that small RNAs can regulate maternal RNAs), *v*) *bcd* clearance is impaired in mutants of Pumilio (regulator of miRNA silencing).

1.3 Thesis Aims

The overarching theme of this thesis is the study of miRNA::target interactions for a better understanding of the cellular pathways that are subject to miRNA-mediated regulation. In particular, this investigation aims to study miRNA::target interactions in their physiological state *in situ*. To this end, this thesis details the development and implementation of *in vivo* detection of miRNA-target interaction events (miR-TRACE), a novel miRNA::target identification platform that permits the interrogation of miRNA response elements (MRE) under native conditions. I aim to demonstrate the reliability and versatility of this technology on the primary target network of the conserved miRNA miR-184. Furthermore, employing the same principles of probing for miRNA::target interactions under physiological conditions, this thesis investigates the translational quiescence of *bicoid* mRNA during oogenesis. In particular, this study interrogates the involvement of short regulatory RNAs, foremost miRNAs, as a potential mechanism that underlies *bcd* silencing. The principle aims of this work are hence:

- **To develop a novel technology for high-throughput discovery of miRNA::target interactions in their native state *in vivo*, with minimal external intervention**
- **To employ miR-TRACE to uncover novel MREs and target networks that have hitherto not been amenable to high-throughput investigation**
- **To decipher the molecular mechanism underlying *bcd* RNA silencing throughout oogenesis**

Chapter 2

This chapter presents the reagents and experimental protocols used throughout this Thesis.

Chapter 3

This chapter introduces *in vivo* detection of miRNA-target interaction events (miR-TRACE) as a novel technique for miRNA::target identification *in situ*. I present the iterative optimisation of this technology and demonstrate the step-wise pipeline employed to generate miR-TRACE scores, a new metric to validate known and identify new MREs. Specifically, I deliver proof-of-concept that miR-TRACE can identify previously defined MREs with high confidence.

Chapter 4

In this chapter I explore the versatility and scalability of miR-TRACE through *i)* scaling the assay for parallel analysis of multiple MREs and *ii)* developing a transcriptome-wide readout. This culminates in the development of high-throughput *in vivo* detection of miRNA-target interactions events by massively parallel sequencing (miR-TRACEseq). miR-TRACEseq is applied to the primary target network of the conserved miRNA miR-184 and validates previously reported MREs as well uncovers novel MREs.

Chapter 5

This chapter investigates the translational quiescence of *bicoid* RNA during *Drosophila* oogenesis, with a specific focus on miRNAs. I employ a range of techniques, including small RNA profiling and an exploratory miR-TRACE experiment spanning the whole *bicoid* 3' UTR. Furthermore, I develop an unbiased *bicoid* RNA

pulldown followed by small RNAseq, and computational analysis, to create the first connection between *bicoid* silencing and miRNAs.

Chapter 6

I provide a general conclusion of my work as well as a future outlook for my research.

Chapter 7

This chapter is the bibliography of work referenced throughout this Thesis.

Chapter 2: Materials and Methods

2.1 Buffers & Solutions

Table 2.1: Solutions and buffers used throughout this Thesis

Buffer	Composition	Notes
Lysis Buffer	20 mM Tris.HCl (pH7.5) 500 mM LiCl 0.5 % LiDS (wt/vol) 1 mM EDTA 12.5 mM DTT	Store up to 3 months at 4 °C To 10 ml buffer add fresh one table cOmplete Proteinase Inhibitor cocktail (Roche) and 10 µl of 20U/µl Supersasin (Ambion)
Buffer 1	20 mM Tris.HCl (pH7.5) 500 mM LiCl 0.1 % LiDS (wt/vol) 1 mM EDTA 5 mM DTT	Store up to 3 months at 4 °C Add fresh 1 µl/ml buffer Supersasin (20U/µl) (Ambion)
Buffer 2	20 mM Tris.HCl (pH7.5) 500 mM LiCl 1 mM EDTA 5 mM DTT	Store up to 3 months at 4 °C Add fresh 1 µl/ml buffer Supersasin (20U/µl) (Ambion)
Buffer 3	20 mM Tris.HCl (pH7.5) 2 mM LiCl 1 mM EDTA 5 mM DTT	Store up to 3 months at 4 °C Add fresh 1 µl/ml buffer Supersasin (20U/µl) (Ambion)
Elution Buffer	10 mM Tris.HCl (pH 7.5) 1 mM EDTA	Store up to 1 year at room temperature Add fresh 1 µl/ml buffer Supersasin (Ambion)
5x Proteinase K Buffer	50 mM Tris.HCl (pH 7.5) 750 mM NaCl 1% (wt/vol) SDS 50 mM EDTA 2.5 mM DTT 25 mM CaCl ₂	Store up to 1 year at room temperature
TRACE Lysis Buffer	50 mM HEPES (pH 7.5) 75 mM NaCl 0.1 mM EGTA 0.5 % <i>N</i> -laurylsarcosine 0.1 % sodium deoxycholate	Store up to 6 months at 4 °C

NEB RNase H Buffer	50 mM Tris-HCl 75 mM KCl 3 mM MgCl ₂ 10 mM DTT	
TRACE Proteinase K Buffer	0.125 M EDTA 0.25 M Tris.HCl (pH 7.5) 2.5 % SDS (w/v) 5 mg/ml Proteinase K (Ambion)	Prepare fresh
Phosphate Buffered Saline (PBS)	137 mM NaCl 2.7 mM KCl 10 mM Na ₂ HPO ₄ 1.8 mM KH ₂ PO ₄	
PBST	1X PBS + 0.01% Triton	
MES	100 mM 2-[N-morpholino]ethane sulfonic acid	
TT	0.25 M Tris.HCl (pH 7.5) 0.01 % Tween®-20	
TE	0.1 M Tris.HCl (pH 7.5) 1mM EDTA	

2.2 *In vivo* detection of miRNA-target interaction events (TRACE)

2.2.1 Material (Lysate) Preparation

2.2.1.1 S2R+ cells

Approx. 5 T175 flasks were grown up for WT and miR-184 KO cells. All cells were pooled into 2 50 ml Falcon tubes and centrifuged for 3 min at 3,000xg. The pellets were resuspended in a total volume of 50 ml PBS and cells were counted in a C-Chip disposable haemocytometer (Neubauer improved, Labtech). 10 million cells were crosslinked in 10 ml 1 % formaldehyde (11586711, Fisher) in PBS for 10 min at 25 °C on a rocker. The cells were spun down for 3 min at 3,000xg and resuspended in 10 ml 0.125 M Glycine in PBS and quenched for 5 min at room temperature on a

rocker. The cells were then moved into the cold room and washed 3 times with ice cold PBS and spun at 3,000xg after every wash. Each 10 million cell pellet was resuspended in 1.1 ml ice cold, SUPERasin and DTT freshly supplemented TRACE Lysis buffer and homogenised using a Qias shredder (2 min centrifugation at 10,000xg). Lysates were snap frozen in 50 µl aliquots in liquid nitrogen and stored at -80 °C until used.

2.2.1.2 Ovary

Female flies (approx. 60) were yeast fattened for 48 h and dissected in ice-cold PBS. The dissected ovaries were washed once with cold PBS and dissociated into individual ovarioles / egg chambers by pipetting up and down 30x with a P-200 pipette. All liquid was removed and the ovaries were crosslinked 10 min at 25 °C with 1 % formaldehyde (11586711, Fisher) in PBS on the rocker. Ovaries were allowed to settle and the crosslinking solution was replaced with the quenching solution (0.125 M Glycine in PBS). The ovaries were incubated for 5 min at 25 °C on a rocker. Next the ovaries were washed three times with ice cold PBS. They were lysed in 1 ml ice cold, SUPERasin and DTT freshly supplemented TRACE Lysis buffer and homogenised using a Dounce homogeniser (10 strokes with pestle A and 10 strokes with pestle B). 50 µl aliquots were snap frozen in liquid nitrogen and stored at -80 °C until used.

2.2.1.3 Embryo

5 medium sized embryo collection cages were set up with grape agar plates. Flies were trained on a light dark cycle (12 h darkness during day and 12 h light during night) to synchronise egg laying and maximise the harvest. At least one pre-lay of 90 min was performed before collections. Embryos were collected for 90 min and

washed into a collection chamber with a paintbrush and water. They were then dechorionated for 2-3 min in bleach (Sainsbury's) and thoroughly washed with water afterwards. The nitrex membrane of the collection chamber was placed into a ~10 ml beaker filled with PBST. The crosslinking solution was prepared in a 50 ml Falcon tube as follows: 9.375 ml PBS + 0.625 ml formaldehyde + 30 ml heptane (Sigma). The nitrex membrane was dried and the embryos were pipetted back onto it and blotted dry. The membrane was then placed into the crosslinking solution, the embryos shaken off and the membrane recovered. The tube was shaken vigorously at room temperature with 10 min by hand. After phase separation the Falcon tube was spun for 1 min at 500xg and the supernatant was removed. Stop solution (125mM Glycine in 0.1% PBST) was added. The tube was shaken for 3-5 min at room temperature. Following another spin at 500xg the supernatant was removed and the embryos were washed 3 x in 50 ml PBST with 500xg spins for 1 min inbetween. Approx. ~100 mg embryos were lysed in 1 ml ice cold, SUPERasin and DTT freshly supplemented TRACE Lysis buffer and homogenised using a Dounce homogeniser (10 strokes with pestle A and 10 strokes with pestle B). 50 µl aliquots were snap frozen in liquid nitrogen and stored at -80 °C until used.

2.2.2 RNase H Mapping

2.2.2.1 solo 'one-ASO:one-MRE'

All lysates were thawed on ice, however, reactions were assembled at room temperature (important to prevent off-target binding). RNA concentration was normalised between WT and KO lysate through dilution of the more concentrated KO lysate. Each reaction was set up in 0.2 PCR tubes (Axygen) using 5 µl lysate, 3.5 µl RNase H buffer (10X), 0.5 µl 100 mM DTT (Fisher, # BP172), 0.15 µl 100 mM MgCl₂, 2 µl RNase H (NEB, # M0297L), 0.25 µl SUPERasin (Ambion, # AM2694), 31.5 µl

H₂O and 2 µl antisense oligonucleotide (100 µM, IDT) or water (–ASO control). The reaction was incubated for 1 h at 37 °C in a thermocycler. Crosslink reversal was performed by adding freshly mixed 2 µl 1 M Tris.HCl, 2 µl 0.5 M EDTA, 2 µl 10 % SDS, 2 µl 20mg/ml Proteinase K (Ambion, # AM2548). The reaction was incubated for 1 h at 50 °C and 30 min at 65 °C. RNA was extracted and converted to cDNA as described in general molecular methods (2.4.2).

2.2.2.2 grouped ‘one-ASO:one-MRE’

ASOs were pooled into groups on the basis of minimal inter-amplicon off-targets. Groupings were derived by a custom algorithm designed to assess all potential binding interactions between ASOs and UTRs. This program was written in R (version 3.2.3) and relied on the string processing library “stringr”, as well as plotting systems “ggplot2” and “d3heatmap”. Each reaction consisted of 5 µl lysate, 3.5 µl RNase H buffer (10X), 0.5 µl 100 mM DTT, 0.15 µl 100 mM MgCl₂, 2 µl RNase H (NEB, # M0297L), 0.25 µl SUPERasin, 13 µl H₂O, and 20 µl 100 µM oligo pool (or H₂O in –ASO control). The reaction was incubated for 1 h at 37 °C in a thermocycler. Crosslink reversal was performed by adding freshly mixed 2 µl 1 M Tris.HCl, 2 µl 0.5 M EDTA, 2 µl 10 % SDS, 2 µl Proteinase K. The reaction was incubated for 1 h at 50 °C and 30 min at 65 °C. RNA was extracted and converted to cDNA as described in general molecular methods (2.4.2).

2.2.2.3 multiplex ‘one-ASO:many-MREs’

A semi-random oligonucleotide was ordered from IDT with the specifications NN-GGACGG-NN (where N = 25% A, 25% B, 25% C, 25%D). Each reaction was set up in 0.2 PCR tubes (Axygen) using 5 µl lysate, 3.5 µl RNase H buffer (10X), 0.5 µl 100 mM DTT (Fisher, # BP172), 0.15 µl 100 mM MgCl₂, 2 µl RNase H (NEB, # M0297L), 0.25 µl SUPERasin (Ambion, # AM2694), 31.5 µl H₂O and 2 µl semi-random ASO (100 µM, IDT) or water (–ASO control). The reaction was incubated for 1 h at 37 °C

in a thermocycler. Crosslink reversal was performed by adding freshly mixed 2 µl 1 M Tris.HCl, 2 µl 0.5 M EDTA, 2 µl 10 % SDS, 2 µl 20mg/ml Proteinase K (Ambion, # AM2548). The reaction was incubated for 1 h at 50 °C and 30 min at 65 °C. RNA was extracted and converted to cDNA as described in general molecular methods (2.4.2).

2.2.2.4 Formulas for RPS and TRACE Score

$$Relative\ Protection\ Score\ (RPS) = \frac{Expression\ Level_{+ASO}}{Expression\ Level_{-ASO}}$$

$$TRACE\ Score = \frac{RPS_{wild\ type}}{RPS_{miRNA\ KO}}$$

2.2.3 TRACeseq - Library preparation and MiSeq

RNA libraries were generated using the NEBNext Ultra Directional RNA Library Kit for Illumina (NEB, # E7420) according to the manufacturer's instructions with the following exceptions. All samples were normalised to 100 ng/µl. 1 µl (5pg/µl) *egfp* was added to 300 ng RNA. First strand cDNA synthesis was performed in the following way. To each 300 ng RNA sample the following was added: 0.5 µl 10 µM *rp49* RT primer, 0.5 µl 10 µM *hsp83* RT primer, 0.5 µl 10 µM *egfp* RT primer, 2 µl 1 µM 6mer seed primer. RNA was fragmented at 94 °C for 15 min and held at 37 °C (to prevent off-target priming at lower temperatures). To each sample (held at 37 °C) the following premix was added: 0.5 µl Murine RNase Inhibitor, 5 µl Actinomycin D and 1 µl ProtoScript II RTase. Following addition of the RT mix, the samples were incubated a further 15 min at 37 °C, followed by 15 min at 42 °C, followed by 15 min at 70 °C, followed by hold at 4°C. The libraries were sequenced using the Illumina MiSeq platform using a 50 cycle v2 kit (25 paired end).

2.2.4 DASH and NextSeq

2.2.4.1 sgRNA generation

We selected the top 6 contaminating ribosomal RNA species and selected suitable target regions using the CRISPR.mit.edu resource of the Zhang lab. Single guide RNAs (sgRNA) were generated by *in vitro transcription* (see 2.4.7 of General Molecular Methods) of a sgRNA double stranded DNA template. Every sgRNA template was generated in an individual PCR reaction that included 5 µl 10 µM CRISPR forward primer (includes T7 promoter and sgRNA sequence), 5 µl 10 µM CRISPR reverse primer (includes backfold structure of tracr RNA), 2 µl 10 mM dNTPs, 20 µl Phusion HF buffer and 1 µl Phusion HF DNA polymerase. Reactions underwent PCR as follows: 98 °C 30 s, 35 cycles of (98 °C 10 s, 60 °C 30 s), and 72 °C 10 min, 4 °C hold. 1 µl was analysed on a 2 % Agarose (Thermo Fisher, # 16500-500), the remainder was gel extracted (Qiagen, # 28706) and measured on a Nanodrop 2000 (Thermo Scientific). sgRNAs for sgRNA 1-18 were pooled in equimolar amounts (400ng of each) and subjected to IVT (T7 MEGASCRIPT, Ambion, # AM1333). Two reactions were set up at room temperature each including 4.25 µl H₂O, 2 µl ATP, 2 µl CTP, 2 µl GTP, 2 µl UTP, 2 µl reaction buffer, 3.75 µl sgRNA template PCR product pool (80 ng/µl). The reaction was incubated for 4 h at 37 °C. 1 µl Turbo DNase was added and the reaction was incubated a further 15 min. The reaction was then stopped through the addition of 115 µl ddH₂O and 15 µl ammonium acetate stop solution. Proteins were extracted through addition of 150 µl phenol:chloroform:isoamyl alcohol (25:24:1) (Invitrogen, # 15593-031) and vortexed for 30 s. Phases were separated through centrifugation at 10,000 x *g* for 3 min at room temperature. The upper RNA phase was transferred to a fresh tube. RNA was precipitated through adding an equal volume (~ 150 µl) of isopropanol (Fisher, # P/7500/17). The reaction was mixed and precipitated at - 20 °C overnight. The RNA

was then centrifuged at 17,000 x *g* for 30 min at 4 °C. The pellet was washed twice with 0.5 ml room temperature 70 % ethanol (VWR, # 20821.330), with centrifugation at 17,000 x *g* for 3 min at 4 °C in-between washed. The remaining liquid was removed and the RNA pellet was dried for 3 min at room temperature. The pellet was resuspended in 30 µl RNase-free ddH₂O (Ambion, # AM9939), its concentration measured by Nanodrop and frozen in 20 µg aliquots at -80 °C. RNA integrity was assessed on 1 % agarose gel. 2 µl RNA were mixed with 8 µl loading dye (NEB 2x RNA loading dye, # B0363S), denatured for 2 min at 90 °C and quick cooled on ice for 2 min.

2.2.4.2 DASH reaction, qPCR and NEXTseq

All 12 libraries (3x WT - ASO, 3x WT + ASO, 3x KO - ASO, 3x KO + AS) were diluted to 10ng/µl. RT-qPCR established the levels of *gfp* spike-in present in each library which was used to normalise and pool libraries in an equimolar amount. 10 µl library pool (100 ng) was added to 4 µl Cas9 10X buffer, 4.32 µl (Cas9 protein, 46 µM, NEB) 8.51 µl sgRNA pool (3.5 µg/µl), and 3.17 µl H₂O. The mixture was incubated at 37 °C for 2 h. Next, 5 µl Proteinase K (Ambion) was directly added to the reaction and incubated a further 15 min at 37 °C. The library was purified using AMPure XP beads (according to manufacturer's protocol) (Beckman Coulter, # A63880) at 0.9X volume and eluted in 20 µl. The DNA was then concentrated using the Zymo DNA Clean & Concentrator kit (Zymo Research, # D4003S). The library was then re-amplified for 10 cycles using P5 and P7 primers (Illumina). The reaction contained 1 µl library, 1 µl primer (10 µM) and 8 µl Phusion Polymerase Mix (NEB) and was cycled at 98 °C for 2 min, then 10 cycles of 98°C for 10 s, 65 °C 30 s, 72 °C 40s, followed by 5 min final elongation at 72 °C. The re-amplified library was then purified again on AMPure XP beads (0.9X volume) and measured on a Qubit Spectrophotometer (16 ng/µl). A no Cas9 control was carried out in parallel under identical conditions. The sequencing

was repeated on the Illumina NextSeq platform with the NextSeq® 500/550 High Output Kit v2 (75 cycles).

2.2.5 TRACeseq bioinformatics and GLM

2.2.5.1 NGS analysis

All FASTA files containing raw paired end sequencing reads were first merged and subsequently aligned to the *Drosophila* dm3 genome with the Bowtie2 algorithm (Langmead and Salzberg, 2012). This generated BAM files (Li et al., 2009). To visualise the alignment results, BAM files were then sorted and indexed with Samtools (Li et al., 2009) and loaded using IGV or UCSC (Karolchik et al., 2011; Robinson et al., 2011). To prevent potential PCR bias, PCR duplicates were removed from BAM files. In order to extract the 6mer motif read counts across the whole transcriptome, we first identified all the chromosome coordinates of the 6mer (CCGTCC) motif in both strands and generated a corresponding a BED file. We then used Bedtools function coverageBED to extract all read counts coverlapping the 6mer motif (Quinlan and Hall, 2010).

2.2.5.2 Logistic regression model

2.2.5.2.1 True positives

We used two sources to generate a set of true positives. The first one was a multiple dosage miR-184 overexpression (OE) and corresponding mRNA transcriptome data (Spearman anti-correlation). The second one was experimentally validated MREs by the genERA approach (Qianxin Wu, *manuscript under consideration*). The following steps were taken:

- 1) The majority of true positives were derived from an integrated anti-correlation analysis between miR-184 levels and transcriptome-wide gene expression levels across 6 bins with increasing miR-184 level (KO<WT<OE1<OE2<OE3<OE4). We

summed the total reads per RefSeq ID including 5'UTR, CDS and 3'UTR, but not introns since these are not part of mature RNA, using the R function 'featureCounts' from the 'Subread' package (Liao et al., 2014). These were converted to FPKM and a Spearman's correlation was calculated using the R function 'cor.test' with method="spearman", correlating FPKM and miR-184 levels as measured by qPCR. We retained only those with negative Spearman's coefficients ($\rho < 0$) with a false discovery rate (FDR) ≤ 0.001 . This resulted in 284 tightly anti-correlated miR-184-target RNA pairs, which exhibited a median Spearman coefficient of $\rho = -0.86$ (range -0.82 to -0.95). 2) We identified 17918 occurrences of the 6mer motif across the *Drosophila* genome (Dm3), which was proportionately reduced to 9510 when considering only those located in transcribed strands (RefSeq). The exclusion of transcripts with low read counts in either the WT and KO no oligo samples (i.e. ≤ 5 mean count) further lowered this set to 1431 sites displaying the 6mer seed motif. 3) Of the 284 miR-184 anti-correlated target pairs, only 87 carried the 6mer motif, indicating a transcriptome wide gene expression perturbation following microRNA overexpression. Among them, 74 target sites revealed a mean read count ≥ 5 (83 without ≥ 5 filter) using TRACEseq NEXTseq data. These had a median anticorrelation of $\rho = -0.89$ (range -0.83 to -0.94). 4) Of note, multiple RefSeq ID could potentially overlap the same 6mer site. Out of the 74 target sites, 38 corresponded to unique 6mers. These 38 6mers thus flagged as true positives. 5) Finally, to capture as many true positives as possible, we also flagged 17 experimentally validated *bona fide* MRE sites that have been found active by the GenERA approach (Qianxin Wu, *manuscript under consideration*).

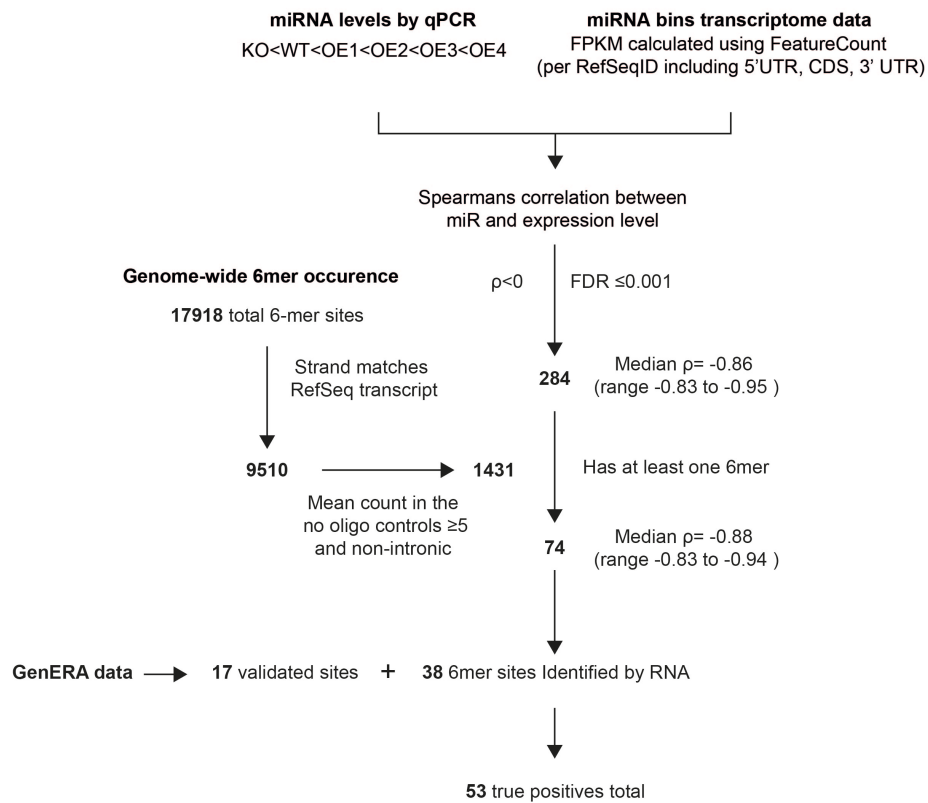


Figure 2.1 | Derivation of true positives for model training. Spearman's correlation was calculated between miR-184 expression and transcriptome data. Only anti-correlated pairs were propagated ($p < 0$) were considered (284). Of these 74 pairs carried at least one 6mer whose mean read count ≥ 5 . 38 of these corresponded to unique sites. 17 genERA validated MREs were added to generate 53 true positives in total. KO = knockout, WT = wildtype, OE = overexpression, FPKM = fragments per kilobase mapped, FDR = false discovery rate

2.2.5.2.2 Logistic regression training

For logistic regression training we used 3 biological replicates for each 6mer, yielding a total data set of 159 true positives and 4,149 true negatives. A random subset of 75% was chosen as a training set while the remaining 25% were retained as a validation set. Using the R function 'glm' with 'link="logit"' we trained the model based on the following predictors *i*) GFP control normalised read count, *ii*) relative protection score for both WT and KO and *iii*) the TRACE score, which were defined as follows:

$$GFP \text{ normalised read count (GNRC)} = \frac{\text{raw read count motif}_i}{GFP \text{ read count}_i}$$

$$\text{Relative Protection Score (RPS)} = \frac{GNRC_{+ASO}}{GNRC_{-ASO}}$$

$$\text{TRACE Score} = \frac{RPS_{WT}}{RPS_{KO}}$$

where i = experimental condition.

2.2.5.2.3 Assessment of GLM

We used the R package ‘ROCR’ to generate area under the curve metrics, F measure, sensitivity and specificity. In order to test for non-equality at the bases 1 and 8 in the *hits* and *non-hits*, we performed chi-square tests using the R function ‘chisq.test’. We then compared the distribution of 6mers within genes between hits and non-hits using another chi-square test.

2.3 *Bicoid* translational quiescence methods

2.3.1 Nanostring analysis

Egg chambers of yeast-fattened Oregon R flies were dissected into four bins (pre-stage 9, 9, post-stage 9, 14) and total RNA was extracted. 100ng total RNA were loaded into the nCounter *Drosophila melanogaster* miRNA Expression Assay Kit (Nanostring). The protocol was carried out according to the manufacturer’s instructions with a 16 h hybridisation time. Data was normalised using the nSolver software (Nanostring), using positive normalisation and background deduction.

2.3.2 miRNA target prediction

An offline version of the miRanda algorithm was executed to predict MREs in *bcd* isoform E (Enright et al., 2003). First prediction was performed with standard score

threshold of 140, second, post-pull down prediction was executed at slightly more permissive threshold of 120.

2.3.3 piRNA analysis

piRNA *bcd* mapping was performed as previously reported (Rouget et al., 2010). In brief, sequenced piRNAs deposited in the Gene Expression Omnibus as GSM327620 (Ovary_rep1_w1118_P) and GSM327625 (Embryo_0-2hrs_rep1_w1118_0_2hr_Emb) (Brennecke et al., 2008) were BLASTed with *E value* of 100 and minimum match of 14 nucleotides against the full length *bcd* sequence (isoform E). piRNA alignment was performed using *muscle* algorithm (Edgar, 2004).

2.3.4 Pulldown probe generation

9 ASO (25mer, $T_m \sim 60$) were tested as putative pull down probes using a previous version of the TRACE assay. We chose sites that are not overlapping with putative miRNA response elements. All ASOs were BLASTed against the *Drosophila melanogaster* genome to avoid off-targets with homology greater than 14nt. ~ 80 Oregon R female flies were yeast fattened over night. Ovary dissections took place the next day. Ovaries were dissected in ice-cold PBS and dissociated into individual ovarioles / egg chambers by pipetting up and down with a P-200 pipettor. The dissociated egg chambers were resuspended in 1 ml of 1 % formaldehyde in PBS and rotated end-over-end for 10 min at room temperature. The crosslinking reaction was quenched through resuspension of the material in 0.125 M Glycine. The reaction was allowed to incubate for 5 min while rotating. The egg chambers were washed 3x in ice-cold PBS and finally resuspended in 300 μ l ice cold, SUPERasin and DTT freshly supplemented TRACE Lysis buffer and homogenized using a Dounce homogenizer (10 strokes with pestle A and 10 strokes with pestle B). The insoluble material was spun down for 10 min at 16,100xg and the cleared lysate (supernatant)

was snap frozen in 100 µl aliquots in liquid nitrogen and then stored at -80 °C. To map the accessible sites, 10 µl lysate was incubated with 70 µl RNase H buffer (1X), 1 µl 1 M DTT, 0.3 µl 1 M MgCl₂, 4 µl RNase H (NEB), 0.5 µl SUPERasin (Ambion) and 4 µl antisense oligonucleotide (100 µM, IDT) or water (–ASO control) for 1 h at 37 °C. A DNase treatment was then performed through adding 1 µl DNase I (Promega) and 0.1 µl 60 mM CaCl₂ and incubating this mix for 10 min at 30 °C. Crosslink reversal and protein digestion was performed through adding 4 µl 1 M Tris.HCl, 4 µl 0.5 M EDTA, 4 µl 10 % SDS, 4 µl Proteinase K (freshly pre-mixed) and incubating the mixture for 1 h at 55 °C and 30 min at 65 °C. RNA was extracted as described above. Samples were then treated with Turbo DNase (Ambion). 3 µl 10X Turbo DNase buffer (0.1 volume) and 1 µl Turbo DNase were added to each 30 µl RNA elution and incubated at 37 °C for 30 min. The enzyme was deactivated through addition of 3 µl inactivation reagent and incubation for 5 min with gentle flicking of the tube every minute. 25 µl RNA were recovered and precipitated through addition of 2.5 µl 3 M sodium acetate (Ambion), 75 µl 100% ethanol and 1 µl glycogen (Ambion). The RNA was incubated for at least 15 min at -80 °C and then spun at maximum speed for 30 min at 4 °C. The supernatant was decanted and the pellet was washed in 70 % ethanol and spun again. All ethanol was removed and the tube allowed to air dry for 5 min. The RNA was resuspended in 10 µl nuclease-free water and its concentration and purity determined on a Nanodrop spectrophotometer. The RNase H sensitivity was determined using control (ORF) primers and primers spanning the cut site of the ASO (target). The ASO which displayed the highest RNase H sensitivity were chosen as pull down probes and ordered from Integrated DNA Technologies (IDT) in the following format: {5'-Oligonucleotide Sequence-3'} – C18 Spacer – Amino Modifier.

2.3.5 Probe - Bead coupling

Dynabeads MyOne Carboxylic Acid (Invitrogen, cat # 65011) were coupled to oligonucleotides in a ratio of 50 nmol oligonucleotide : 1 ml Dynabeads in batches of 1ml or multiples thereof. Dynabeads were resuspended and 1 ml was transferred to a low-retention tube microcentrifuge tube (Sorenson, cat # 39640T). The tube was inserted into a magnet for 2min and the supernatant was removed. The Dynabeads were washed twice in 1ml 100mM MES buffer and resuspended in 100 μ l thereof. In a new tube, 50 μ l 5' amino modified oligonucleotides (25 nmol oligonucleotide 5 plus 25 nmol oligonucleotide 8) were mixed with 40 μ l 1.25 M EDC in 100mM MES pH 4.8, and made up to 100 μ l. The activated oligonucleotide mix was added to the Dynabeads and briefly vortexed. The mixture was incubated overnight on a roller mixer (~12-16h). The coupled Dynabeads were then washed three times in 1ml TT buffer for a period of 30 min each. Finally, the coupled Dynabeads were resuspended to a concentration of 10 mg/ml and stored at 4 °C until used.

2.3.6 Material Generation

2.3.6.1 Ovary

Approximately 200 ovaries were dissected from yeast fattened female Oregon R flies (2-4 days old) in ice-cold PBS. They were washed in PBST and dissociated into individual egg chambers through pipetting up and down 20 x with a P 200 pipettor. The egg chambers were allowed to settle; the majority of the liquid was removed. The egg chambers were then transferred with a P-200 with a cut tip (larger orifice avoids egg demolition) on a 90 mm petri dish, arranged in an even distribution with as much liquid removed as possible, and placed on ice. The material was UV crosslinked (254 nm) on a Stratalinker with 1 Joule per cm^2 . After crosslinking the egg chambers were carefully washed back into a 1.5 ml tube with PBST. Liquid was removed as much as possible and the material was weighed on a tared, scientific

scale. 200 mg ovaries were lysed in 4 ml ice-cold lysis buffer freshly substituted with protease inhibitor tablets and RNase inhibitor. The ovaries in lysis buffer were transferred to a 2 ml Dounce homogeniser and homogenised with 20 strokes with pestle A and 20 strokes with pestle B. Finally the lysate was passed 5 x through a 21G needle. The lysate was snap frozen in liquid nitrogen and stored at -80 °C.

2.3.6.2 Embryo

Young Oregon R flies (1-7 days of age) were set up in population cages with an excess of female flies, approx. 80,000 flies/cage. Flies were trained on a 12 h light dark cycle to synchronise egg lays in the morning (darkness). At least one prelay was performed for a minimum of 60 min to eliminate any embryos that initiated development *in utero*. They were then dechorionated for 2-3 min in bleach (Sainsbury's) and thoroughly washed with water afterwards. They were then processed identically to the ovaries (see above).

2.3.7 RNA Capture

2.3.7.1 Capture bead preparation

1.4 ml of ASO coupled Dynabeads in a 5 ml tubes (Eppendorf) were placed on a DynaMag™ -2 Magnet until all beads were caught on the side wall (~ 1 min). The liquid was removed and the beads were equilibrated in 1.4 ml lysis buffer (freshly substituted with protease inhibitor tablet and SUPERasin RNase inhibitor). 70 µg of *Xenopus* RNA were added for pre-absorption and the beads were incubated for 1 h at 40 °C. The beads were then washed 1 x with 5 ml buffer 1 and 3 x with 1 ml elution buffer at 40 °C. The beads were re-equilibrated in 1.5 ml lysis buffer and washed twice in that buffer.

2.3.7.2 Lysate preparation

700 µl uncoupled empty beads were equilibrated in 1.4 ml lysis buffer. 4 ml lysate was thawed on ice and split into 2 x 2 ml tubes (Axygen) and centrifuged at 4 °C for 20 min to clear insoluble material. The lysate was then incubated for 20 min @ 60 °C in a heat block. Approx. 3500 µl lysate was added to the empty beads and incubated for 20 min at 40 °C to pre-clear the lysate.

2.3.7.3 Pull Down

10 µl pre-cleared lysate were taken off and snap frozen as the input fraction. The remaining lysate was added to the pre-absorbed, re-equilibrated capture beads. The antisense purification was performed for 3 h at 40 °C in a hybridisation oven. 10 µl flow through were snap frozen and saved for analysis. The beads were washed 1 x with 40 °C pre-warmed, SUPERasin substituted (1 µl/ ml buffer) lysis buffer for 5 min at 40 °C. The beads were then washed twice with 5 ml SUPERasin substituted (1 µl/ ml buffer) pre-warmed buffer 1 for 5 min at 40 °C. This was followed by two more hot washes with buffer 2 and two more washed with buffer 3. Finally, the RNA was eluted in 400 µl elution buffer at 95 °C for 3 min.

1/4th volume 5X proteinase K buffer and 5 µl proteinase K (Ambion) was then added to all samples (input, flow through, eluate and beads) and proteins were digested for 1 h at 55 °C. Next the eluate fraction was concentrated in a Amicon Ultra YM-3 column (Millipore). 500 µl eluate was added to the column which was then spun for 30 min at 14,000 xg at room temperature. Immediately after the spin, the column was inverted and placed into a new Amicon 2 ml collection tube and the concentrated liquid was eluted for 2 min at 8,000xg centrifugation. RNA was extracted from all samples using the miRNeasy kit as described below. ~ 5 ng RNA was converted to cDNA and qPCR was performed for quality control of RNA enrichment and integrity.

2.3.8 Library Preparation and NGS Analysis

The NEB Next Small RNA library kit (E7300) was used according to the manufacturer's instructions to generate small RNA libraries of the eluate RNA. As the eluate contained very small amounts of RNA (~ 10 ng vs. recommended minimum of 100 ng) all primers and adaptors were diluted 1:20 in nuclease free water. The library was size selected on the Pippin Prep (Sage Bioscience) set to isolate sizes 125-165 bp. The library was checked for correct fragment size (~ 143 bp) on a high sensitivity DNA bioanalyzer chip (Agilent, # 5067-4626) and sequenced on a NextSeq DNA sequencer (Illumina). Raw reads were trimmed using cutadapt (v1.8.3) (`-g <5p adaptor> -a <3p adaptor> --minimum-length 10 -n 2 --discard-untrimmed`) and aligned using BUTTER (v0.3.2) (`--mismatches=1`) (Axtell, 2014). miRNA expression box plots were generated using 'featureCounts' `-a /databank/igenomes/Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -t exon -g gene_id -o features_counted -s 0 merged_R1_5p3p_trimmed.bam` (Liao et al., 2014). We used the "transcript_biotype" field in the genes.gtf file to generate the pie chart displaying different annotations.

2.4 General Molecular Methods

2.4.1 RNA Work

Prior to beginning all pipettes and work surfaces were cleaned with RNase Zap (Ambion). Total RNA was extracted with RNeasy Mini Kit (Qiagen) or miRNeasy Mini Kit (Qiagen) if short RNAs were required according to the manufacturer's instructions. For RNA preparations not requiring small RNAs, 300 µl RLT lysis buffer was added to a maximum of 1×10^7 cells or 50 µl TRACE processed lysate. 300 µl of 70% ethanol was then added and the suspension was mixed by pipetting. Up to 700

µl were transferred to an RNeasy Mini spin column situated in a supplied collection tube. The column was spun for 30 s at 8,000xg and the flow through was discarded. Any remaining lysate was applied to the column and the previous step was repeated. The column was then washed with 700 µl RW1 buffer and spun at 8,000xg. The flow through was discarded. Next, the column was washed twice with buffer RPE each followed by centrifugation at 8,000xg for 30 s and 2 min, respectively. The column was then transferred to a clean 2 ml tube and spun again at full speed for 1 min to remove residual fluid and dry the membrane completely. The column was then placed into a fresh 1.5 ml tube and the RNA was eluted in 30 µl nuclease free water (1 min incubation followed by 1 min spin at 8,000xg). The concentration and purity (A260/280 and A260/230) was detected on a Nanodrop spectrophotometer.

The miRNeasy protocol contained the following modifications:

Cells or eluates were lysed in 700 µl Qiazol and incubated for 5 min at room temperature. 140 µl chloroform were added and tubes were vigorously shaken for 15 s. The tubes were then spun at 12,000xg for 15 min. The aqueous supernatant was removed and transferred to a fresh 1.5 ml tube. 1.5 volumes of 70 % ethanol were added and the mixture was pipetted up and down. Up to 700 µl were then added to an RNeasy spin column and centrifuged for 30 s at 8,000xg. Any remaining liquid from the previous step was added to the column. Next, the column was washed with 700 µl RWT buffer and centrifuged again for 30 s at 8,000xg. The column was then washed twice with 500 µl buffer RPE and spun for 30 s and 2 min at 8,000xg, respectively. The column was then placed in a fresh 2 ml collection tube and spun for 1 min at full speed to remove residual liquid and fully dry the membrane. The RNA was eluted by adding 30 µl nuclease free water, letting the column stand for 1 min and then spinning the column for 1 min at 8,000xg. The concentration and RNA purity was assessed using the Nanodrop spectrophotometer.

2.4.2 cDNA synthesis

In most cases 100 ng total RNA were reverse transcribed using the QuantiTect kit (Qiagen, #205310). Reactions were set up on ice. 100 ng RNA were incubated with 1 μ l gDNA wipeout (7x) in 7 μ l total volume for 10 min at 42°C. Following this DNA removal, 2 μ l 5x reaction buffer, 0.5 μ l primer mix and 0.5 μ l reverse transcriptase were added and the tube was incubated for another 30 min at 42 °C. This was followed by a 3 min heat inactivation at 95 °C. All cDNA was diluted 1:2, i.e. to 10 μ l was added 10 μ l nuclease free water.

2.4.3 qPCR

Quantitative polymerase chain reaction (qPCR) was performed in 10 μ l reactions (always technical triplicates) which were set up at room temperature as follows (per replicate): 5 μ l SsoAdvanced Universal SYBR Green Supermix (BioRad, #172.5271), 0.5 μ l (500 nM) forward and reverse primer, 3 μ l nuclease free water and 1 μ l cDNA. The reactions were loaded into 384 well hard shell plates (BioRad), sealed with adhesive foil and briefly spun down. Standard cycling conditions on a BioRad CFX384 qPCR machine were as follows: 98 °C for 2 min (enzyme activation), 40 cycles of 98 °C for 10 s and 60 °C for 20 s, with a plate read after each cycle. The data was normalised using the $\Delta\Delta C_t$ method using a housekeeping gene (*rp49* in most cases). All primers were designed with Primer Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) with amplicon length set to 50-150 bp and melting temperature set to 60 °C.

2.4.4 miRNA cDNA synthesis

miRNAs were reverse transcribed using the TaqMan Small RNA Assays (Applied Biosystems, # 4440885) protocol according to the manufacturer's instructions with the reaction volume halved. Each 7.5 μ l reaction contained 2.5 μ l RNA (usually 10

ng), 3.5 µl master mix (per reaction: 0.075 µl 100 mM dNTPs, 0.5 µl MultiScribe Reverse Transcriptase, 0.75 µl 10X reverse transcription buffer, 0.095 µl RNase Inhibitor and 2.08 µl nuclease free water) and 1.5 µl 5X RT primer. The reaction was incubated 5 min on ice. Next, the tube was placed into a thermal cycler and run with the following parameters: 30 min at 16 °C, 30 min at 42 °C and 5 min at 85 °C.

2.4.5 Taqman miRNA qPCR

10 µl qPCR reaction were assembled on ice as follows: 0.5 µl TaqMan Small RNA Assay (20X), 0.665 µl RT product, 5 µl TaqMan Universal PCR Master Mix II (2X) no UNG, 3.835 µl nuclease free water. Technical triplicates were loaded into 384 well hard shall plates (Bio-Rad) and sealed with an adhesive foil. The reaction was cycled as follows on a BioRad CFX384 machine: 10 min at 95 °C followed by 40 cycles of 95 °C for 15 s and 60 s at 60 °C, with a plate read after each cycle. The data was normalised using the $\Delta\Delta C_t$ method using 2S RNA as a control.

2.4.6 Digital droplet PCR (ddPCR)

ddPCR reactions were set up in 22 µl reactions with 11 µl ddPCR EvaGreen Supermix (BioRad, #1864034), 0.4 µl forward and reverse primer (400 nM), 9.2 µl nuclease free water and 1 µl cDNA. Droplets were either generated with the QX200 manual droplet generator (BioRad) or with the QX300 auto droplet generator (BioRad) by loading 20 µl (manual) or 22 µl (auto-DG) using droplet generation oil (BioRad). The droplets were carefully transferred into a semi-skirted 96 well plate (Eppendorf) which was then heat sealed. The reaction was cycled as follows: 5 min at 95°C, 40 cycles of 30 s at 95 °C and 1 m at 59 °C, 5 min at 4 °C, 5 min at 90 °C. The data was analyzed using QuantaSoft (BioRad).

2.4.7 *In vitro* Transcription

The *Xenopus eef1a* RNA used for capture bead pre-absorption was generated with the T7 MEGASCRIPT *in vitro* transcription kit (Ambion, AM1334) with a slightly modified protocol (Bassett and Liu, 2014). 20 µl reactions were assembled by combining the following in order, at room temperature (to avoid precipitation): 6 µl nuclease free water, 2 µl ATP, 2 µl CTP, 2 µl GTP, 2 µl UTP, 2 µl 10x reaction buffer, 2 µl (supplied *eef1a* template plasmid) and 2 µl enzyme mix. The reaction was then incubated for 4 h at 37 °C. After 4 h, 1 µl Turbo DNase was added and the reaction was incubated 15 more minutes. The reaction was then stopped through the addition of 115 µl H₂O and 15 µl ammonium acetate stop solution. The RNA was extracted through addition of 150 µl phenol:chloroform:isoamyl alcohol (25:24:1) and thoroughly vortexing for 30 s. The phases were separated by a 3 min centrifugation step at 10,000xg at room temperature. The upper phase was transferred to a fresh tube. The RNA in the aqueous phase was then precipitated by adding 150 µl (equal volume) of isopropanol. The solution was mixed and incubated at -20 °C for at least 15 min. The RNA was then collected via centrifugation at 17,000xg for 30 min at 4 °C. The pellet was washed twice in 0.5 ml 70% ethanol at room temperature with centrifugations at 17,000xg for 3 min at 4 °C in between. The pellet was then air dried to evaporate residual ethanol for approx. 3 min. The pellet was resuspended in 30 µl nuclease free water and frozen at -80 °C following Nanodrop concentration measurement.

The *egfp* spike-in control used in the TRACEseq was generated with the T7 mMessage Machine *in vitro* transcription kit (Ambion, AM1344) using a minimally changed protocol (Bassett and Liu, 2014). 10 µg of DNA template were linearised with EcoRI in a 50 µl reaction with 5 µl enzyme, and 5 µl 10x NEB reaction buffer for 2 h at 37 °C. The linearisation was terminated through the addition of 1/20 volume 0.5 M EDTA and 1/10 volume 3 M sodium acetate. The DNA template was then

precipitated through addition of 2 volumes ethanol and incubation at -20 °C for a minimum of 15 min. The DNA was collected at 17,000xg for 30 min at 4 °C. Residual fluid was removed and the pellet was air dried for 5 min at room temperature. The dried pellet was resuspended in 12 µl nuclease free water and the concentration was measured on a Nanodrop spectrophotometer and adjusted to 500 ng/µl. For the *in vitro* transcription, a 20 µl reaction was assembled with 4 µl nuclease free water, 1 µg linearised plasmid, 10 µl 2x NTP/CAP reagent, 2 µl 10x reaction buffer and 2 µl enzyme mix. The reaction was mixed and incubated for 2 h at 37 °C. Next, 80 µl nuclease free water were added to the reaction mixture and the RNA was purified. 350 µl RLT buffer was added and the solution was mixed. 250 µl 100% ethanol were added next and the mixture was pipetted up and down. The sample was transferred to a RNeasy spin column and centrifuged for 15 s at 10,000xg at room temperature. The flow-through was discarded. The column was washed twice with 500 µl RPE buffer spinning at 10,000xg for 15 s and 30 s, respectively. The column was transferred to a clean 2 ml tube and spun for a further 2 min to further dry the column. The column was then placed in a 1.5 ml collection tube and nuclease free water, pre-warmed to 37°C, was applied to the column center. The column was incubated for 1 min and then spun for 1 min at 10,000xg at room temperature. The concentration was determined on a Nanodrop spectrophotometer and the RNA was frozen at -80 °C in 10 µg aliquots.

2.4.8 Flow Cytometry

Flow cytometry was performed 48 h after transfection on a Fortessa LSR (BD Bioscience). Cells were washed once in 1x PBS to remove residual media, with a pelleting centrifugation at 500 xg for 3 min. Cells were gated to remove debris and isolate singlets. 50,000 events were recorded per sample. Relating to the *bcd* 3'UTR sensor in chapter 5, settings for EYFP were excitation at 488nm with a 530/30A

emission filter. For mCherry, settings were excitation at 561nm with filter at 610/20. Cells were gated for EYFP and mCherry expression. Normalised mCherry fluorescence was then calculated: mean mCherry fluorescence/mean EYFP fluorescence.

2.5 Cell Culture

2.5.1 Cell maintenance and passage

S2R+ cells were maintained at 25 °C in *Drosophila* Schneider's medium (Gibco, #21720024) supplemented with 10% Fetal Bovine Serum (Gibco, #10500) and 0.5% Penicillin-Streptomycin (Gibco, #15140122). Cells were passaged twice per week with a dilution of 1:5 (cell suspension : fresh media).

2.5.2 Transfections

Transfection experiments were carried out as follows: 10^6 S2R+ cells are plated on day 1 with full Schneider's *Drosophila* media (Gibco, #21720024) + 10% heat-inactivated Fetal Bovine Serum (Gibco, #10500) in a 12-well plate (~ 80 % confluency). On day 2 (day of transfection) the media is exchanged for 0.5 ml Schneider's medium without FBS (critical for high transfection efficiency). 500 ng DNA is mixed with 2.5 μ l TransIT2020 transfection reagent (Mirus, # MIR5400), made up to 25ul volume with nuclease free water in a 1.5 ml tube and incubated at room temperature for 20 min. The mixture is then pipetted gently and evenly onto the cells. 4 h post transfection 0.5 ml full Schneider's medium is added to the cells. Cells are grown for 48 h and then harvested.

2.6 Drosophila Methods

2.6.1 Fly husbandry

Oregon R flies were maintained on standard cornmeal agar medium at 25 °C.

2.6.2 Ovary dissections

Young female flies (2-3 days after eclosure) were fattened with active dry yeast for 2 days in the presence of a few males (approx. 10 females / 1 male). On the day of the dissection the flies were anaesthetised on CO₂ pads (FlyPad) and dissected using forceps in ice-cold PBS for no longer than 1 h per dissection batch.

2.6.3 Embryo collections

Young Oregon R flies (1-7 days of age) were set up in population cages with an excess of female flies, approx. 80,000 flies/cage. Flies were trained on a 12 h light dark cycle to synchronise egg lays in the morning (darkness). Juice plates were made from 666 ml H₂O, 222 ml apple squash (Waitrose), 22 g sucrose, 20 g agar per batch. The mixture was microwaved for ~ 5 min or until all solutes were dissolved and poured immediately into 90 mm petri dishes (Sterilin). At least one prelay was performed for a minimum of 60 min to eliminate any embryos that initiated development *in utero*.

Chapter 3: *In vivo* Detection of miRNA-target Interaction Events (miR-TRACE) Decodes Functional miRNA Response Elements

3.1 Introduction

Post-transcriptional regulation of the genetic programme is achieved through intricate interactions between *cis* regulatory elements and *trans* factors. microRNA response elements (MREs) are an extensively studied class of *cis* elements that confer repression upon RNAs they are embedded in (Bartel, 2009). The efficacy of target repression depends on a multitude of factors that include local site architecture, RNA secondary structure, transcript abundance, competition with other MREs, and the interplay with RNA binding proteins (Ho and Marsden, 2014; Jens and Rajewsky, 2015). Hence, deciphering these interactions requires technologies that permit their unbiased interrogation under native, physiological conditions.

miRNAs typically recognise their targets predominantly by virtue of the 'seed' region, nucleotides 2-7 of the miRNA (Bartel, 2009). This recurring pattern led to the development of a host of bioinformatics prediction algorithms that screen conserved 3' UTRs for complementarity to the seed region (Li and Zhang, 2015). Consequently, *in silico* target prediction typically yields an overwhelming number of predicted targets, including many false positives (Rajewsky, 2006). Employing miRNA mutagenesis (loss-of-function) or miRNA overexpression (gain-of-function) coupled with RNAseq, proteomics or ribosome profiling can confine a large number of predicted targets to a smaller set of physiologically relevant ones. However, profiling approaches are typically unable to distinguish direct (primary) from indirect (secondary) targets in which an upstream regulator of a putative target RNA is

deregulated. Furthermore, they are carried out under altered stoichiometry (e.g. miR overexpression) which may engender compensatory effects and perturb homeostasis (Khan et al., 2009). While physical binding information from CLIP experiments or tagged RNA pulldowns can reveal primary targets, CLIP data is frequently confounded by the incidence of artefacts caused by two separate library preparations (one long mRNA and one short RNA) and subsequent bioinformatics *de novo* pairing. Hence, the functional relevance of the target interaction under truly endogenous conditions remains difficult to establish.

This chapter details the development of *in vivo* detection of miRNA-target interaction events (miR-TRACE), a novel validation platform for miRNA targets *in situ*. The underlying principle of miR-TRACE is RNase H cleavage of *in vivo* crosslinked RNAs. We reasoned that one of the most marked differences between a biologically relevant miRNA::target interaction (active MRE) and other stochastic binding events, is manifested by the robust assembly and prevalence of the miRISC complex at the MRE locus. This interaction can be stabilised *in situ* via chemical crosslinking and maintained even through high detergent lytic conditions (Hassan et al., 2013). To evaluate whether an MRE is bound or not, antisense oligonucleotides (ASOs) are hybridised to the corresponding RNA domains under stringent conditions (high salt, high detergent) (Fig. 3.1). The binding of each ASO to the target MRE is measured by RNase H digestion, which takes place following dilution of the detergent concentration. Therefore, since hybridisation can only occur at unoccupied regions, only such sites are subject to cleavage. On the contrary, RNA sites that are bound by protein (e.g. the miRISC) are expected to be unavailable for hybridisation and thus refractory to RNase H cleavage. Here I present the direct comparison of RNase H protection scores at candidate MREs in cell lysates derived from wild type (WT) S2R+ and miR-184 knock out (KO) cells, to distinguish active from inactive MREs.

RNase H is an endoribonuclease which cleaves the RNA strand in a RNA:DNA heteroduplex (Donis-Keller, 1979; Stein and Hausen, 1969). It is part of the family of nucleotidyl-transferases that also comprise argonaute nucleases, transposases, retroviral integrases and the Holliday junction resolvase (Nowotny et al., 2005; Pallan and Egli, 2008). The roles of endogenous RNase H include discarding RNA primers from Okazaki fragments and restoring DNA topology *via* resolving R loops (Broccoli et al., 2004; Cerritelli and Crouch, 2009). RNase H null mice do not survive beyond day 8.5 due to an inability to synthesise mitochondrial DNA (Cerritelli et al., 2003).

Due to its ease of targeting, RNase H has been abundantly exploited in oligonucleotide therapy. Its activity in cells underpins the conceptual basis for supplying antisense DNA oligonucleotides to target mRNAs for a reduced protein output. For example, Mipomersen (Kynamro™) is a 20nt ssDNA ASO targeting *apop* mRNA via recruiting endogenous RNase H, thereby reducing apolipoprotein B levels in patients suffering from familial hypercholesterolemia (Raal et al., 2010).

In addition to down-regulating gene expression, ASOs can also be used to direct RNase H cleavage in a manner that is useful for assaying RNA secondary structure and accessibility. Since ASOs can only hybridize with sterically accessible, complimentary RNAs, therefore subsequent cleavage by RNase H serves as a useful proxy for RNA accessibility. This principle has been directed towards the study of spliceosomal RNA structures. Lazar *et al.* probed the 5' end of the U1 RNA with an antisense oligonucleotide and successfully digested the heteroduplex with RNase H under non-denaturing conditions, which revealed that this portion of the U1 RNA is accessible and available to undergo enzymatic splicing (Lazar et al., 1982). In 1984, Ruskin and colleagues identified an inaccessible, RNase H resistant region that forms a lariat structure critical for splicing (Ruskin et al., 1984). The first structure of an entire RNA (apolipoprotein II mRNA) was probed using tiled antisense

oligonucleotides in 1989 (Hwang et al., 1989). About a decade later, the first randomised ASO libraries have been employed to identify effective target regions for antisense inhibition (Birikh et al., 1997; Ho et al., 1998).

Up to this time, the majority of mapping protocols used *in vitro* transcribed RNA or *in vitro* assembled ribonucleoprotein particles (RNP). While this may possibly recapitulate endogenous RNA folding and thus be appropriate for designing ASOs avoiding higher order intramolecular RNA structures, it is unlikely that this approach reflects the RNA in its truly native state. Neither does this approach consider the dimension of RNA binding proteins (RBP) that frequently decorate and regulate mRNAs. Hence, to identify *bona fide* accessible sites, it is important to consider the RNA target in its endogenous context. To this end, Scherr *et al.* mapped the RNA encoding mouse DNA methyltransferase in freshly prepared cell lysates, at least partly reflecting its native physiological context (Scherr and Rossi, 1998).

One inherent problem of performing biochemistry under non-denaturing conditions is the random association of negatively charged RNA with positively charged proteins *via* non-specific electrostatic interactions. Hence, in crude cell lysates it remains difficult to evaluate whether a protected site is the result of a cell programmed protein-RNA interaction or simply an artefact of a non-specific association that occurred post-lysis. To overcome this limitation, *in vivo* crosslinking can be employed to 'snap-freeze' physiological interactions *in situ*, and maintain these throughout cell lysis under denaturing conditions. This strategy abrogates non-physiological binding and ensures that only *bona fide* RNA binding proteins remain associated due to their covalent attachment to a target RNA.

The first combination of *in vivo* crosslinking and subsequent RNase H accessibility mapping was performed by the Kingston lab to identify accessible RNA regions and guide the design of antisense probes for pull down assays (Simon et al., 2011).

Surprisingly, they reported RNase H activity in denaturing lysis buffer containing high detergent concentrations (Simon et al., 2011). Here, I have adapted and optimised this conceptual framework to develop a high-resolution RNA mapping platform directed at the study of RNA *cis* regulatory elements, particularly miRNA response elements (MRE).

3.2 Chapter Aims

- Implementation and optimisation of RNase H mapping strategy under denaturing conditions.
- Proof-of-concept that miR-TRACE reflects MRE occupancy by miRISC
- Assessment of the potential of miR-TRACE to distinguish between previously characterised functional and non-functional MREs.

3.3 Results

3.3.1 RNase H protocol overview

In vivo formaldehyde crosslinking ‘snap-freezes’ and stabilises the ternary complex of mRNA, miRNA and RISC (Fig. 3.1a). Cells are then lysed under high salt and detergent conditions, favouring nucleic acid interactions while disrupting any non-physiological protein associations (i.e. non-crosslinked). As a consequence, a site that is not bound by an RBP or the RISC complex should be accessible to antisense oligonucleotide (ASO) hybridisation (“unprotected”), if the secondary structure is permissive. On the contrary, a biologically relevant MRE (bound by a cognate miRNA/RISC complex) is expected to be unavailable for ASO hybridisation (“protected”) (Fig. 3.1b).

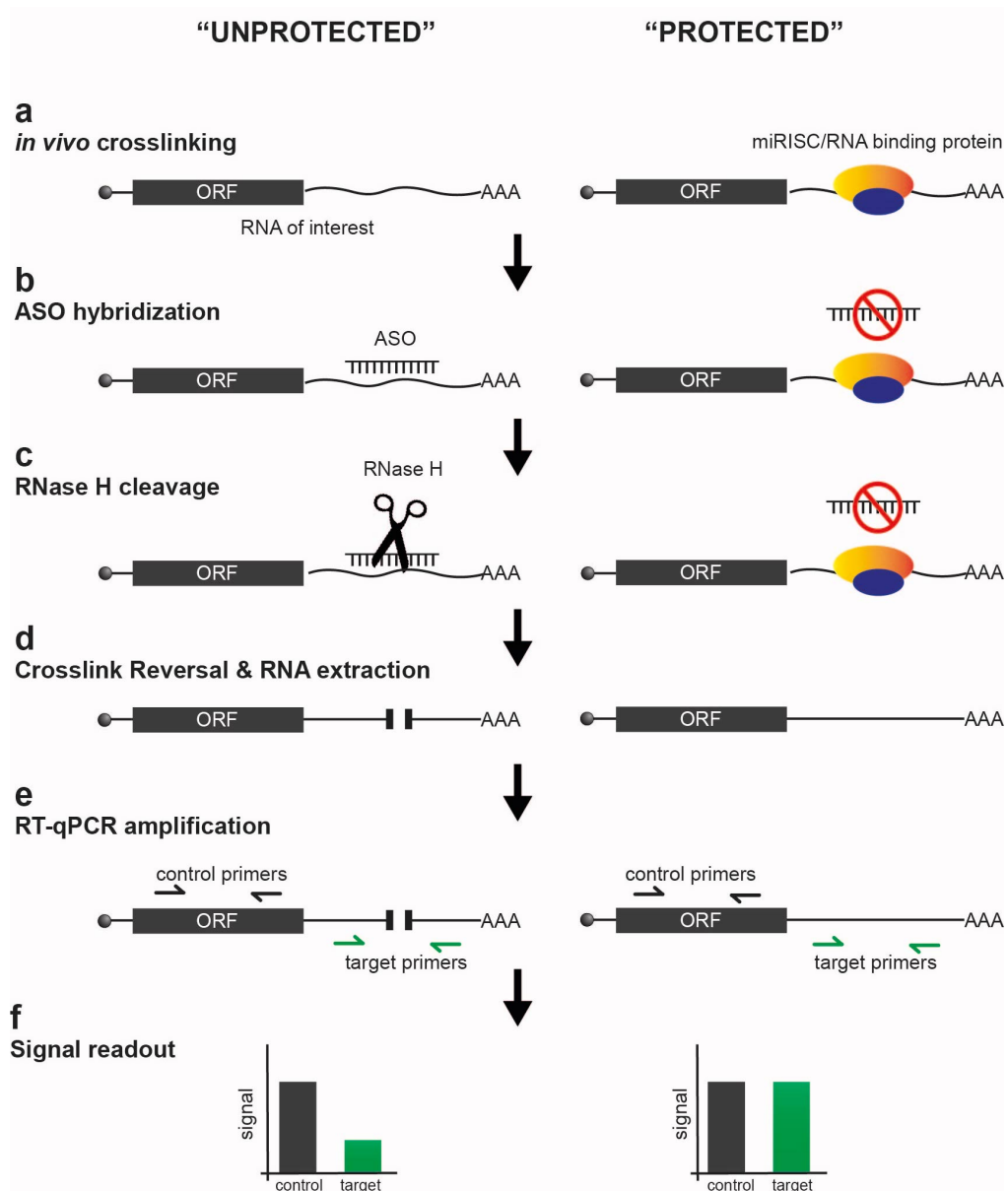


Figure 3.1 | RNase H can distinguish between unprotected and protected RNA elements on the basis of antisense oligonucleotide (ASO) hybridisation. a, Intact cells are formaldehyde crosslinked *in vivo*. **b,** ASO are targeted against regions of interest (e.g. MRE) and can hybridise depending on RNA binding protein occupancy. **c,** Hybridisation events are “captured” by RNase H cleavage of the RNA strand. N.B. protected, e.g. RISC-bound RNAs cannot be cleaved due to steric occlusion of the RNase H. **d,** Crosslink reversal and RNA extraction are performed prior to analysis. **e,** qPCR detects cleavage events using control primers against a non-targeted region and target primers spanning the cut site. The ΔC_t between the primer pairs reflects the cutting at the MRE and is an indirect read out of protection. **f,** A difference between control and target primer indicates an unprotected, unbound MRE while similar primer signals would reflect a bound, regulated MRE. Green colour is signal from target primers, black graph is data from control primers.

To assess whether hybridisation takes place or not, the detergent concentration is reduced by dilution and recombinant RNase H is delivered to cleave RNA strands that are engaged in ASO heteroduplexes (Fig. 3.1c). Following RNase H digestion, all crosslinks are reversed by proteinase K treatment and total RNA is extracted (Fig. 3.1d). The RNA is then reverse transcribed and subjected to quantitative PCR (qPCR) analysis with primer pairs spanning the cut site (target primers) and primer pairs amplifying an upstream uncut region (control primers) of the same cDNA molecule (Fig. 3.1e). A protected site should yield no signal difference between the two primer pairs, while a cut (unprotected) site is expected to have a lower signal for the target primer (Fig. 3.1f).

3.3.2 Overview of miR-TRACE assay

To study miRNA::target interactions using this approach, we performed a paired RNase H protection analysis of candidate MREs in *Drosophila* S2R⁺ wild type (WT) cells and miR-184 knockout (KO) cells (Fig. 3.2). The conserved miR-184 is one of the most abundant miRNAs in S2R⁺ cells and its predicted primary target network encompasses at least 77 predicted MREs (Qianxin Wu, *manuscript under consideration*). We hypothesised that an active MRE, which is occupied by the miR-184 RISC complex, is inaccessible to ASO hybridisation and hence protected against RNase H cleavage due to steric occlusion (Fig. 3.2 [1]). However, if the miRNA is removed from cells *via* genomic ablation (KO) we expect miRISC to no longer engage with this site, rendering a decrease in local RNase H protection (Fig. 3.2 [1]). On the contrary, if an MRE is predicted to be targeted but not physiologically bound by the miRISC complex we expect this site to be equally unprotected in both WT and miR-184 KO cells (Fig. 3.2 [2]). A third option is if the MRE is occluded by higher order RNA structure or masked by an RBP preventing miRISC access. Such a site is expected to appear protected across WT and KO cells (Fig. 3.2 [3]). According to this

principle, a reduction in relative protection from WT to KO is only expected for *bona fide* MREs (Fig. 3.2 d).

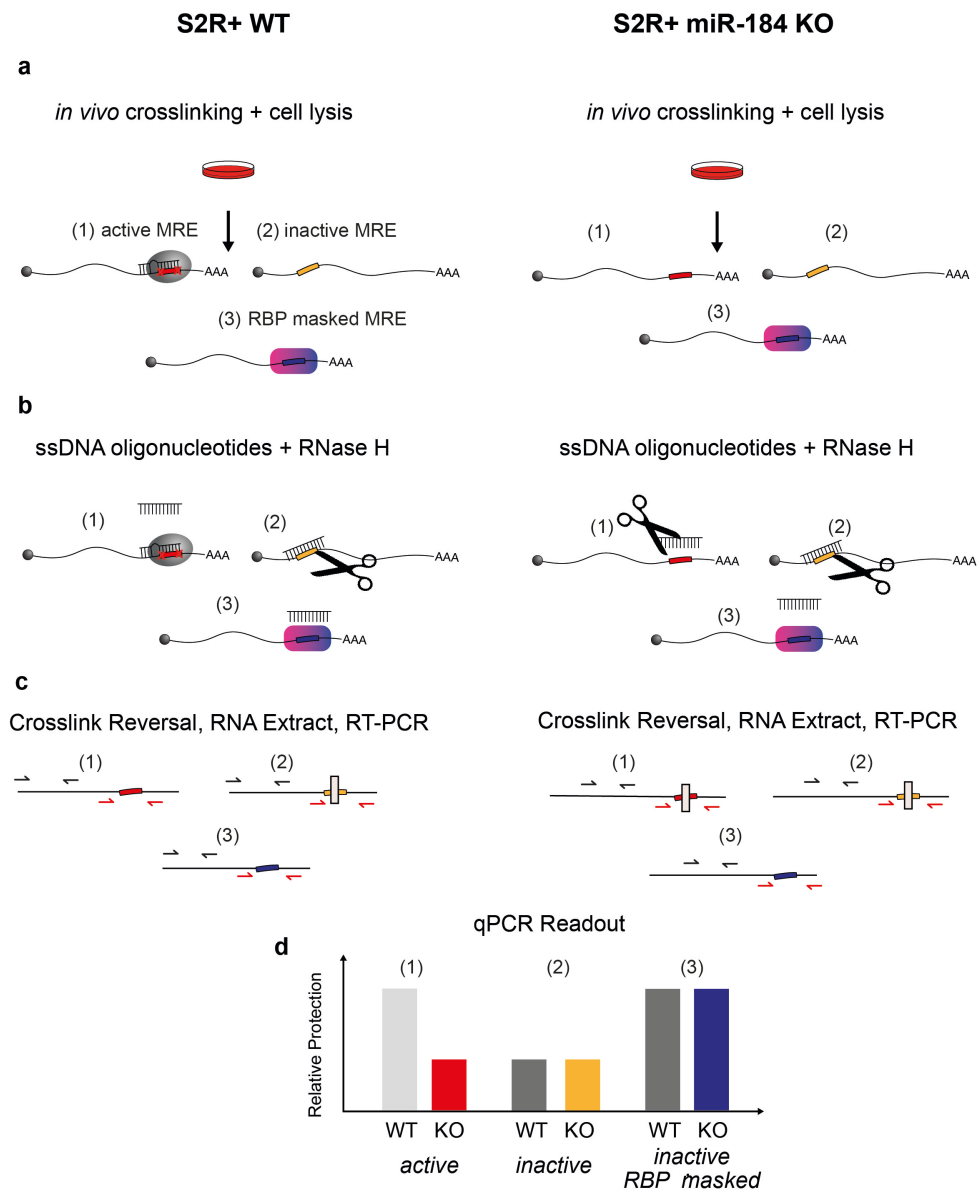


Figure 3.2 | The miR-TRACE assay measures differential accessibilities at miR-184 predicted target sites. a, S2R+ WT and miR-184 KO cells are processed in parallel (formaldehyde crosslinking and cell lysis). **b,** 25mer antisense oligonucleotides (ASO) targeting the seed region of predicted MREs are hybridised under stringent conditions and successfully formed DNA::RNA heteroduplexes are digested by RNase H. **c,** Crosslinks are reversed, the RNA is extracted and reverse transcribed. N.B. if MRE is cut, cDNA synthesis is interrupted. qPCR is performed with two primer pairs (control = black and target = red). **d,** The relative protection of each MRE under WT and KO is calculated (see Fig. 3.4 for details).

3.3.3 Optimisation of RNase H miR-TRACE Mapping Protocol

Before comparing the relative MRE protection across WT and KO cells we evaluated the technical performance of the previously published RNase H protocol on non-crosslinked *Drosophila* ovary lysates (Simon, 2013). The accuracy of the RNase H mapping depends on a large ΔC_t dynamic range between control and target primers in which different hybridisation events are compared (Fig. 3.3e). This requires conditions where the catalytic activity of the RNase H is solely dependent on the ability of the DNA oligonucleotide to hybridise to the target RNA. Surprisingly, according to the published protocol, this enzymatic reaction was hitherto carried out under high detergent conditions present in the lysis buffer (Simon, 2013; Simon et al., 2011). Our assessment revealed that the catalytic activity of RNase H in this buffer was significantly compromised (Fig 3.3a). Thus, under these conditions, no meaningful ΔC_t between a primer pair spanning the cut site (A) and an upstream control primer (ORF) could be obtained even when using non-crosslinked RNA as substrate (Fig. 3.3a).

Based on this observation, we reasoned that reducing detergent concentration post-lysis would create a more favorable environment, thus improving the catalytic activity of RNase H. To assess this experimentally, we diluted the reaction volume 1:4 (0.25X) (Fig. 3.3b, c) and 1:8 (0.125X) (Fig. 3.3d) in NEB supplied RNase H buffer (1X) immediately before RNase H delivery. To account for larger reaction volumes, we increased the amount of DNA oligonucleotides and RNase H accordingly. Analysis of duplex formation on non-crosslinked RNA revealed a progressive increase in detection sensitivity with a maximum ΔC_t dynamic range of 8.65 cycles, approaching naked RNA under optimal conditions (Fig 3.3f). A 1:8 (0.125X) dilution was eventually selected as the standard condition for all mapping experiments.

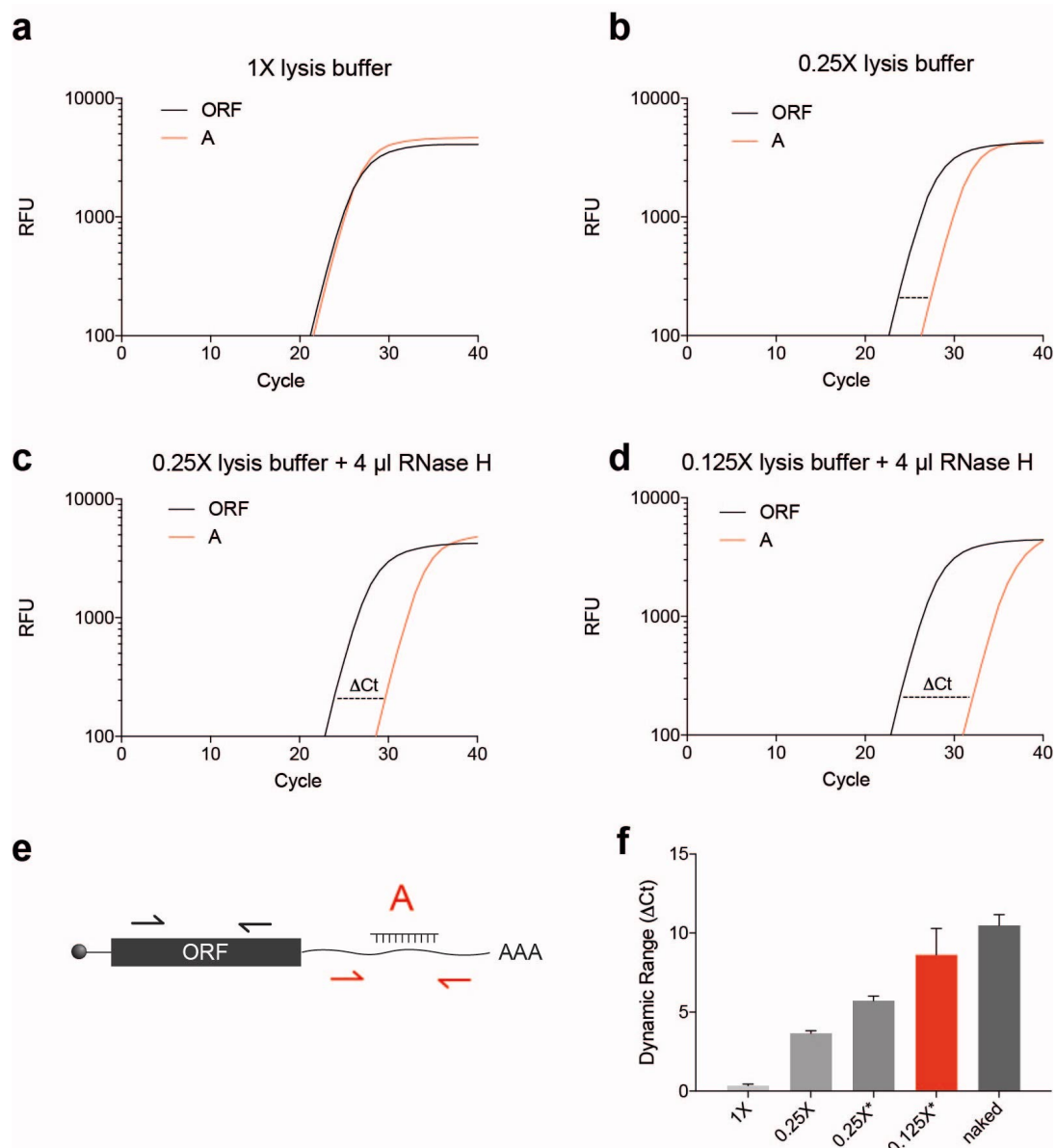


Figure 3.3 | Dilution of the lysis buffer immediately prior to RNase H addition improves dynamic range. a,e, In the original lysis buffer RNase H was incapable to induce a meaningful Δ Ct shift between a control primer pair (ORF, black) and a target primer pair (A, red) on RNA from non-crosslinked lysate. **b,c,d,** Dilution of the lysis buffer through the addition of NEB RNase H buffer improved the enzyme's ability to cleave duplexes, as observed by an increase in Δ Ct. **f,** 0.125X (1:8 dilution) original lysis buffer concentration was chosen for subsequent experiments which approached the cutting efficiency observed for naked RNA. For simplicity only one replicate is shown in the qPCR amplification plot. The read out was performed in technical triplicates. RFU = relative fluorescence units, *= 4 μ l RNase H, error bars in (f) denote standard deviation.

3.3.4 Stepwise Analysis Pipeline Used To Calculate TRACE Score

To compare RNase H protection by miRISC across a candidate MRE, we designed the following analysis pipeline. For each MRE, WT and miRNA-KO cell lysates are processed in the presence or absence of a corresponding ASO (Fig. 3.4a). Each of the four samples undergoes amplification with control primers and target primers (Fig. 3.4a). The abundance of the MRE hybridisation site is calculated in each sample through the transformation of ΔC_t values (target primer C_t minus control primer C_t) (Fig. 3.4b). To account for differences in the abundance of the RNA, WT and miR-184 KO expressions are first normalised to their respective controls without ASO (Fig. 3.4c). Finally, the cut samples (+ ASO) are normalised to the WT +ASO sample (which itself becomes 1) to define a **Relative Protection Score (RPS)** for every MRE in WT and miR-184 KO conditions in the presence of the ASO (Fig. 3.4d). To facilitate data analysis we further defined a **TRACE Score** as the RPS of the WT sample divided by the RPS of the miR-KO (Fig. 3.4e). Thus, the lower the RPS in the KO (expected with loss of miRISC) the larger the TRACE Score.

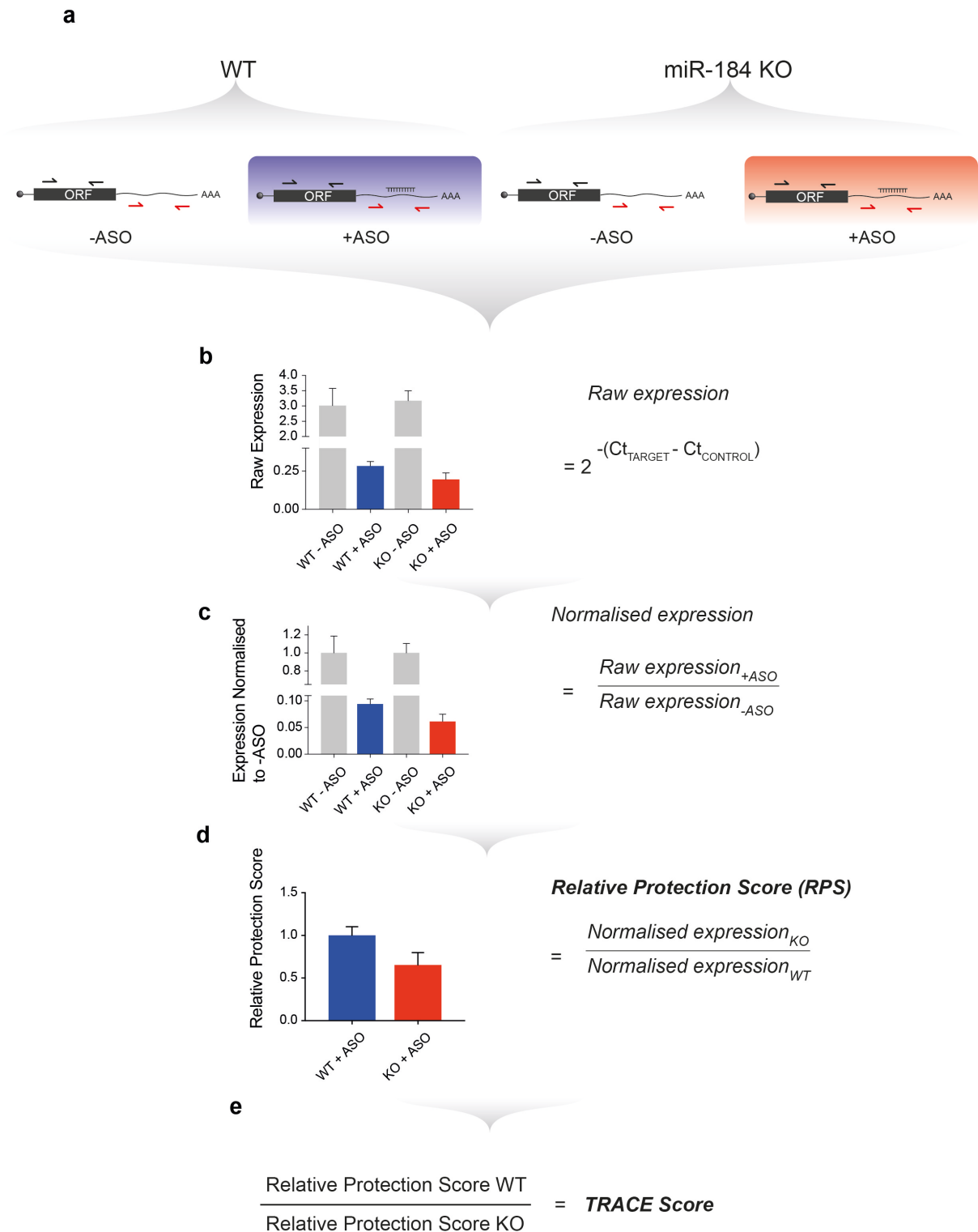


Figure 3.4 | Stepwise calculation of relative protection score and TRACE Score. **a**, Four samples are processed for each MRE: WT minus ASO, WT plus ASO, KO minus ASO and KO plus ASO. Red primers are target primers spanning the hybridisation site. Black primers are control primers detecting an upstream region in the ORF. **b**, Transformation of the ΔCt between target and control primers generates the raw expression for all four conditions. **c**, To account for differences in transcript abundance, WT and miR-184 KO samples are

normalised to their respective no ASO control (set to 1). **d**, Finally, the ASO cut samples are normalised to WT (set to 1) to define a **Relative Protection Score (RPS)** for both WT and KO. **e**, The TRACE Score of any MRE is defined by dividing the RPS of WT by the RPS of KO. The lower the RPS in the KO, the bigger the TRACE Score.

3.3.5 Proof Of Concept – miR-TRACE reflects miRISC binding at *crok* MRE

Drosophila miRNA-184 is one of the most abundant miRNAs in S2R+ cells (Ruby et al., 2007). Since we formulated TRACE as an indirect read out of miRISC abundance at an MRE of interest, we first evaluated the assay on cells expressing four different miR-184 levels: 1) WT S2R+, 2) miR-184 KO S2R+, 3) miR-184 KO S2R+ transfected with 500 ng miR-184 genomic rescue fragment and 4) miR-184 KO S2R+ transfected with 1000 ng miR-184 genomic rescue fragment (Fig. 3.5a). miRNA levels were confirmed by Taqman qPCR and displayed 93 % loss in the KO, 237 % overexpression in the 500 ng and 776 % overexpression in the 1000 ng KO plasmid transfection (Fig. 3.5a).

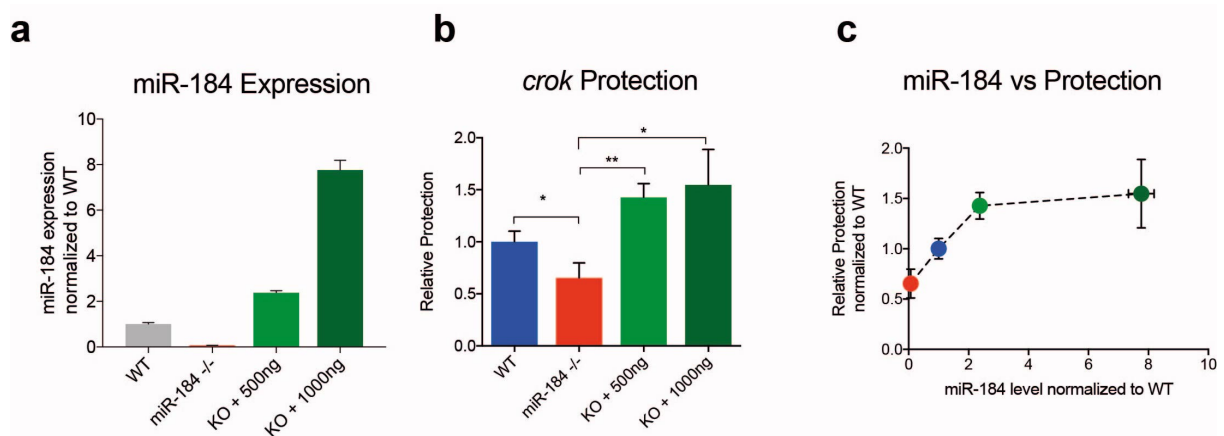


Figure 3.5 | miR-TRACE detects miRISC occupancy at *crok* MRE. **a**, Four conditions with different miR-184 levels were used to test the TRACE assay: WT S2R+ cells, miR-184 genomic KO S2R+ cells, KO cells transfected with 500 or 1000ng miR-184 genomic rescue plasmid. **b**, For the *crok* MRE, the **Relative Protection Score (RPS)** was significantly lower in miR-184 KO cells, reflecting a loss of the RISC complex at this site. Overexpression of miR-184 from a genomic rescue fragment reversed this effect and restored protection. **c**, The relationship between miR-184 abundance and RNase H protection appears linear up to 2.5x of the WT amount and plateaus off thereafter. n=3, error bars denote standard deviation, * $P < 0.05$, ** $P < 0.005$. Statistical significance was calculated using student's t-test.

As previously mentioned, the predicted canonical miR-184 MRE network encompasses at least 77 different MREs in S2R+ cells (Qianxin Wu, *manuscript under consideration*). We decided to test our system on a miR-184 target site embedded in the 3'UTR of the *crooked* (*crok*) mRNA. This MRE carries a predicted 8mer seed, thus belonging to the strongest group of MREs (Grimson et al., 2007). Furthermore, it has been functionally validated and shown to confer robust repression of *crok* mRNA (Qianxin Wu, *manuscript under consideration*). Based on these observations, the *crok* MRE should be occupied by the miRISC complex in WT cells and be refractory to RNase H cleavage (protected site). On the contrary, in miR-184 KO cells the RISC complex can no longer be targeted to this MRE, which should result in an increase in RNase H cleavage (unprotected site, decreased RPS). To test this, we measured the activity of a 25mer ASO centered on the 8mer seed in WT and miR-184 KO cells. Indeed, we observed a marked reduction in the RPS in the KO sample by 35% (Fig. 3.5b). We reasoned that if this effect is specific to miR-184 miRISC association, miR-184 transfection into miR-184 KO cells should restore MRE protection. As expected, transfection of a miR-184 genomic rescue fragment into the miR-184 mutant cells at concentrations of 500 ng and 1000 ng revealed restoration of the RPS, even exceeding WT levels, consistent with higher miR-184 expression in the rescue (Fig. 3.5b).

To assess the correlation between miRNA levels and RPS, we plotted miR-184 levels against RPS values. Interestingly, we noted a linear relationship only up to ~2.5x of the WT miR-184 levels (Fig. 3.5c). The RPS appears to plateau after this point as transfection with 1000 ng rescue fragment does not lead to a proportionate increase in RPS, suggesting all *crok* MREs may have been saturated at this point (Fig. 3.5c). From this pilot experiment we concluded that the TRACE assay reliably reflects RISC binding at a previously validated MRE.

3.3.6 Stepwise analysis pipeline for TRACE using *rp49* as normalisation factor

Prior to expanding the training data set and validating our method on more targets we tested our system using a different means of normalisation. Previously, we employed intragenic control primers to calculate the signal difference relative to the primer pair spanning the ASO cut site. Although this approach should exhibit superior sensitivity, it relies on the assumption that no further RNA degradation takes place post-cutting. In theory, this should be a valid assumption since RNase H cutting takes place under stringent conditions and in the presence of broad-spectrum RNase inhibitors. However, if further degradation takes place following RNase H cleavage, a possible consequence is the removal of the intragenic control primer binding site, leading to a loss of dynamic range.

To mitigate this risk we decided to replace upstream control primers with primers detecting a non-targeted, unrelated RNA. On the basis of minimal perturbation in transcript level between WT and miR-184 KO (data not shown), we chose the housekeeping gene *rp49* although other housekeeping genes e.g. *hsp83* would also be suitable (Fig. 3.6a). Employing a housekeeping gene as the normalisation control harbors two additional advantages. First, while performing the cutting experiment one can readily measure the global fold change of transcript levels between WT and miRNA KO cells (itself a commonly used readout for miRNA target identification). Second, the number of primers required is reduced to half since normalisation now employs a universal housekeeping gene.

In the sample data provided here, the minus ASO controls provide information on the relative transcript abundance. In the example provided, *crok* is 2.5-fold upregulated in miR-184 KO cells (Fig. 3.6b). To derive the RPS and TRACE score, we used a similar approach to intragenic normalisation, namely setting the minus ASO samples

to 1 and normalizing the plus ASO samples accordingly (Fig. 3.6c). This permits a direct comparison between the RPS of WT and miR-184 KO cut samples (Fig. 3.6d).

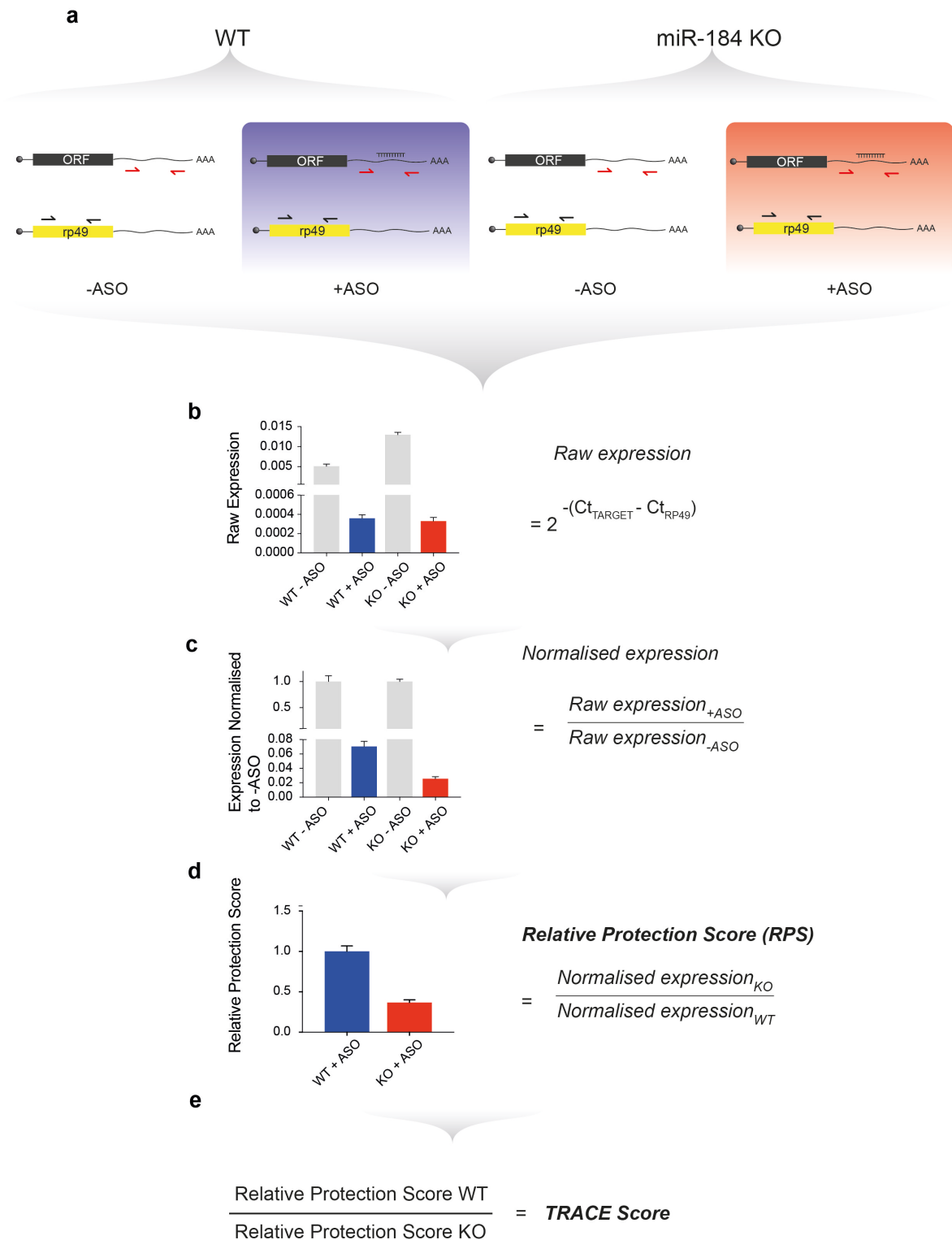


Figure 3.6 | Alternative calculation of TRACE Score using *rp49* normalisation. **a**, Four samples are processed for each MRE: WT +/- ASO and KO +/- ASO. **b**, Transformation of the ΔC_t between the *rp49* and target primers generates the raw expression for all four conditions. **c**, The WT and miR-184 KO data is then normalised to its respective no ASO control (set to 1). **d**, Finally, the ASO cut samples are normalised to WT (set to 1) and the difference in expression between WT and KO reflects the relative protection of the MRE.

When *rp49* was used as the normalisation control, comparing WT and miR-184 KO protection of the *crok* MRE revealed a significant reduction in the KO RPS by 63% (Fig. 3.7a). Although transfection of the miR-184 genomic rescue fragment into KO cells restored the RPS in a dose-dependent manner, in this case WT levels were not surpassed (Fig. 3.7a). Comparing the two normalisation methods, we found they both displayed congruent reduction trends in protection levels from WT to KO (Fig. 3.7b). Given the higher sensitivity observed in the *rp49* normalisation, as well as the advantages of using the housekeeping gene outlined above, we decided to employ this strategy for further studies.

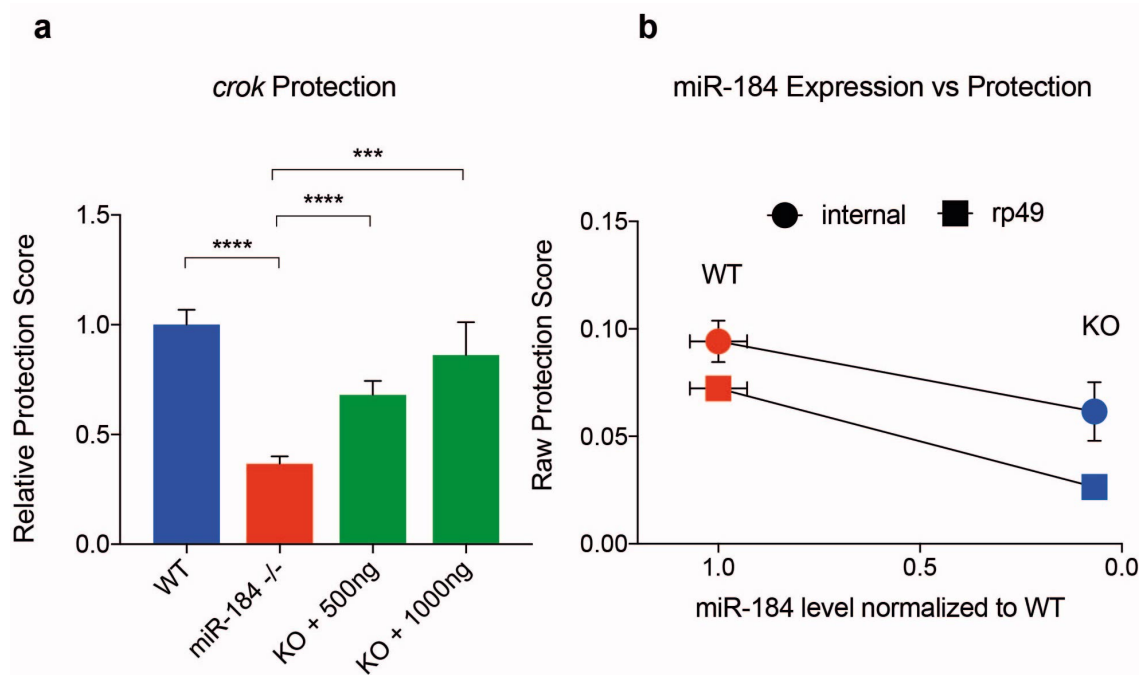


Figure 3.7 | Normalisation using *rp49* displays superior sensitivity over intragene normalisation. **a**, RPS is significantly reduced in miR-184 KO cells. Transfection with miR-184 rescue fragment restores protection levels in a dose-dependent manner to wild type levels. **b**, Both internal and *rp49* normalisation methods display similar loss of protection with reduced miR-184 levels. $n \geq 5$, error bars indicate standard deviation, $***P < 0.0005$, $****P < 0.00005$. Statistical significance was calculated using student's t-test.

3.3.7 Generation of a training dataset

We next sought to validate our technique further by expanding the training data to include eleven additional MREs providing us with a total of twelve candidate MREs. Wu *et al.* have investigated the primary target network of miR-184 in *Drosophila* S2R+ cells (*manuscript under consideration*). They performed multiplex MRE mutagenesis using the CRISPR/Cas9 system to disrupt all putative miRNA::target interactions underlying the predicted miR-184 MRE network. Consequently, if an MRE is repressed by miR-184, the transcript abundance should increase in the miR-184 KO. Briefly, by measuring levels of the unique deletion patterns (UDP) of paired genomic DNA and cDNA they calculated MRE scores according to the following formula:

$$\text{MRE score} = (x^{KO}/y^{KO})/(x^{WT}/y^{WT})$$

where x = cDNA and y = gDNA levels.

MREs have henceforth been divided into three score groups. The *high* MRE score group consists of MREs that confer strong repression upon their targets. The *medium* group harbours MREs that lead to more moderate yet significant repression. The *low* score group contains MREs that are virtually inactive. To validate the TRACE assay we chose to test four targets from each group comprising a training set of twelve targets. We based our sub-selection of MREs encoded in mRNAs with similar FPKM values to *crok* (RNAseq data not shown), and displaying an even coverage across

the range of MRE scores (1-22) (Fig. 3.8). For the *high* group we chose MREs in CG12880, *crok*, CG10217, *pck*. In the *medium* group we included *gli*, *lqfR* (Epsin), *cont* and *nrx-IV*. The *low* group comprised CG8010, *yrt*, *l(2)35Bg* and *cdc5*.

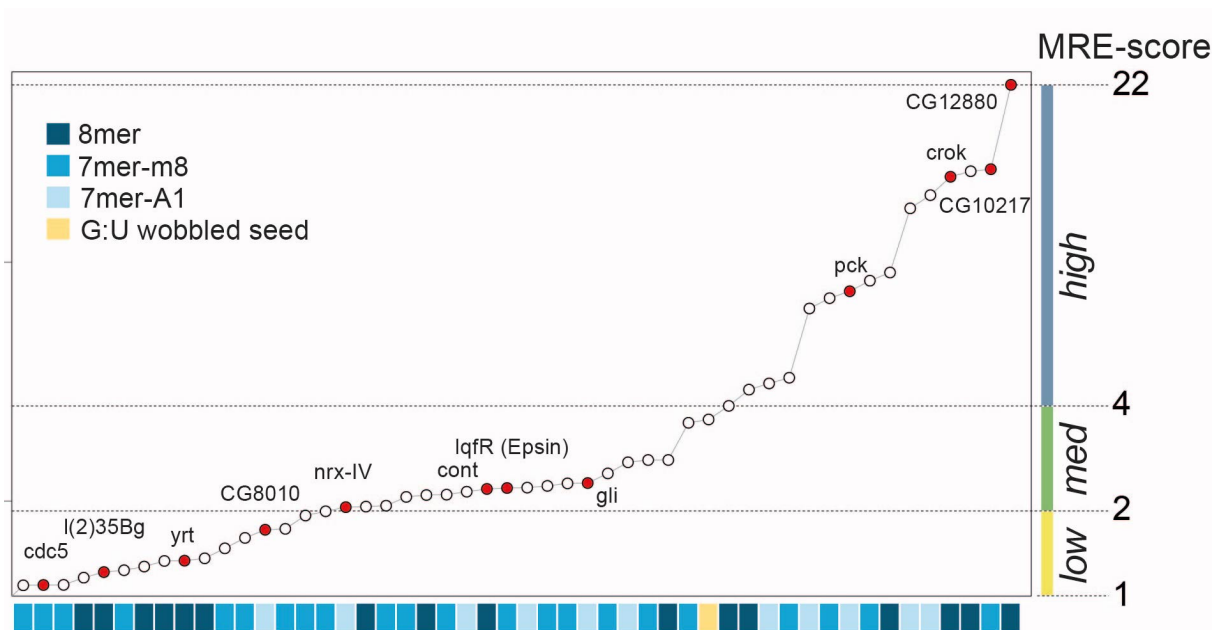


Figure 3.8 | Selection of miR-TRACE training data. The miR-184 target network is ranked according to individual MRE scores. The colour code at the bottom indicates the MRE seed type. To test the TRACE assay twelve MREs (named) were selected on the basis of similar FPKM values to *crok* (not shown) with an even distribution of all scores. The turquoise group contains MREs with scores between 4 and 22. The green group contains medium MRE scores between 2 and 4. The yellow group comprises low MRE scores less than 2. Data from (Qianxin Wu, *manuscript under consideration*)

3.3.8 miR-TRACE analysis closely reflects MRE score

We next conducted a thorough assessment of protection changes between WT and miR-184 KO at these twelve MREs. As expected, the *high* MRE score group revealed the largest RPS reduction between WT and miR-184 KO: 63% for *crok*, 55% for *pck*, 70% for CG12880 and 57% for CG10217 (Fig. 3.9a). The *medium* group displayed more moderate reductions: 29% for *gli*, 28% for *cont*, 50% for *nrx-IV*, and 14% for *lqfr(epsin)* (Fig. 3.9b). In contrast, the *low* group exhibited no significant changes in RPS values (Fig. 3.9c).

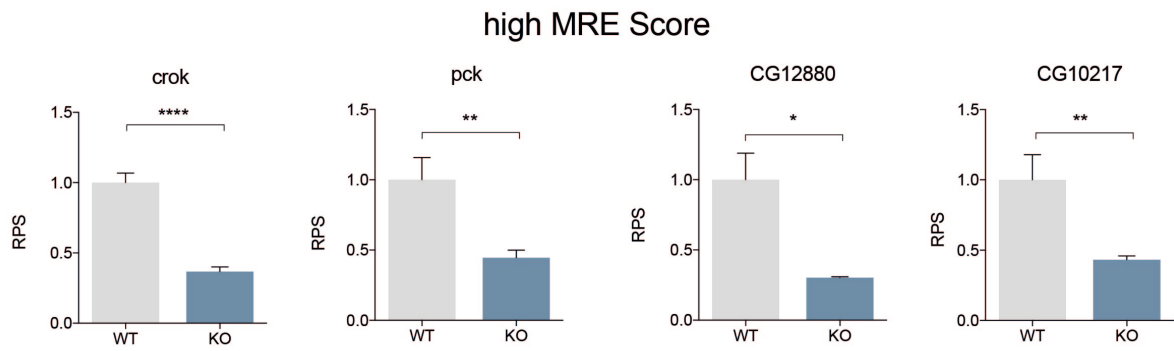
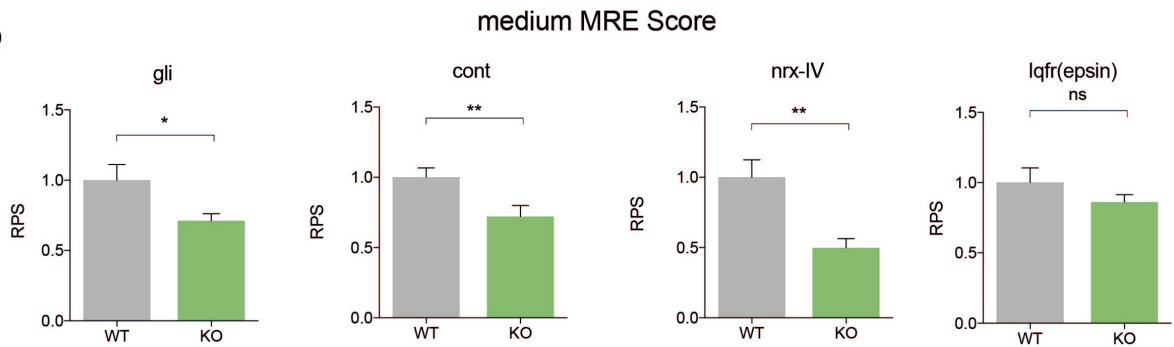
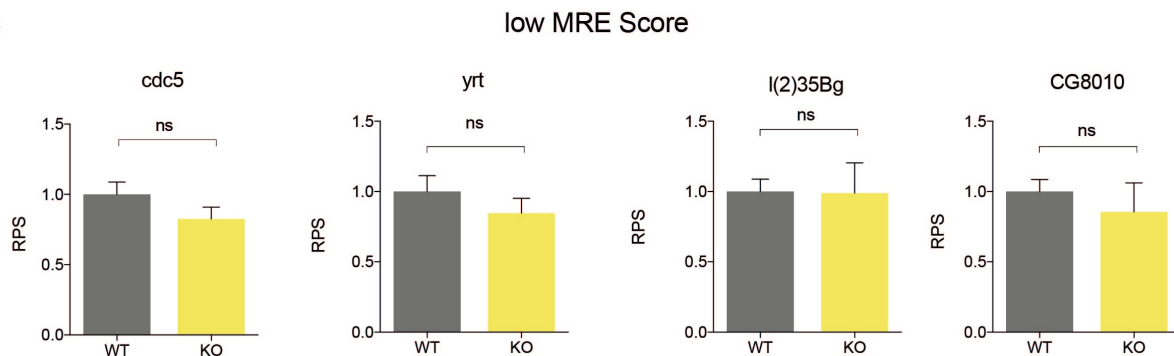
a**b****c**

Figure 3.9 | miR-TRACE scores of twelve MREs reflect their respective MRE class.

a, The high MRE score group displays strong reductions in their RNase H protection (RPS) between WT and miR-184 KO (turquoise box plots). **b**, The medium MRE score group exhibits moderate, yet in most cases significant reductions in their protection (green box plots). **c**, The low group shows no significant change in the RNase H protection. Error bars = standard deviation. * $P < 0.05$, ** $P < 0.005$. ns = not significant. $n = 3$, statistical significance was calculated using student's t-test. The grey scale bars represent the WT Relative Protection Score.

Calculating the mean average of the TRACE scores for the *high*, *medium* and *low* MREs revealed significant differences between these three groups (Fig. 3.10a). The

high group has a TRACE score mean of 2.65, which is 1.78 fold higher than the *medium* mean (1.49) and 2.32 fold higher than the *low* group mean (1.14) (Fig. 3.10a).

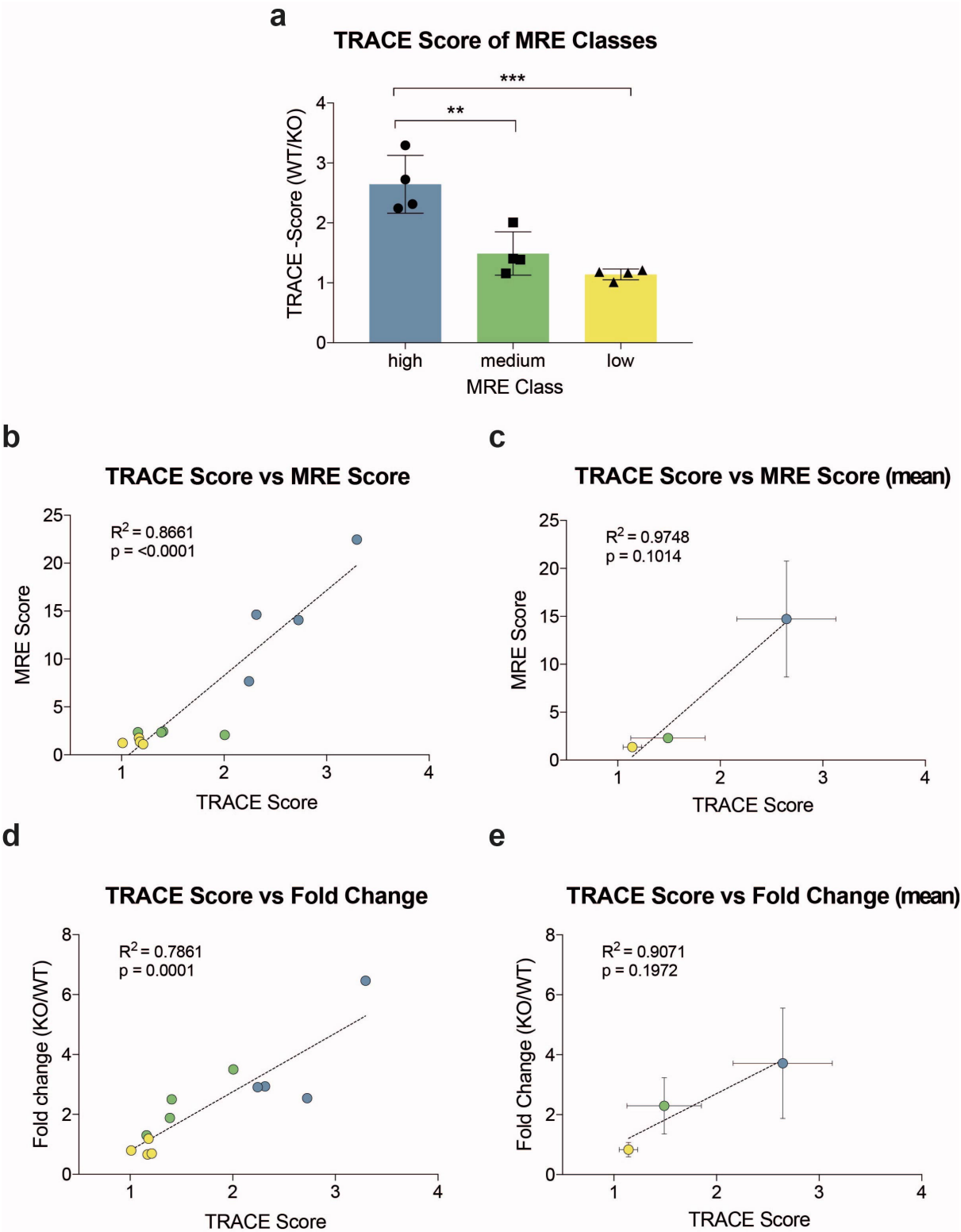


Figure 3.10 | The TRACE score metric performs well in comparison to the MRE score and KO/WT fold change. a, The TRACE score (WT protection/ KO protection) of the three

MRE classes falls into three distinct groups. **b**, The direct comparison of individual TRACE with individual MRE scores reveals a positive linear relationship with a significant correlation. **c**, Taking the mean average of the high, medium and low groups displays the same positive linear relationship. **d**, Assessing individual TRACE scores vs. the fold change (KO/WT) of each MRE exhibits a significantly positive relationship. **e**, Calculating the mean average for both TRACE score and fold change (KO/WT) maintained the linear relationship albeit no longer significant. Error bars = standard deviation, *** $P < 0.0005$, ** $P < 0.005$. Statistical significance for a) was calculated using student's t-test. Linear regression was calculated using Graphpad Prism.

We next sought to evaluate the performance of TRACE score head-to-head against the MRE score metric and KO/WT transcript fold change (a commonly used metric to measure the potency of miRNA::target interactions). Plotting the individual TRACE scores of our twelve MREs against their respective MRE scores revealed a significant linear correlation (linear regression, $p = <0.0001$, $R^2 = 0.87$) (Fig. 3.10b). Taking the mean of each class and plotting TRACE score versus MRE score resulted in the same trend (linear regression, $p = 0.1014$, $R^2 = 0.97$) (Fig. 3.10c). We next assessed how the TRACE score performed against fold change (miR-184 KO transcript levels/ WT levels) and observed a significant linear correlation (linear regression, $p = 0.0001$, $R^2 = 0.79$) (Fig. 3.10d). Averaging these scores into *high*, *medium* and *low* groups also exhibited a positive correlation albeit with a non-significant p-value (linear regression, $p = 0.1972$, $R^2 = 0.91$) (Fig. 3.10e). From these data we concluded that miR-TRACE is a reliable method to discern active from inactive MREs.

3.4 Discussion

Here, for the first time, we have directed the principles of *in vivo* crosslinking and RNase H protection analysis towards the study of miRNA::target interactions *in situ*, without any external intervention. The TRACE technique has successfully recapitulated previously functionally validated MREs and performed competitively against other miRNA target identification read outs, such as transcript fold change (KO/WT) and the CRISPR-Cas9 generated MRE score (Qianxin Wu, *manuscript under consideration*).

3.4.1 Optimisation of RNase H cleavage conditions

Interestingly, our evaluation of the only published protocol (Simon et al., 2011) employing *in vivo* crosslinking followed by RNase H cleavage revealed that the catalytic activity of RNase H was heavily compromised under denaturing conditions (1X lysis buffer). This is perhaps not surprising, given that biochemical laws should forbid the activity of an enzyme at high concentrations of two detergents present in the lysis buffer: sodium deoxycholate and *N*-laurylsarcosine. To circumvent this problem we tried to dilute the lysate in RNase H reaction buffer (NEB) immediately before addition of RNase H. Using a 1:8 dilution we observed a marked improvement in cutting efficiency and thus dynamic range. We were aware that a reduction in detergent concentration bears the uncertainty of re-formation of non-specific RNA (negatively charged) -protein (positively charged) interactions. However, we reasoned this risk is mitigated as long as the initial cell lysis takes place under stringent conditions (1X lysis buffer), and the lysate is only diluted immediately prior to RNase H cutting. As a more elegant future solution, we considered a buffer exchange using poly(dT) capture beads to isolate all polyadenylated transcripts as an

ideal compromise between stringent lysis conditions and effective enzyme catalysis (Castello et al., 2013b).

3.4.2 TRACE reflects previously calculated MRE Scores

We chose twelve previously characterised MREs to serve as test beds to evaluate the TRACE assay. This group included four very active MREs (*high MRE class*), four MREs with moderate activity (*medium MRE class*) and four MREs with marginal activity (*low MRE class*), displaying no increase in cDNA level upon seed mutation as shown by Wu *et al.* (*manuscript under consideration*) (Fig. 3.8). We generated TRACE scores for each MRE and found that they closely reflect the MRE scores. We further compared our method against fold change of KO transcript over WT and similarly observed a strong correlation. All MREs with a high TRACE score also displayed a high fold change of KO/WT transcript levels (Fig. 3.10 b,c). MREs with low TRACE scores (close to 1) revealed no significant fold change (Fig. 3.10 d,e).

3.4.3 TRACE Score as a new metric for miRNA::target interactions

The kinetics of various types of miRNA::target interactions (8mers to non-canonical sites) can be extremely varied. Hence, an accurate metric for scoring these interactions necessitates a broad dynamic range in which different miRNA targeting events are compared. Between the limited number of MREs analysed, the TRACE scores ranged from 3.29 (CG12880) to 1.01 (l(2)35Bg). All MREs of the highly active group displayed a score of 2.31 or larger, suggesting 2.31 could be set as a threshold above which to consider an MRE active. However, with the restricted number of MREs considered, a universal threshold is more difficult to gauge. More MREs will need to be assessed in order to confidently define such a cut-off.

Although our data should reflect the loss of miRISC complex at an active MRE, we were surprised that the difference in protection score between WT and KO was not

greater. Even the most potent MREs displayed only a 2-3 fold reduction in their protection score (Fig. 3.9a). For example at the potent *crok* MRE, we would expect loss of all miRISC binding, which should be reflected in a large reduction in the protection score. The discrepancy between this expectation and the actual observation could be explained by several reasons:

First, alternative repression mechanisms could compensate for the loss of miR-184 and hence mask miRISC loss at the MRE. Compensation has been abundantly observed for miRNA studies where the removal of a single miRNA frequently engendered no phenotypic consequence due to compensation *via* redundant miRNA family members. However, this explanation is unlikely because miR-184 is not part of a family and its seed region is sufficiently unique amongst other miRNAs expressed in S2R+ cells, ruling out the possibility that a related miRNA compensates for the loss of miR-184 RISC binding at this site. Nonetheless, it remains possible that an unrelated RBP substitutes miRISC binding in the KO, hence obstructing ASO hybridisation leading to a lowered reduction in the protection score. Second, the concentration of *in vivo* regulated MREs is typically higher than that of the corresponding miRISC, suggesting that miRISC binding is inherently limited to a subset of its available targets (Denzler et al., 2014). Third, physical binding between miRISC and MREs is transient (e.g. AGO2 in mouse remains bound to MRE for ~5 min at 37°C), suggesting it is possible that only a small fraction miRISC::MREs is crosslinked at any one time (Salomon et al., 2015). Fourth, upon miRISC binding target mRNAs are often degraded rapidly, indicating that only a small fraction of miRISC-bound mRNAs are assayed. These four accounts provide an explanation for the fact that even the most active MREs exhibited only a 'modest' deprotection (compared to the theoretical expectation) upon miR-184 KO as assayed by TRACE.

Nevertheless, the fact that TRACE detects deprotection reliably makes it a valuable and useful technology.

3.4.4 Technical considerations for miR-TRACE

In its current implementation, the TRACE assay uses 1% formaldehyde *in vivo* crosslinking for 10 min at room temperature. While this dosage and length has been successfully used in other short RNA capture approaches (Hassan et al., 2013; Simon et al., 2011), the efficiency of chemical crosslinking has recently been challenged (Gavrilov et al., 2015). Evidence accumulated that formaldehyde crosslinking efficiency can drastically vary and oftentimes requires extensive optimisation, causing it to be dubbed a “black box” (Gavrilov et al., 2015). Recently, the field of small non-coding RNAs has seen the development of a host of techniques taking advantage of ultraviolet (UV) crosslinking (254 nm) as opposed to chemical crosslinking (Chi et al., 2009; Hafner et al., 2010; Helwak et al., 2013). UV crosslinking requires direct contact (minimal distance) between RNA and proteins and does not crosslink proteins to each other, rendering this method a ‘cleaner’ approach (Pashev et al., 1991). Similarly, for our purpose of mapping miRISC associated MREs, the refined nature of UV crosslinking may be advantageous. However, the analysis of UV crosslinked material is hampered by two limitations. First, UV crosslinking is notoriously inefficient; as little as 5% of RNA-protein interactions appear to be efficiently crosslinked (Pashev et al., 1991). This poses a problem for the TRACE assay. If only 5% of the interactions are captured, the majority of MREs will still be accessible to RNase H cutting, leading to a false negative result (high cutting despite physiological miRISC binding). Second, it is difficult or in most cases impossible to completely reverse UV induced crosslinks. Even following an extensive proteinase K treatment, peptide stumps remain associated to the RNA (Chi et al., 2009). While this phenomenon is exploited in some

applications, it would hinder the passage of reverse transcriptase required in our case. Consequently, incomplete cDNA strands would be generated. If the primer-binding site is not present in cDNA due to premature truncation, even a non-cut RNA would not be detectable, leading to ample false negative results.

Interestingly, structural and kinetic studies of human AGO2 have shown that base pairing beyond miRNA nucleotide (nt) position 16 is not essential for target binding (Wee et al., 2012). This suggests, in principle, that those nucleotides may be accessible to ASO hybridisation even if the seed and the upstream bases have been crosslinked to AGO. However, even if target RNA bases beyond position 16 of the miRNA were available for hybridisation, it is likely that steric hindrance of the crosslinked miRISC complex (AGO molecular weight alone is 105 kDa) would prevent RNase H access. To avoid this uncertainty in the future, shorter ASO could be employed although this may come at a loss of cleavage efficiency (see chapter 4). Another solution to guarantee that only the AGO-bound seed of the target MRE can be cut, would entail the use 2'O-methyl RNA/DNA chimeras (Inoue et al., 1987; Rukov et al., 2015). Presently, we use 25mer DNA oligonucleotides where every single RNA::DNA base pair acts as substrate for RNase H. In contrast, 2'O-methyl RNA/DNA chimeras can hybridise to the RNA target, recruit RNase H and yet only elicit cleavage of the target RNA that is base paired to the unmodified DNA part of the chimera. Consequently, a chimeric oligonucleotide could be designed which incorporates DNA bases corresponding to the seed region, flanked by 2'O-methyl RNA nucleotides on each side. This would constrict the cutting to the seed only, with the additional benefit of higher affinity binding to the seed-adjacent regions.

In its present state, TRACE cannot distinguish between secondary structure interference and RBP occlusion. One solution to address this issue in future implementations is to assay in parallel naked or deproteinised RNA isolated from WT

and miRNA KO cells. If under these conditions an inactive MRE still appears inaccessible, it suggests RNA secondary structure is preventing miRISC access. Another option to incorporate the assessment of secondary structure contribution would be to include bioinformatics structure prediction of MRE loci. Several such algorithms already exist (Maragkakis et al., 2009) and could be readily integrated into the miR-TRACE analysis platform.

3.4.5 Comparison of miR-TRACE to other miRNA target MRE identification and validation techniques

TRACE offers an attractive and facile alternative to existing target identification and validation platforms. Arguably, the simplest target identification technology, profiling approaches measuring mRNA fold change in wild type and miRNA knockout cells have the caveat of lacking the ability to discriminate between direct (primary) and indirect (secondary) miRNA targets. Furthermore, modulating miRNA levels through overexpression frequently engenders substantial perturbations of the transcriptome, thus potentially masking *bona fide* target sites from detection (Khan et al., 2009). TRACE circumvents these caveats as the assay, per definition, quantifies differential miRISC occupancy only at primary targets. Moreover, TRACE works on native cells whose homeostasis has not been affected through transfection.

A commonly used approach to validate miRNA-MRE interactions is the heterologous target-MRE-sensor (Kuhn et al., 2008). In this setup, naïve cells are transfected with a reporter constructs that carries the MRE or in some cases entire 3'UTRs of interest, whose output is measured in the presence or absence of a miRNA. However, the main limitation of this system is its inherently artificial configuration. miRNA::target interactions are probed at non-physiological stoichiometries resulting from ectopic delivery or artificial promoters. Furthermore, these synthetic constructs ignore important contextual features such as RBP occupancy, RNA secondary

structure or alternative poly-adenylation (Kertesz et al., 2007; Nam et al., 2014). In contrast, TRACE assesses miRNA binding at endogenous MREs *in situ*. Furthermore, TRACE is not limited to transient systems as it does not require transfection of foreign material into any of its samples. Its only requirement is that sufficient input material can be gathered, which is feasible in cell culture systems and most *Drosophila* tissues.

Genome editing of predicted MREs followed by RT-qPCR has recently been proposed to uncover functional MREs *in situ* (Bassett et al., 2014). This approach has hitherto been the only method to assess whether a miRNA::target interaction is relevant in an endogenous context, under cellular homeostasis. However, this approach remains labour intensive, expensive and presently difficult to scale up. Furthermore, it is still confined to predicted targets, whilst TRACE can be carried out in an exploratory manner (see chapter 4).

3.4.6 Final Conclusions

I have here presented the development of TRACE, a novel miRNA::target binding validation platform which reliably distinguishes active from inactive MREs *in situ* without any external intervention. Hence, in principle, any accessible tissue or cell line can be readily assayed by TRACE. We assessed the assay performance against two other metrics, MRE score and target transcript fold change and found a strong positive correlation. However, the dataset of 12 MREs is presently too limited to define a reliable universal TRACE-score threshold above which an MRE can be considered active. Nevertheless, within the tested MREs a clear cut-off value for active and inactive MREs could be inferred, suggesting that this strategy has the potential to become a useful addition to the small RNA exploratory toolkit.

Chapter 4: Design and Implementation of high-throughput *in vivo* detection of miRNA-Target Interaction Events by Massively Parallel Sequencing (miR-TRACEseq)

4.1 Introduction

Following the development of TRACE to decode individual MREs, we next explored the scalability of this technology, with the goal to assess multiple *cis* regulatory elements in parallel. Our candidate TRACE approach described in the previous chapter relied on 25mer ASOs specifically targeted to individual MREs embedded in target mRNAs. In this instance, MRE-specific relative protection scores (RPS) reflecting the local miRISC occupancy are generated from individual experiments (i.e. one ASO per reaction) (Fig. 4.1a). Although this approach is feasible for validating small sets of MREs, the size of most miRNA target networks calls in question the cost efficiency of such approach, especially during the exploratory phase of a project. Although increasing the throughput of experimental platforms frequently comes at a cost of resolution, we wondered whether we could assay more than one MRE in parallel, without compromising individual dynamic range. If successful, MRE protection could be measured in a multiplexed manner, thereby reducing the cost and time associated with exploratory studies. We envisioned two possible avenues to implement this: *i*) pooling full-length (25mer) oligonucleotides into batches with minimal inter-amplicon off-targets (Fig. 4.1b) or *ii*) using a single universal oligonucleotide that can target any seed-bearing MRE (Fig. 4.1c).

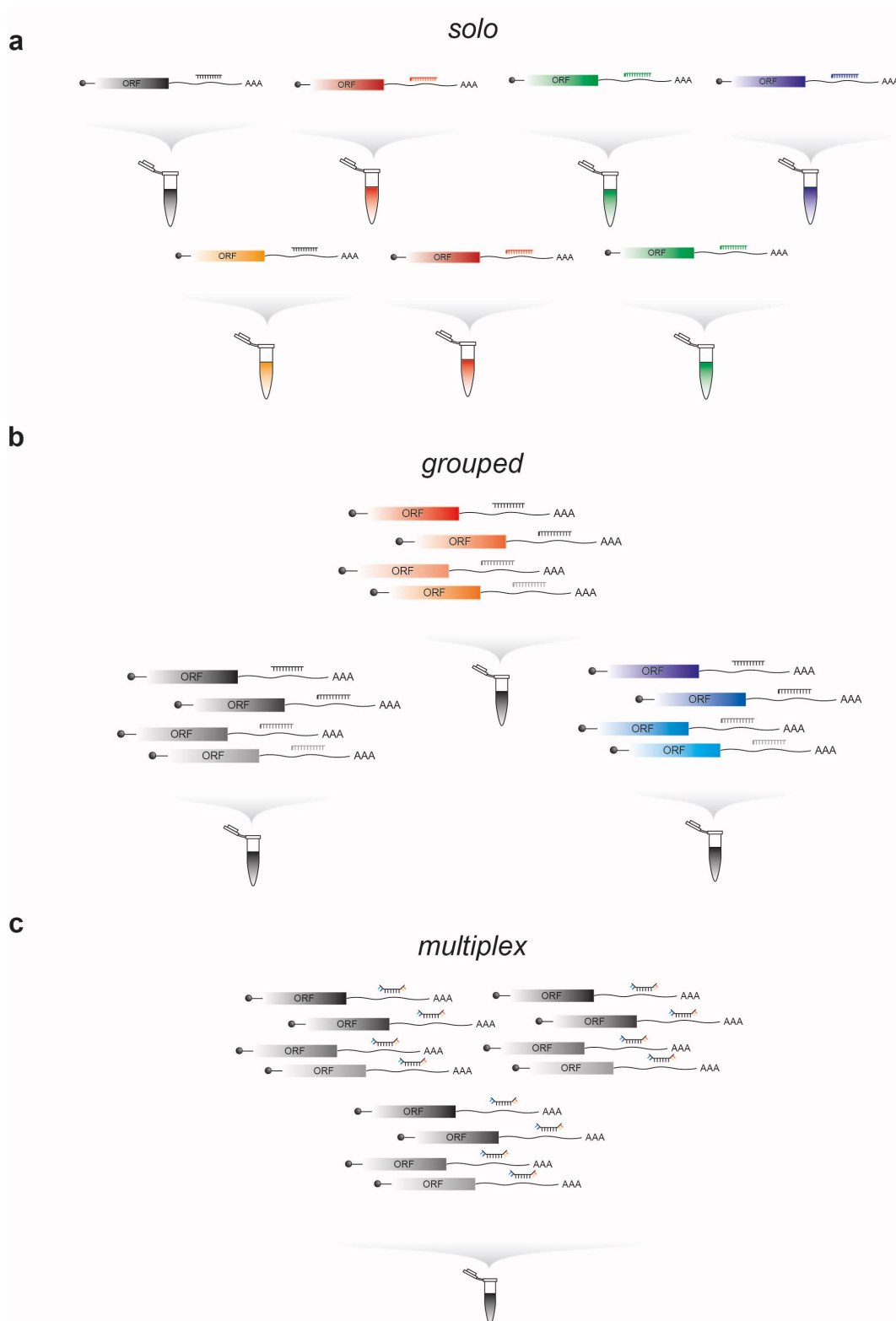
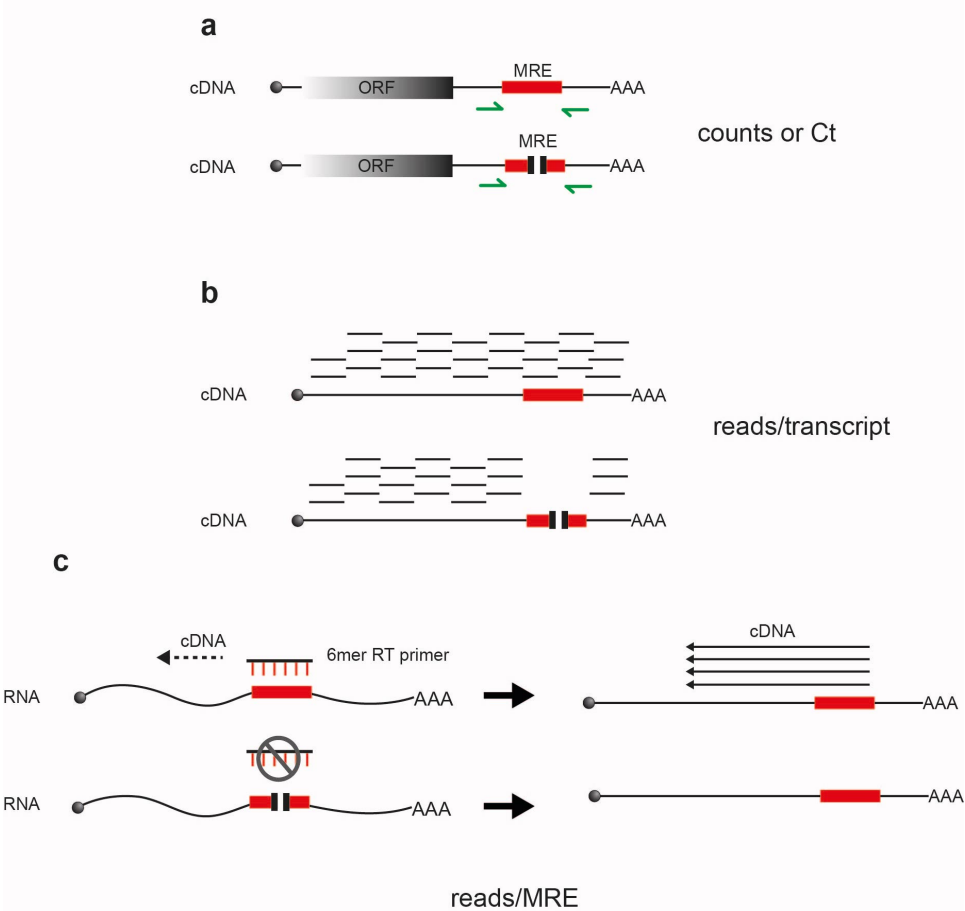


Figure 4.1 | Multiplexing strategies to improve TRACE throughput. **a**, the original TRACE protocol relied on a single ASO per MRE **b**, ASO groups with minimal off-target potential can reduce the number of reactions. **c**, a universal seed-targeting could potentially assess the entire predicted target network of a miRNA simultaneously.

Arguably, all nucleotide-based approaches, including those used for therapy (e.g. antisense inhibition) and biotechnology (e.g. CRISPR-Cas9 or RNAi), have the caveat of off-target effects. The minimum RNA::DNA heteroduplex that can serve as a substrate for *in vitro* RNase H cleavage is 4 base pairs (Donis-Keller, 1979), suggesting that in theory a 25mer can recognise and cleave 22 different tetramers. A previous *in vivo* experiment showed that injecting antisense DNA into *Xenopus* oocytes induces at least partial RNA degradation if as few as 10 consecutive bases are paired (Woolf et al., 1992). Random pooling of 25mers would thus engender widespread unpredictable RNA cleavage rendering specific analysis of on-target events extremely difficult. However, considering that our readout is targeted amplification of ~ 150 bp regions, we reasoned ASO pooling is feasible as long as no off-target events occur between different amplicons. To this end, we designed an algorithm that rationally groups 25mer ASOs into batches with minimal off-target potential.

To explore the second strategy entailing a universal ASO targeting all MREs within a complete network, we took advantage of two factors: *i)* all MREs of the predicted miR-184 primary target network share the same 6mer region corresponding to position 2-7, the seed sequence of mature miR-184. *ii)* the ASO minimum length required to elicit target RNA cleavage is shorter than 25 nt. The critical question was whether a single short ASO spanning only the 6mer MRE seed could serve as a universal RNase H agent. This would enable simultaneous assessment of relative protection scores for all target sites bearing the 6mer MRE seed sequence in a single TRACE assay reaction.

Finally, we sought to also develop a high throughput read out system in addition to multiplexing the biochemistry of the assay. Hitherto, following crosslink reversal and proteinase K treatment, the readout system has been based on targeted amplification using qPCR with primer pairs spanning the cut site (Fig. 3.1 and Fig. 4.2a). While this offers a rapid and simple analysis platform for candidate targets, expanding the TRACE approach to genome-wide studies requires the implementation of a next-



generation sequencing readout.

Figure 4.2 | Increasing the assay throughput *via* multiplexing the readout. **a**, hitherto cutting events have been readout *via* qPCR amplification over MRE of interest. **b**, classical RNAseq could provide unbiased coverage of UTRs but most likely will lack sufficient depth for intra-transcript resolution (i.e. cut site identification). **c**, seed-targeted RT priming initiates cDNA synthesis only at protected MREs, constricting sequencing reads to MREs only thus increasing meaningful sequencing bandwidth.

We first considered a standard RNAseq experiment to count mapped reads over an MRE of interest and compare these between WT and miR-184 KO cells in the presence or absence of an ASO. Albeit an apparently simple solution, a careful evaluation of this approach revealed that even a 400 million NextSeq experiment would almost certainly lack sufficient intra-transcript depth and resolution to map precise cut sites in an unbiased manner (Fig. 4.2b). In a typical RNAseq experiment, sequencing reads tend to map in a scattered fashion to genomic locations which is considered in the standard unit of FPKM (Fragments Per Kilobase of transcript per Million mapped reads) (Fig. 4.2b). However, to quantify cutting events reliably, absolute read counts covering the target region (6mer) are essential. Performing a standard RNAseq experiment and selecting only those counts mapping to the 6mer seed motif will most likely limit TRACE score calculation in terms of insufficient statistical power (low read counts).

To overcome this problem and improve sequencing bandwidth over the 6mer seed motif we developed a fundamentally different selection approach rooted in seed-initiated reverse transcription (RT) (Fig. 4.2c). This concept resembles *primer extension* approaches where a gene-specific RT primer (GSP) is hybridised to the 3' end of an RNA with the goal to elucidate an unknown 5' end of a particular RNA species (i.e. transcription start site) (Carey et al., 2013; Ghosh et al., 1978). In contrast to primer extension, which typically focuses only on a single RNA molecule, we attempted to reverse transcribe in a combinatorial manner multiple RNA species that share a common 6mer motif.

To this end, we sought to take advantage of the fact that all miR-184 predicted targets carry a 6mer complementary region to nucleotides 2-7 of the miRNA (Agarwal et al., 2015). We reasoned that if an MRE is unprotected, ASO hybridisation would enable RNase H-mediated destruction of its seed. Following RNA extraction,

hybridisation of a specific RT primer against this cleaved MRE seed would be prohibited and as a consequence no cDNA synthesis would be initiated at this site. In contrast, a protected MRE that withstands RNase H cleavage would be available for seed specific RT priming and initiation of cDNA synthesis.

As such, the readout was changed from full-length cDNA amplification (reads/counts per MRE) to seed targeted RT priming and sequencing (i.e. only intact 6mer motifs in the RNA can produce cDNA) (Fig. 4.2c). We reasoned that such a seed-targeted sequencing approach should increase the meaningful bandwidth of the sequencing run *via* constricting, in theory, all cDNA synthesis initiation to a single 6mer motif. If successful, we believe this multiplex approach should be amenable to unbiasedly assess not just miRISC occupancy of a predicted primary miRNA target network, but also any 6mer (or longer) RBP motif in the wider context of the transcriptome.

4.2 Chapter Aims

- Implementation of multiplexed TRACE biochemistry *via i)* a grouped 25mer and *ii)* a single universal ASO strategy for parallel MRE protection analysis.
- Assessment of the grouped and single multiplex ASO assay on previously characterised functional and non-functional MREs and comparison of its efficacy to individual 25mer TRACE
- Implementation of a high throughput readout to detect cutting events *via* next generation sequencing (NGS)
- Explore miRISC occupancy of the miR-184 6mer seed in a transcriptome-wide context

4.3 Results

4.3.1 Creating ASO groups streamlines TRACE

We programmed an algorithm for the rational grouping of ASOs with minimised off-target potential. In the original protocol, cutting events were measured by RT-qPCR amplification across an MRE of interest with a typical amplicon length of ~ 150bp. RNA degradation following RNase H cutting has not been observed (see Fig. 3.3, ORF control primers do not exhibit Ct shift). We hence reasoned that any off-target cutting outside the respective amplicons would not engender target degradation and loss of qPCR signal. We thus limited our off-target search to amplicons only.

For the efficient grouping of ASOs, each 25mer was dissected *in silico* into smaller consecutive oligomers using a 'sliding window' function. For example, by setting a 10nt off-target threshold, every 25mer can be separated into 16 different 10mers (10nt ASO_j, Fig. 4.3a). All possible short oligomers were aligned against the full-length amplicon of every MRE in the target network. To provide a working example, if any of the 10mers derived from ASO A have perfect (10bp) complementarity to another amplified UTR B, ASO A and ASO B cannot be assigned to the same ASO group. Instead a new group is created for ASO B. An interaction matrix is then created that defines ASO groups (shaded) that do not bear off-targets to any other group member for 10nt or more (Fig. 4.3a). Desired on-targets are shown diagonally (blue or white dot, e.g. ASO A – UTR A) (Fig. 4.3a). We executed the algorithm with three threshold levels: no off-targets longer than 12nt, 11nt or 10nt (Fig. 4.3b). The 12nt cut-off suggested only 4 groups that would markedly reduce the reaction cost. In this case, the predicted primary target network of miR-184 consisting of 90 MREs could in theory be assessed in as little as 4 reactions, a 22.5-fold efficiency optimisation. However, in line with the findings of Woolf and colleagues (Woolf et al., 1992) we reasoned that 12mer cutting still bears a significant risk of off-target

cleavage. Thus, albeit at a slightly higher reaction number (9 groups instead of 4) we decided to proceed with a 10nt cut-off.

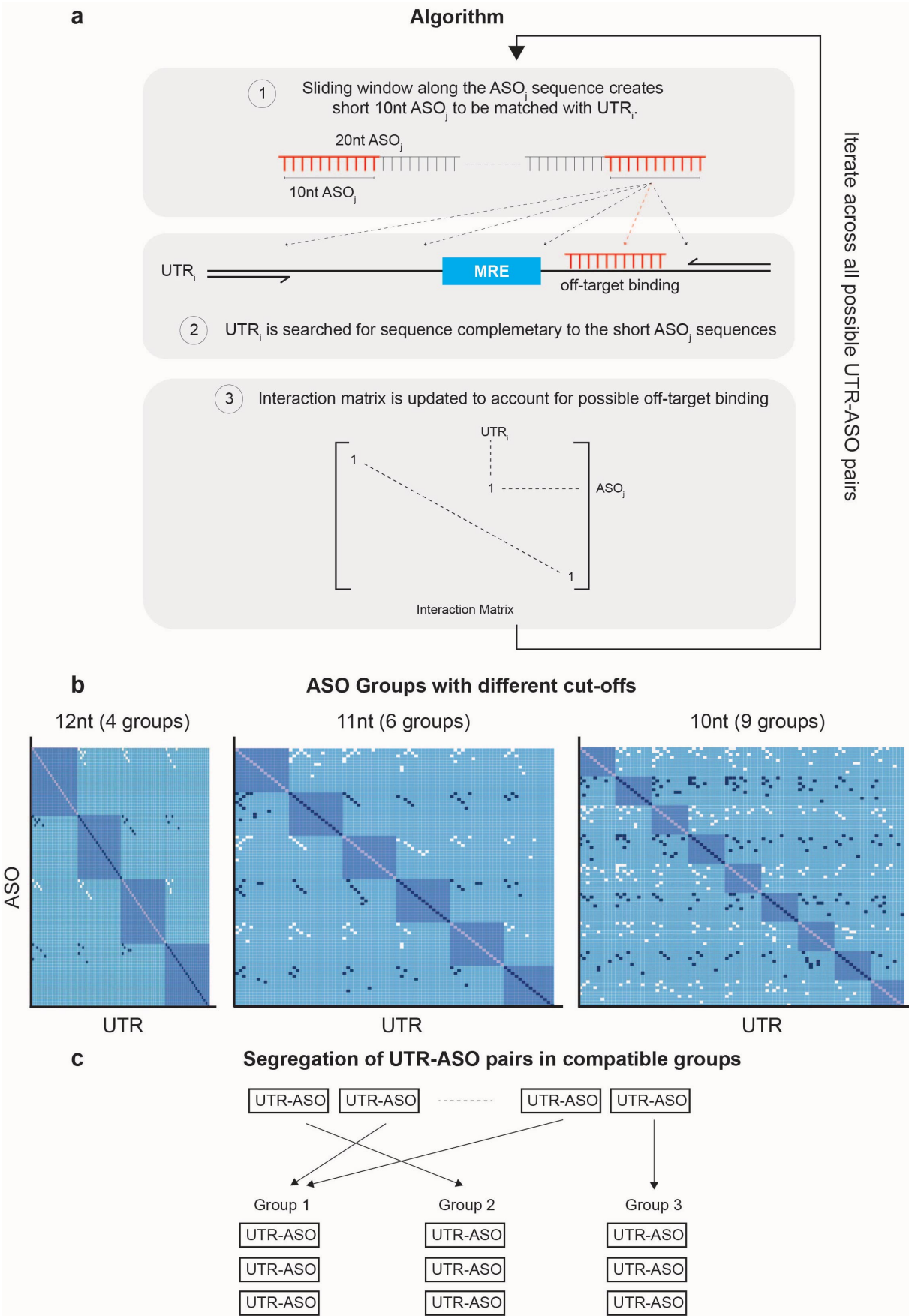


Figure 4.3 | Overview of group selection algorithm. **a**, 1) a sliding window extracts every consecutive 10mer of the full length 25mer. 2) all UTRs (full length amplicon) are searched for complementarity against every 10mer originating from each ASO. 3) An interaction matrix is generated to define ASO groups without off-targets with more than 10bp complementarity. **b**, Results of executing the algorithm with 3 different off-target thresholds (<12nt, <11nt, <10nt). N.B. number of groups increases with more permissive off-targeting criteria. **c**, The ASOs are pooled equimolarly according to the calculated groups.

4.3.2 TRACE on ASO groups maintains TRACE Scores observed in individual assays

To evaluate the efficacy of our grouping strategy, we generated nine equimolar ASO pools according to the algorithm described above, with an off-target limit of 10bp. We tested the same 12 MREs for which we had previously generated TRACE scores using single ASO reactions (*solo* cutting, chapter 3), but this time in their respective ASO pools (i.e. in the presence of other ASOs). Our analysis revealed that the *high* MRE Score group continues to display the largest reductions in relative protection scores (RPS) from WT to miR-184 KO: 42% for *crok*, 64% for *pck*, 73% CG12880, and 77% for CG10217 (Fig. 4.4a). The *medium* MRE group exhibited more moderate yet significant reductions in the RPS: 23% for *gli*, 48% for *cont*, 55% for *nrx-IV*, and 37% for *lqfr(Epsin)* (Fig. 4.4b). Interestingly, the *low* group revealed a decrease for all four MREs although the result was statistically significant for only two of them: 48% for *yrt*, 30% for *l(2)35Bg* (Fig. 4.4c).

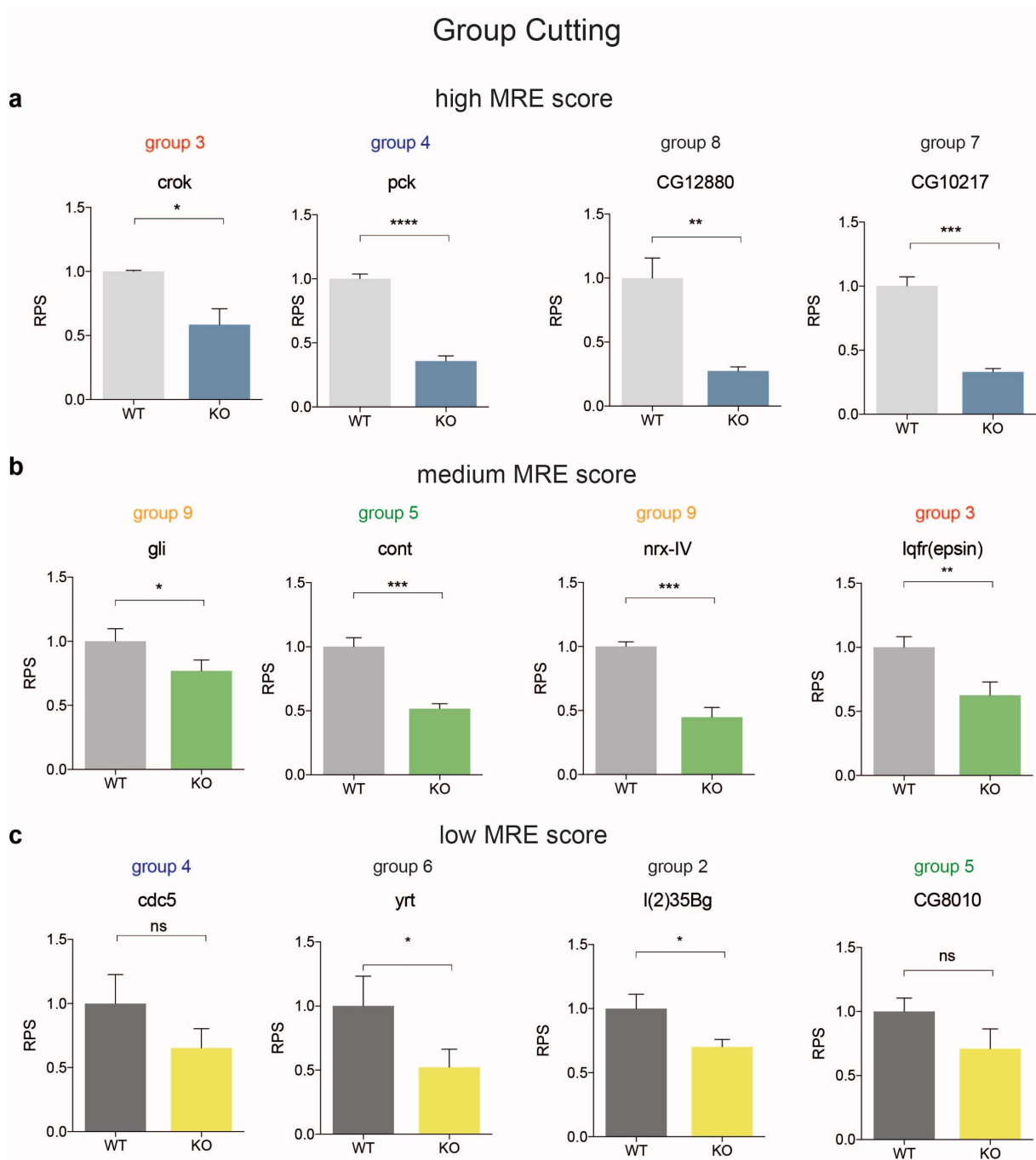


Figure 4.4 | Grouped TRACE scores of different MRE classes. **a**, MREs of the highest class (turquoise) reveal largest loss of RPS between WT and KO. **b**, MREs of the medium MRE score class exhibit smaller yet significant reductions of RPS from WT to KO. **c**, MREs of the lowest class display smallest reductions that were only statistically significant in 2 out of 4 cases. Coloured groups were assayed in the same reaction tube (e.g. *crok* and *lqfr(epsin)*). $n=3$, error bars denote standard deviation, $*P < 0.05$, $**P < 0.005$, $***P < 0.0005$, ns = not significant. Statistical significance was calculated using student's t-test. RPS = Relative protection score.

Calculating the mean of the TRACE scores for the *high*, *medium* and *low* MRE score classes revealed a significant difference between the *high* and the *low* group (Fig. 4.5a). The *high* group has a TRACE score mean of 2.79, which is 1.57 fold higher than the *medium* mean (1.77) and 1.78 times higher than the *low* group mean (1.57) (Fig. 4.5a). In this case however, the *medium* class mean average was not significantly different from the *low* class.

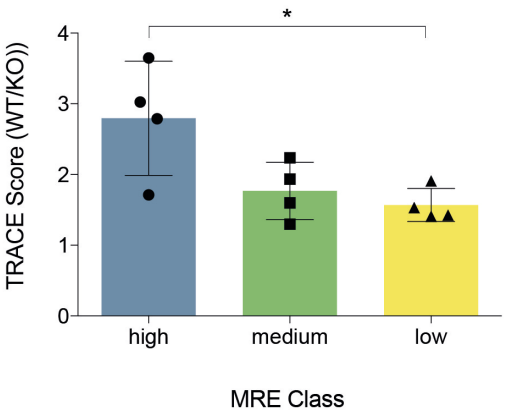
To assess the assay performance, we plotted the TRACE scores derived from the *grouped* experiment against their corresponding TRACE scores from the *solo* experiment. We noted a significant linear correlation (linear regression, $p = 0.0017$, $R^2 = 0.6429$), suggesting that the assays performed with similar specificity (Fig. 4.5b). It is noticeable that the MRE classes separate more clearly on the *solo* axis whilst on the *grouped* axis the *low* and *medium* MRE classes are at least partially overlapping (Fig. 4.5b).

We next evaluated the *grouped* assay sensitivity against the MRE score. Plotting the *grouped* TRACE score *versus* MRE score for each of the 12 MREs individually, revealed a statistically significant linear positive relationship (linear regression, $p = 0.0020$, $R^2 = 0.6336$) (Fig. 4.5c). Charting the means of both TRACE score and MRE score also displayed a statistically significant positive relationship ($R^2 = 0.7948$, $p = <0.0001$) (Fig. 4.5d).

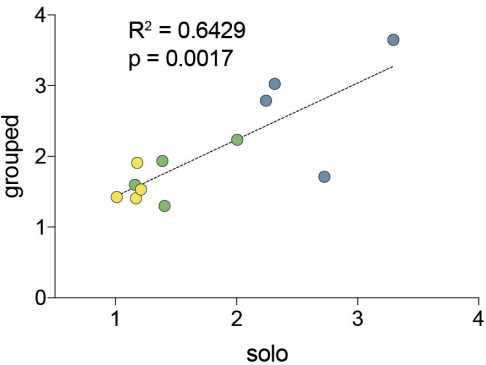
Furthermore, we examined the TRACE score against fold change of transcript levels (miR-184 KO divided by WT). Again, we noticed a statistically significant positive relationship for individual scores (linear regression, $p = 0.0007$, $R^2 = 0.6477$) (Fig. 4.5e). Plotting the means of the TRACE scores of each class against their respective mean fold change also revealed a statistically significant linear correlation (linear regression, $p = 0.0122$, $R^2 = 0.4824$) (Fig. 4.5f). From this data we concluded that, although the *grouped* TRACE assay is capable of distinguishing highly active MREs

from less or inactive ones, the resolution of the assay is not as sensitive as the *solo* strategy.

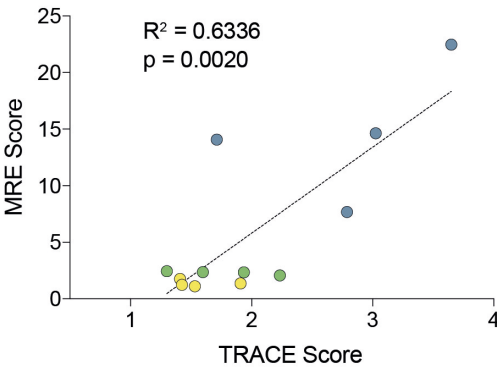
a TRACE Score of MRE Classes (grouped)



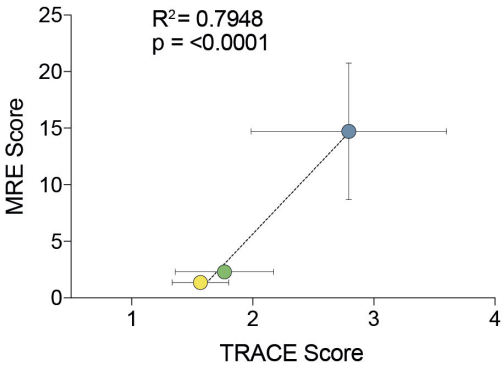
b TRACE Score (solo vs grouped)



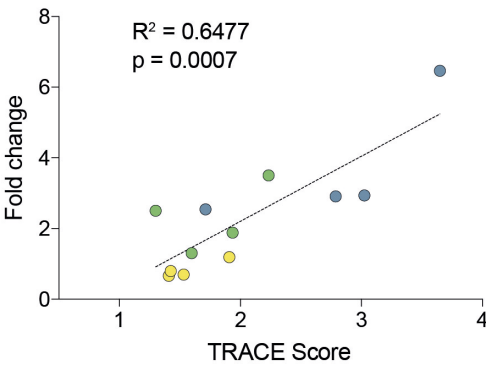
c grouped TRACE Score vs MRE Score



d grouped TRACE Score vs MRE Score (mean)



e grouped TRACE Score vs Fold Change



f grouped TRACE Score vs Fold Change (mean)

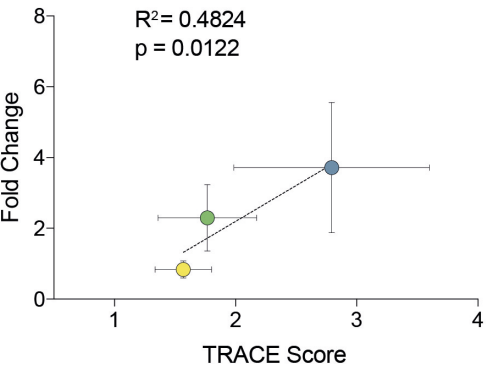


Figure 4.5 | Grouped TRACE assay reflects performance of solo TRACE and two other metrics. **a**, mean *grouped* TRACE scores of each MRE class reflect highest average in *high* class that is significantly different from *low* class. **b**, *grouped* TRACE scores reflect TRACE scores obtained through *solo* analysis. **c**, *grouped* TRACE score correlates with MRE score. **d**, mean TRACE scores correlates with mean MRE scores. **e**, *grouped* TRACE scores correlate significantly with fold change (KO/WT transcript levels). **f**, mean TRACE scores of MRE classes also correlate with mean MRE score. Correlation analysis was done by linear regression where p = probability that slope is non-zero. $*P < 0.05$. Error bars = standard deviation.

4.3.3 Short ASOs sustain efficient RNase H-mediated cleavage at MRE sites

Albeit at a cost of sensitivity, the grouped approach offered an improved throughput for the TRACE assay. To generate an even more streamlined strategy we sought to determine if a single short ASO centred on the 6mer seed sequence could serve as a universal agent for RNase H-mediated cleavage, thus replace the necessity of using a one ASO per MRE (i.e. 90 in the case of the miR-184 target network). In this manner, the 6mer motif could be assayed transcriptome-wide. To implement this strategy, we first assessed the cleavage rate of ASOs with varying lengths containing a 6mer identical to the miRNA-184 seed (Fig. 4.6a,b). To determine the cutting efficiency of these ASO variants, we measured the relative protections of *crok* and *cdc5* MREs using crosslinked WT lysates (Fig. 4.6c,d). Data was normalised to the no ASO control (set to RPS of 1). The 25mer was included as a positive control to benchmark the maximum cutting efficiency for shorter ASO variants. To assess the minimal ASO length requirements for effective cutting, we first tested a 6mer ASO that targets nucleotides 2-7 of the seed corresponding region. This architecture revealed poor cutting of the *crok* MRE and we considered a dynamic range of 1 to 0.6 too narrow to conduct a meaningful TRACE score calculation (RPS of 0.60 compared to 0.08 for 25mer) (Fig. 4.6c). We hence extended the ASO to 10nt and observed an almost 2-fold improvement in cutting efficiency compared to the 6mer.

11mer and 12mer ASOs exhibited only marginally higher cutting efficiencies than the 10mer (Fig. 4.6c). We validated the shortened ASO strategy on an unrelated MRE (*cdc5*) and noted equally adequate cutting levels (Fig. 4.6d). From this data we concluded that a 10mer is capable of eliciting substantial RNA cleavage.

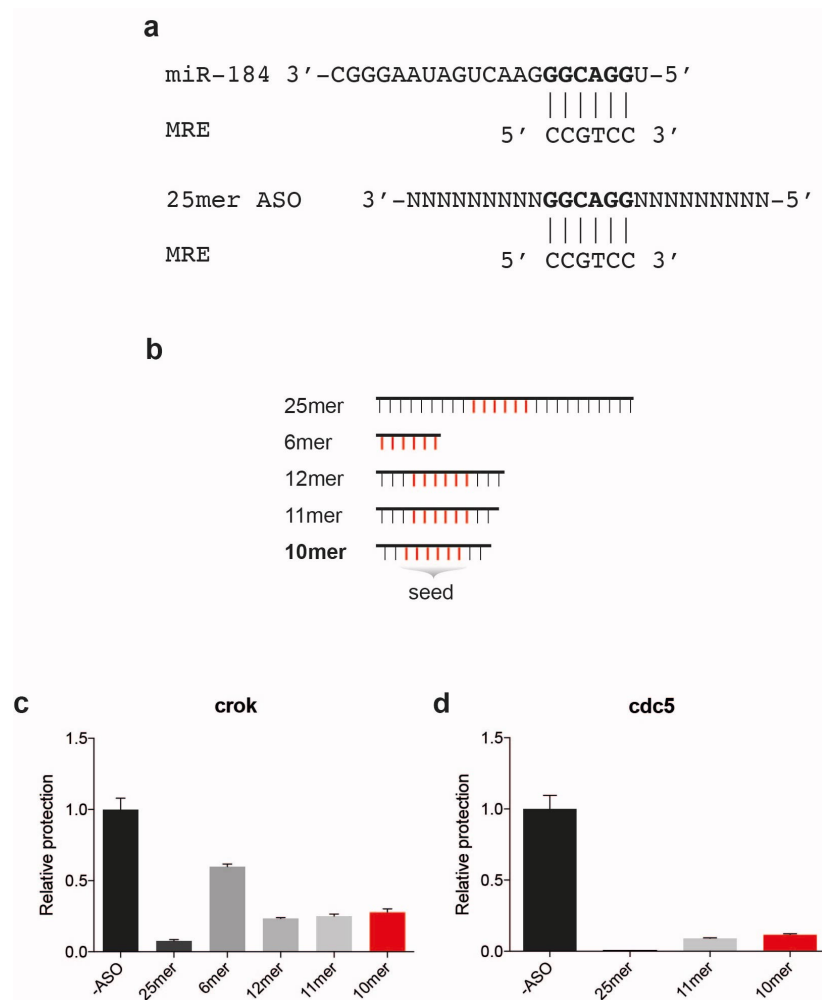


Figure 4.6 | Shorter ASOs can still cleave MREs. **a**, mature miR-184 sequence with seed in bold that is complementary to every predicted MRE. 25mer ASOs are centered on seed corresponding region and fully complementary to target mRNA. **b**, shortened ASO that remain centered on seed (red lines). **c**, assessment of cutting potential of shortened ASOs targeting *crok* MRE in WT lysate (data is normalised to no ASO control). 10mer (red) displays best trade-off between length and potency. **d**, 10mer cognate to *cdc5* MRE also induces substantial cleavage. Error bars = standard deviation. MRE = miRNA Response Element.

4.3.4 Short ASOs effectively uncover differences in RPS between WT and KO

The 10mer ASO demonstrated efficient cutting on WT cell lysates. However, it was important to establish whether the 10mer can also reliably detect the difference in protection between WT and KO and render a meaningful TRACE score. To address this question we re-assessed the RPS of the *crok* MRE in WT and KO cells with a 10mer ASO. This analysis revealed a significant reduction in RPS (37%), albeit not as strong as the 25mer (63%) (Fig. 4.7a,b). The 10mer generated a TRACE score of 1.58 compared to 2.73 for the 25mer.

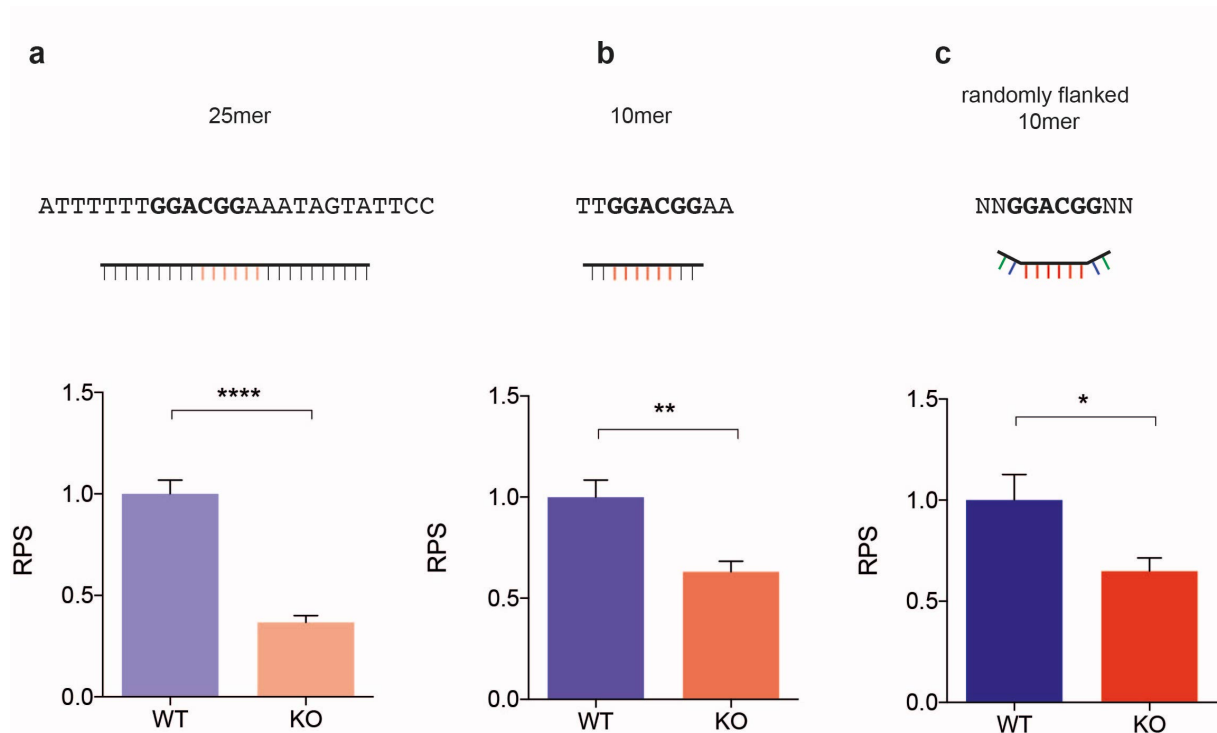


Figure 4.7 | Shortened 10mer and randomly flanked 10mer still detect difference in *crok* MRE RPS. **a**, 25mer ASO specific to *crok* reflects large loss of RPS between WT and KO. **b**, Less pronounced, 10mer also detects differential RPS. **c**, randomly flanked 10mer (NN-GGACGG-NN) produces a reduction in RPS of similar potency to the cognate 10mer. $n \geq 3$, error bars = standard deviation, * $P < 0.05$, ** $P < 0.005$, **** $P < 0.00005$. Statistical significance was calculated using student's t-test. RPS = Relative protection score.

While this was an important milestone, the ultimate requisite for universal seed cutting is RNase H cleavage with a single ASO in parallel for all MREs of the miR-184 target network. As the 6mer revealed inadequate cutting we extended it by two randomised nucleotides at each end to generate a randomly flanked 10mer ASO of the sequence NN-GGACGG-NN. The central six nucleotides are identical to the miR-184 seed and hence complementary to every MRE of the predicted miR-184 target network. N represents 25% chance incorporation of either A, C, G, or T. (Fig. 4.7c). This pool inevitably contained the cognate 10mer for every MRE, albeit at a 256-fold (4^4) lower concentration than a defined 10mer. We treated WT and KO cell lysates with the randomly flanked ASO and noticed a significant reduction in RPS of 35%, comparable to the cognate 10mer (37%) (Fig. 4.7c). The TRACE score for *crok* generated with the randomly flanked 10mer was 1.54, again closely reflecting the TRACE score derived from the cognate 10mer.

From this data we concluded that a randomly flanked 10mer carrying a miR-184 seed-identical central sequence can be used to measure MRE protection in the *crok* mRNA. Following the proof-of-concept experiment, we returned to our training data set and tested 11 additional MREs with varying activities.

4.3.5 Analysis of 12 target MREs by multiplex TRACE

To evaluate the general performance of the randomly flanked 10mer strategy, we revisited our training data set containing 12 MREs. Confirming the reliability of this approach, the *high* MRE score group exhibited the largest reduction in RPS from WT to KO for all four MREs: 35% for *crok*, 43% for *pck*, 53% for CG10217, and 60% for CG12880 (Fig 4.8a).

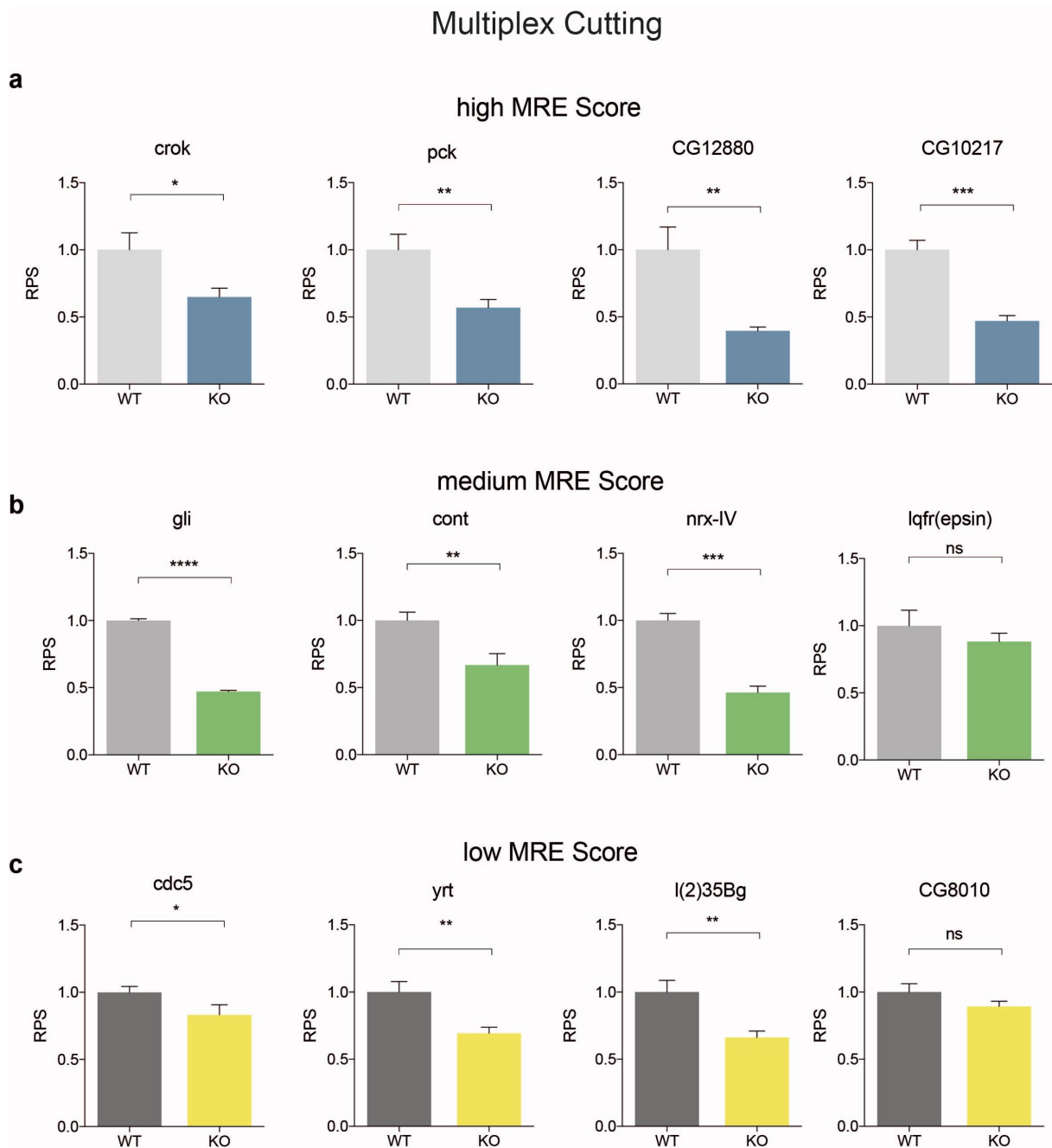


Figure 4.8 | Single ASO, *multiplexed* TRACE scores of different MRE classes. a, MREs of the highest class (turquoise) reveal substantial loss of RPS between WT and KO. **b,** MREs of the medium MRE score class display moderate reductions of RPS from WT to KO. **c,** MREs of the lowest class display smallest reductions. $n = 3$, error bars = standard deviation, $*P < 0.05$, $**P < 0.005$, $***P < 0.0005$, $****P < 0.00005$, ns = not significant. Statistical significance was calculated using student's t-test. RPS = Relative protection score.

The *medium* group displayed similarly strong decreases in RPS between WT and KO, 53% for *gli*, 33% for *cont*, 54% for *nrx.IV*, with the exception of *lqfr* (Epsin) which appeared to not be differentially protected (Fig. 4.8b). Interestingly, the *low* group, albeit attenuated, revealed a significant decrease in RPS for 3 out of 4 MREs: 17% for *cdc5*, 31% for *yrt*, 34% for *I(2)35Bg* (Fig. 4.8c). Overall, we noted a constant decrease in the TRACE score from the *high* group to the *medium* group and *low* group. The *high* group has a TRACE score mean of 1.99, which is 1.15 fold higher than the *medium* mean (1.73) and 1.51 fold higher than the *low* group mean (1.32) (Fig. 4.9a).

To directly compare the multiplex assay performance against previously defined MRE scores we plotted TRACE score *versus* MRE score and noted a significant linear correlation (linear regression, $p = 0.0277$, $R^2 = 0.3983$) (Fig. 4.9b). Although calculating the mean of both parameters for each class also displayed a positive relationship, this trend was not statistically significant (linear regression, $p = 0.3742$, $R^2 = 0.6926$) (Fig. 4.9c).

We next sought to compare the multiplex TRACE score against the fold change (KO/WT) for each MRE. Plotting these two metrics for every MRE individually revealed a statistically significant linear relationship (linear regression, $p = 0.0204$, $R^2 = 0.4308$) (Fig. 4.9d). However, repeating this analysis with the mean of each class produced a non-significant positive correlation (linear regression, $p = 0.2808$, $R^2 = 0.8177$) (Fig. 4.9e).

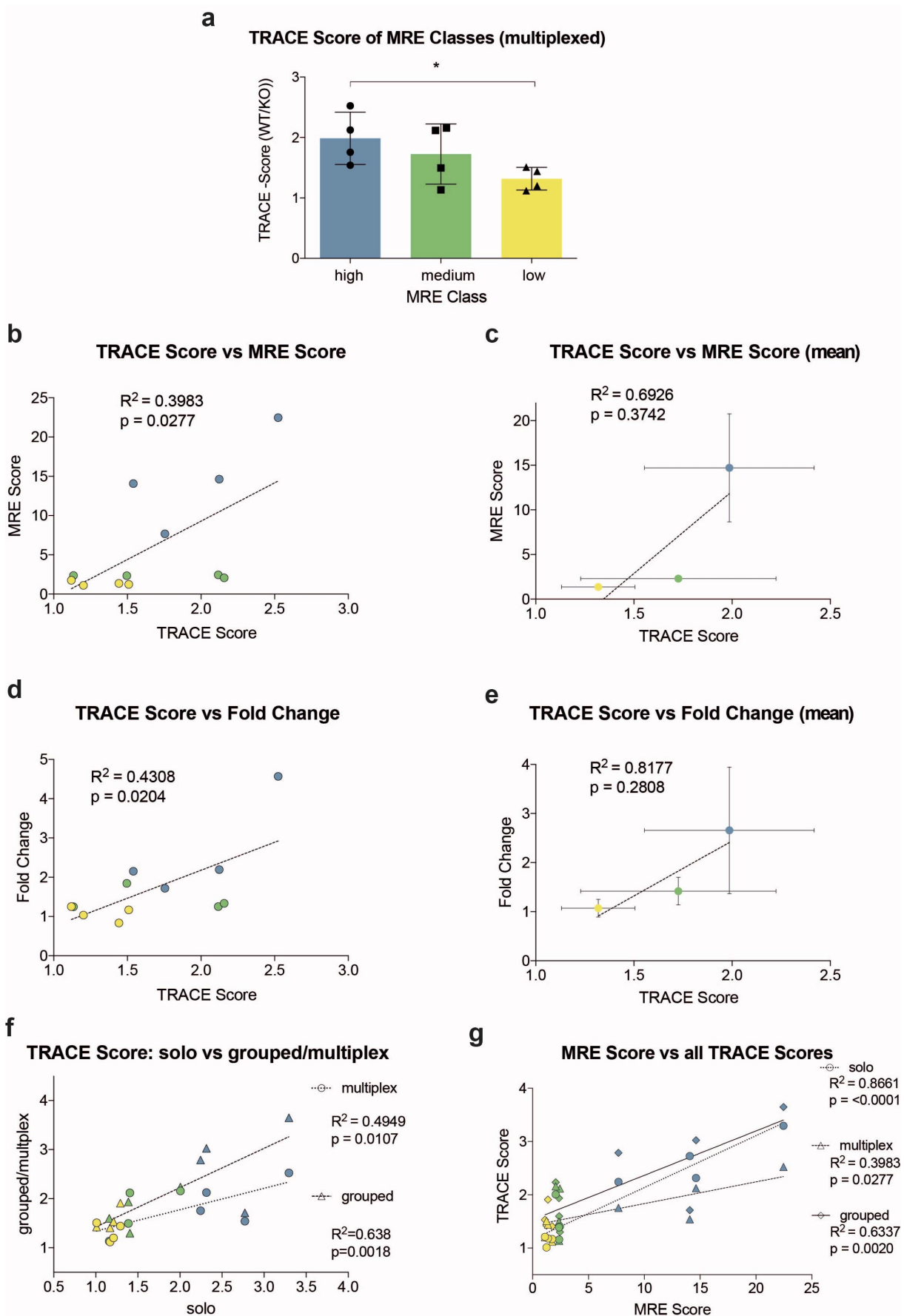


Figure 4.9 | Single ASO, *multiplex* TRACE assay reflects performance of *solo* TRACE and two other metrics. **a**, mean *multiplex* TRACE scores of each MRE class shows highest average in *high* class that is significantly different from *low* class. **b**, *multiplex* TRACE scores correlate with MRE scores, especially for *high* and *low* classes. **c**, mean *multiplex* TRACE score corresponds to mean MRE score. **d**, *multiplex* TRACE scores correlate significantly with fold change (KO/WT transcript levels). **e**, mean TRACE scores of MRE classes also correlate with mean fold change. **f**, Comparison of *grouped* and *multiplex* TRACE scores *versus* corresponding *solo* scores. Both correlations are significant, *grouped* performs slightly better. **g**, Head-to-head comparison of MRE score *versus* TRACE scores derived by *solo*, *grouped* and *multiplex* assays. *Solo* correlates best followed by *grouped*, followed by *multiplex*. Correlation analysis was done by linear regression where p = probability that slope is non-zero. $*P < 0.05$. Error bars = standard deviation.

To evaluate which high throughput method (*grouped* or *multiplex*) best matched the *solo* approach, we charted individual TRACE solo values against both those derived from the *grouped* assay as well as from the *multiplex* assay (Fig. 4.9f). We noted a significant correlation for the *grouped* TRACE scores (linear regression, *grouped* $p = 0.0018$, $R^2 = 0.638$). This performance was slightly superior compared to the *multiplex* TRACE which nonetheless was still statistically significant (linear regression *multiplex* $p = 0.0107$, $R^2 = 0.4949$). Finally, we conducted a head-to-head comparison of *solo*, *grouped* and *multiplex* modes against MRE score. This analysis revealed that *solo* cutting is most predictive of MRE activity, followed by *grouped* cutting, followed by *multiplex* cutting (Fig. 4.9g). Together, these results indicate that, in principle, both the *grouped* and *multiplex* strategies are functional and could be used for the implementation of high-throughput TRACE analysis.

4.3.6 Development of a high-throughput readout for multiplex TRACE analysis

So far we have demonstrated the feasibility of multiplex TRACE, either through pooling ASOs into one reaction or employing a single universal ASO that can target the entire miR-184 primary MRE network. To increase the detection efficiency of TRACE we next focused on restructuring the readout assay from qPCR cDNA amplification (reads/counts per MRE) to seed targeted RT priming followed by sequencing. If successful, this modification would permit the implementation of a NGS readout, thereby allowing an unbiased transcriptome-wide analysis of miR-184 target binding. As previously hypothesised, based on the conceptual framework of this strategy, only intact (i.e. uncut) 6mer motifs in the RNA should be able to initiate cDNA synthesis (Fig. 4.2c). We reasoned the chance of success of this experiment depends on optimal specificity of cDNA generation. Limiting cDNA synthesis to cognate 6mers only, should in theory also improve sequencing read depth, as other RNAs (including very abundant rRNAs) lacking the motif should not be reverse transcribed and thus will not enter library preparation. We hence optimised the RT process to select the most favourable conditions to specifically produce cDNA from motif bearing RNAs.

Standard reverse transcription protocols typically employ a pool of random hexamers to initiate reverse transcription. This process is efficient as the majority of RNA molecules will base pair with cognate hexamers which will initiate the synthesis of full-length cDNA molecules. We assessed if a defined 6mer RT primer sense to the miR-184 seed (seed primer) can initiate reverse transcription at target MREs, by virtue of qPCR amplification of an upstream sequence in the target RNA (Fig. 4.10a).

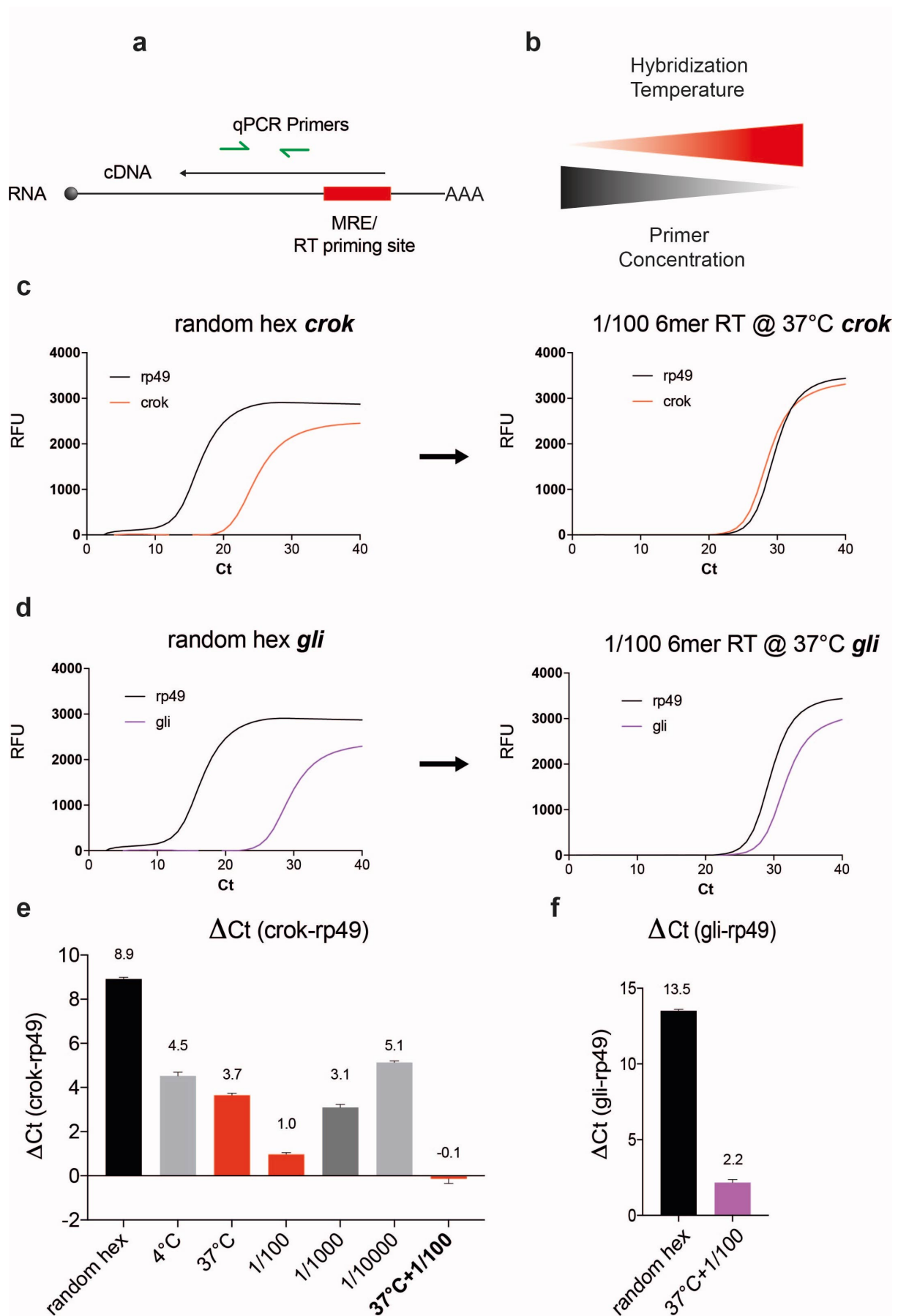


Figure 4.10 | Seed specific priming engenders effective reverse transcription. **a**, To assess whether a defined hexamer can initiate cDNA synthesis at an MRE we designed qPCR primers ~ 100bp upstream of the RT primer binding site. **b**, increasing hybridisation temperature and reducing primer concentration engenders favourable primer binding kinetics. **c**, RT priming at 37°C with 100-fold diluted primer induces 9 cycle Ct shift for *crok* detection relative to *rp49*. **d**, the same conditions also a 11.3 cycle shift in *gli* detection. **e**, quantification of ΔCt values (*crok* - *rp49*). Bold conditions were chosen for subsequent RT reactions. **f**, *gli* cDNA production is also improved by stringent conditions (37°C priming + lowered primer concentration). Error bars = standard deviation. RFU = relative fluorescent units.

We chose *rp49* to serve as a reference against which we measured the relative enrichment of two target 6mer bearing transcripts, *crok* and *gli*. As expected, using random hexamers, abundant *rp49* emerged significantly earlier than the two candidate targets *crok* and *gli* (Fig. 4.10c,d,e). However, when random hexamers were substituted for our seed primer, we noted a decrease in ΔCt to 4.5 cycles for *crok*, suggesting a specific enrichment of this transcript (Fig. 4.10e). While this constituted a promising start, we sought to further improve this enrichment. Primer kinetics are influenced by hybridisation temperature as well as primer concentration. To increase seed primer binding specificity and reduce off-target *rp49* priming, we increased the hybridisation temperature from 25°C to 37°C and noted a further decrease in ΔCt (*crok* - *rp49*) to 3.7 cycles. In parallel, we tested the effect of primer dilution on the relative detection of *crok*. We observed that a 100-fold dilution of the recommended hexamer concentration markedly improved the detection (Fig. 4.10e). Performing the RT at the increased temperature of 37°C with the reduced primer concentration revealed an additive effect, resulting in detection of *crok* transcripts before *rp49*, and an overall shift in ΔCt of 9 cycles from the original conditions (Fig. 4.10c,e). Applying the same optimisation steps also resulted in a considerable improvement in the detection of *gli* transcripts (ΔCt *gli-rp49* from 13.5 (random hex)

to 2.2 (6mer)) (Fig. 4.10d,f). From this data we concluded that cDNA synthesis of target mRNAs could be initiated specifically at MREs using a seed targeted 6mer primer. We hence replaced the standard RT step in the downstream library preparation protocol with this modified step.

4.3.7 Normalisation by an exogenously delivered mRNA

The preparation of individual libraries from multiple samples (12) requires effective normalisation to eliminate bias inevitably introduced during the preparation procedure. To identify differences in miRISC binding with high confidence we must control for technical variation throughout a complex library preparation as well as RNA content. Since the implementation of seed priming RT conditions deliberately depletes commonly used RNA content controls (*rp49* and other housekeeping genes) we considered the addition of 25mer gene-specific RT primers (GSP) for *rp49* and *hsp83* (Fig. 4.11a). Including these two controls permits total RNA content normalisation for sequencing. Furthermore, we decided to introduce a spike-in *in vitro* transcribed (IVT) RNA as a means of normalising for technical variation introduced during library preparation and PCR. Based on minimal homology to the *Drosophila* genome we chose the *egfp* gene for this purpose (Fig. 4.11a).

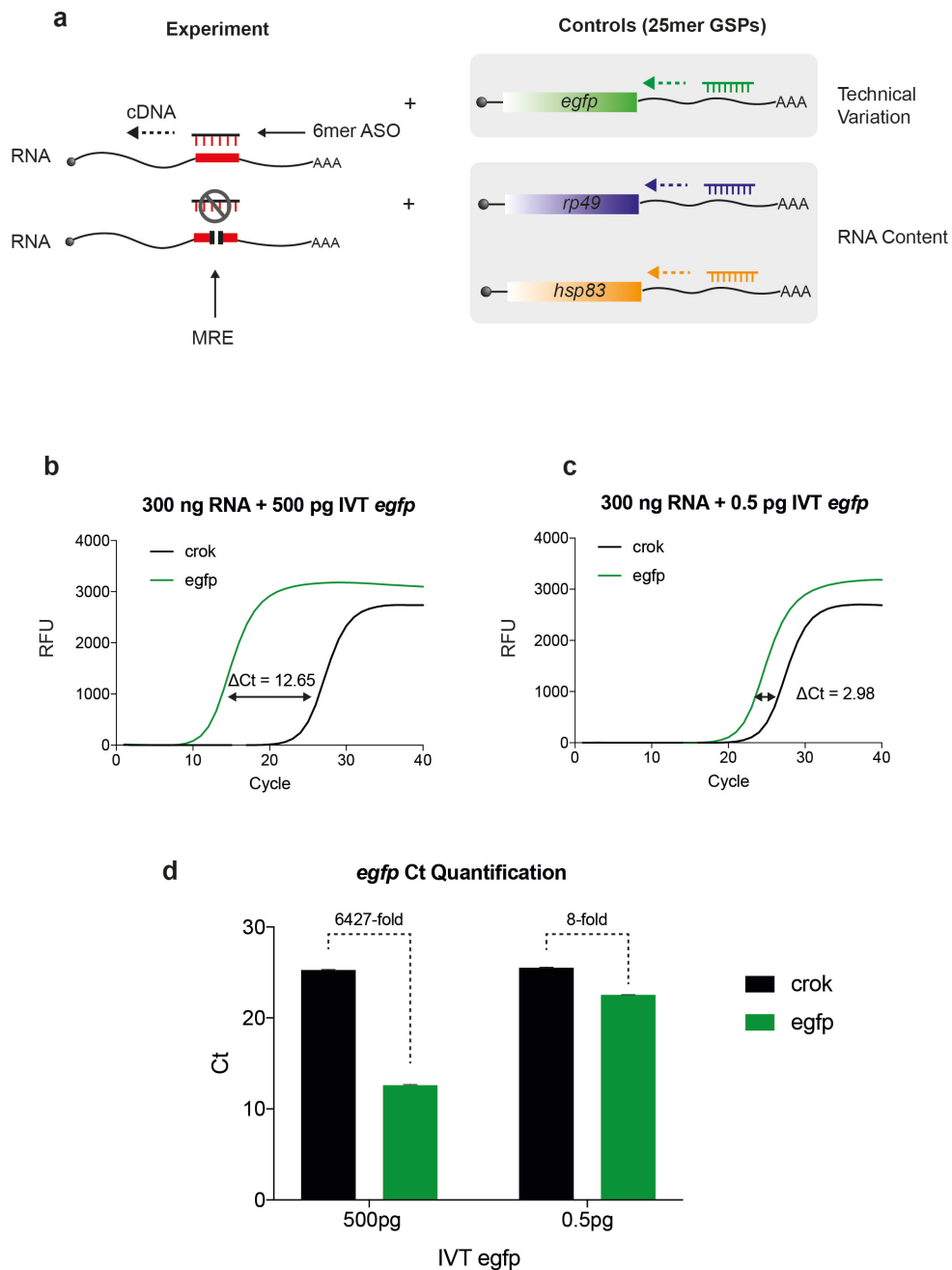


Figure 4.11 | High-throughput TRACE experiment is performed with RNA content normalisation controls and a spike-in control. a, cDNA synthesis is performed through addition of seed corresponding 6mer RT primer, *in vitro* transcribed *egfp* RNA for technical normalisation, *egfp* 25mer GSP as well as two endogenous RNA content control primers (*rp49* and *hsp83*). **b,** 500pg *egfp* spike-in exceeds relevant endogenous RNAs (e.g. *crok*) and was deemed as too concentrated. **c,** 0.5pg *egfp* spike-in delivers similar control RNA levels to *crok* and was hence chosen as concentration for spike-in. **d,** actual Ct values and fold-difference of spike-in *egfp* vs. *crok*. Data shows qPCR triplicates.

To closely assess library preparation performance and to provide a wide and robust normalisation range, we sought to introduce *egfp* at a ten times higher concentration than endogenous *crok*. To estimate the required amount of spike-in, we tested two concentrations of *egfp* (500 pg and 0.5 pg) added to 300 ng total RNA which was subjected to reverse transcription. We measured the relative abundance of both *crok* and IVT *egfp* by qRT-PCR and noted a ΔC_t (*crok* - *egfp*) of 12.65 for 500 pg spike-in, suggesting an 6427-fold excess *egfp* (Fig. 4.11b,d). The 0.5 pg spike-in resulted in a ΔC_t (*crok* - *egfp*) of 2.98, which corresponds to a 8-fold higher abundance of *egfp* (Fig. 4.11c,d). Consequently, the 0.5 pg concentration was chosen for the final protocol. Thus, in the library preparation protocol each final reverse transcription reaction contained 300 ng total RNA, 0.5 pg IVT *egfp* RNA, 6mer miR-184 seed RT primer, a *egfp* 25mer gene specific primer (GSP), a *rp49* 25mer GSP, and a *hsp83* GSP.

4.3.8 The miR-184 6mer seed motif is significantly enriched in TRACE libraries

Six samples treated with the universal randomly flanked 6mer and six uncut controls (3x *WT*-ASO, 3x *WT*+ASO, 3x *KO*-ASO, 3x *KO*+ASO) entered RNA library preparation according to the standard NEB Next (E7530) protocol with the aforementioned modifications to the first strand cDNA synthesis and were sequenced on an Illumina MiSeq sequencer. Hitherto, 6mer enrichment was assessed using biased RT-qPCR which was informative with regards to *crok* and *gli* but lacked transcriptome-wide, unbiased detection. Hence, we first assessed the performance of global seed primer initiated cDNA synthesis. Towards this end, using the BWA algorithm, we mapped all reads to the *Drosophila Dm3* genome and searched for aligned reads with the 6mer RT primer motif (GGACGG).

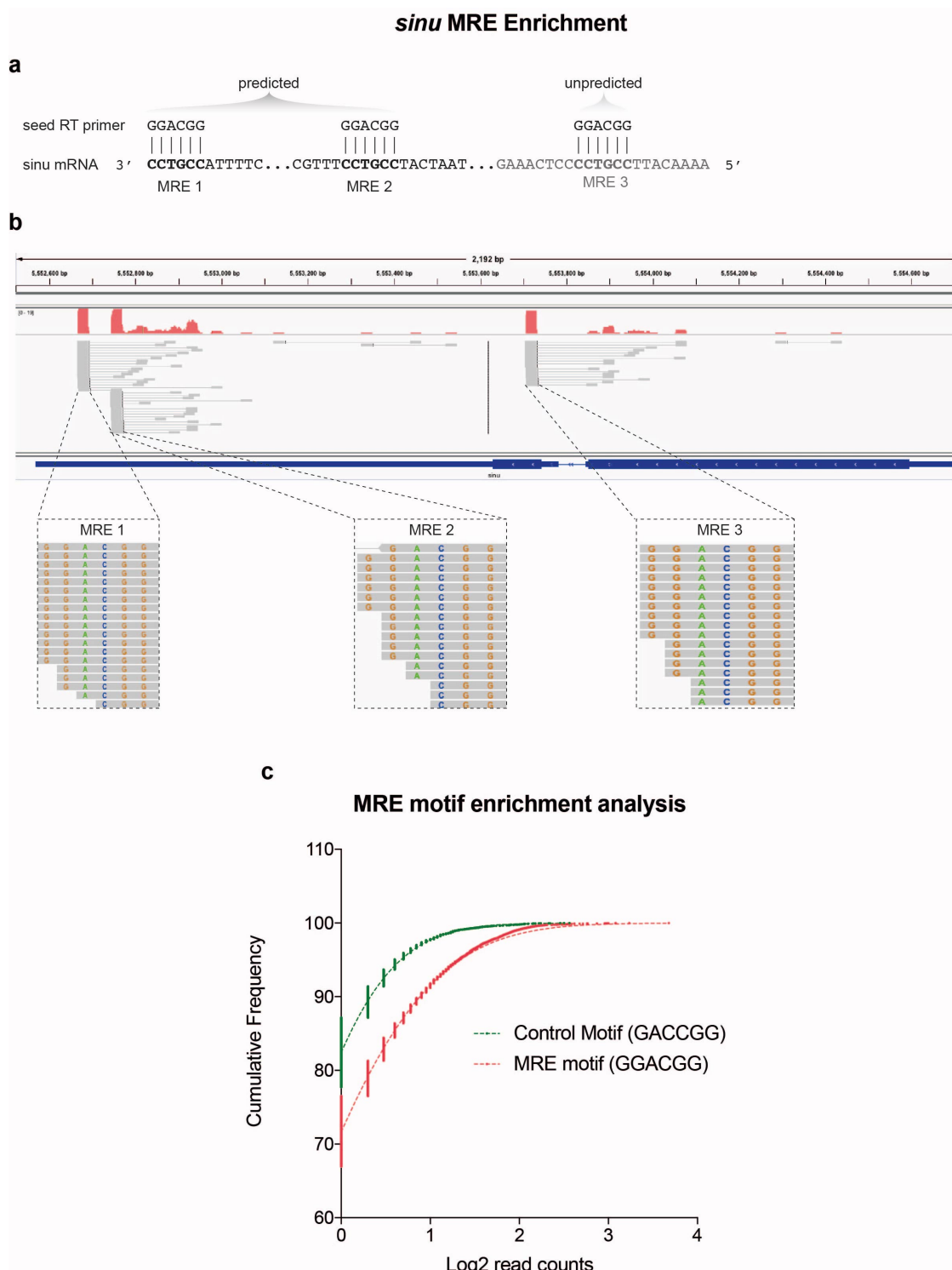


Figure 4.12 | 6mer RT primer motif is enriched in library preparation. **a**, the *sinu* locus contains 2 predicted MREs and 1 unpredicted, non-canonical MRE (ORF). **b**, all of them are the origin of sequencing starting with the RT primer sequence (grey). **c**, calculation of the genome-wide cumulative frequency of the 6mer motif revealed a significant difference to a control motif of similar GC%.

As an example, the *sinu* mRNA contains predicted tandem miR-184 target sites in its 3'UTR as well as an additional putative target site in its ORF, which is not normally detected by standard prediction algorithms. Using the IGV software to visualise raw reads at the *sinu* locus revealed that both MREs predicted in the 3'UTR contained sequencing reads starting at the seed motif (Fig. 4.12). Furthermore, we noted that additional reads clustered at the miR-184 6mer target site in the ORF. Together, these two observations suggested that the readout is not limited to predicted MREs within the 3' UTR but in fact capable of assessing, in principle, any 6mer motif across the transcriptome.

We further quantified the 6mer motif cumulative frequency across all sequencing reads from pooled samples and found its overall read count to be substantially higher than a randomly chosen control motif with similar %GC percentage (Fig. 4.12c). From this data we concluded that seed priming is an effective strategy for targeted cDNA synthesis.

4.3.9 Depth of sequencing limits the ability to calculate meaningful RPS values.

Once we established that the 6mer miR-184 seed motif was significantly enriched in the NGS libraries, we assessed the read depth for our 12 candidate MREs. Zooming in on the *crok* locus revealed an extremely low read count with no reads detected in one of the WT plus ASO replicates (Fig. 4.13a). Comparing reads spanning the *crok* MRE between the 12 samples (3x WT-ASO, 3x WT+ASO, 3x KO-ASO, 3x KO+ASO) showed very low read counts across all samples rendering the calculation of statistically significant TRACE scores unfeasible (Fig. 4.13b). This low coverage was also reflected throughout other loci. Across the 12 target MREs none revealed sufficient depth to conduct a meaningful statistical analysis of TRACE scores (Fig. 4.13c). The mean average raw read count for our 12 candidate MREs ranged from 0.5 for CG12880 to a maximum of 9.5 counts for *nrx-IV*.

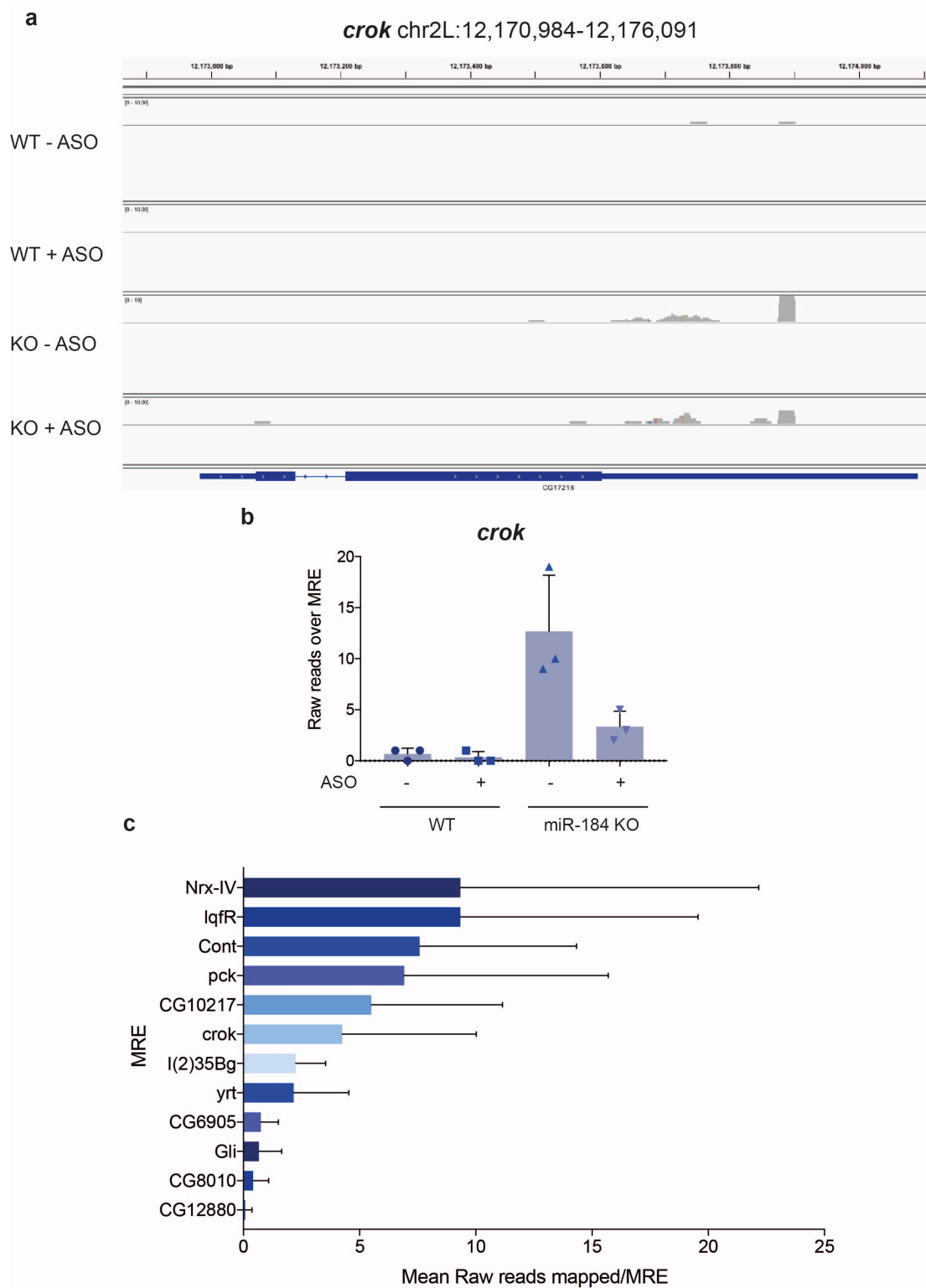


Figure 4.13 | First sequencing run provided insufficient depth to calculate TRACE scores. **a**, raw mapped reads on the *crok* locus from 4 experimental conditions. N.b. coverage is very low. **b**, quantification of triplicates from raw reads mapped to *crok* locus. It is not possible to calculate TRACE scores from this data. **c**, mean raw mapped reads reveal extremely low coverage across 12 candidate MREs. Error bars = standard deviation. ASO = antisense oligonucleotide.

Although this initial MiSeq run generated the expected number of reads (12,911,713), the coverage of target mRNAs of interest appeared to be very poor. To understand the cause of this phenomenon, we sought to scan the genome for ‘hotspots’, which could potentially consume bandwidth at the expense of our candidate targets. To this end, we *in silico* divided the fly genome into 500bp intervals, and calculated the sum of total reads for each 500bp window (Fig. 4.14a). This analysis revealed very high read counts overlapping with ribosomal RNA (rRNA) loci on chromosome *XHet*, which added together accounted for the largest share of reads obtained in this experiment (Fig. 4.14b).

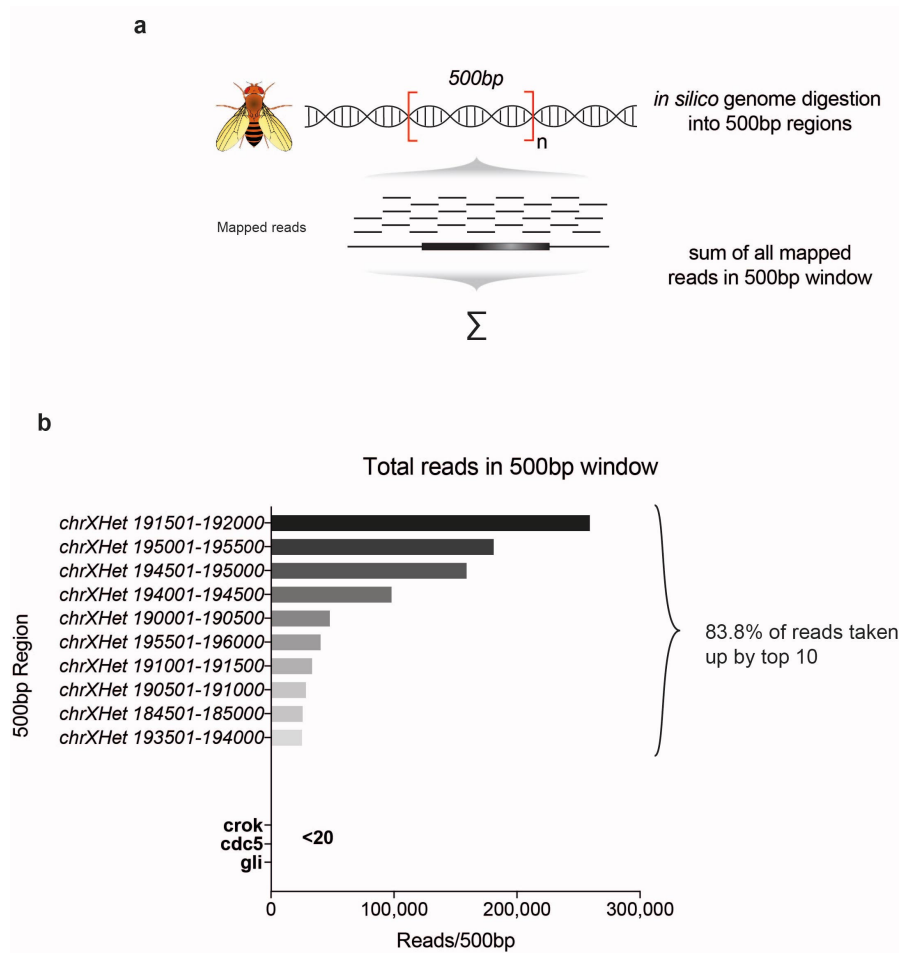


Figure 4.14 | Vast majority of reads map to ribosomal RNA regions. **a**, the *Drosophila* genome was split into 500bp regions *in silico* and total reads from all 12 samples were summed up for each window **b**, x-axis shows reads/500bp and y-axis shows ‘hotspot’ 500bp regions in decreasing order. Top regions on chromosome *XHet* (rRNA loci) dominate library reads and account for majority of sequencing reads, depriving target MREs of essential bandwidth (*crok*, etc).

Constituting the most abundant cellular RNAs, rRNA frequently absorb the majority of sequencing reads, rendering analysis of mRNAs extremely difficult and often impossible. However, as seed initiated RT should limit cDNA synthesis to only those transcripts that contain the 6mer motif, we speculated that target mRNAs carrying the seed would be sufficiently represented in the library preparation. Surprisingly, the 6mer motif CCGTCC was not found in the top contaminating RNA species suggesting that our approach could engender cDNA synthesis with mismatch 6mer base pairing and that rRNA must be eliminated separately.

To prevent the accumulation of rRNA in library preparations in the first place and thereby creating meaningful bandwidth for non-rRNAs, several strategies have been developed. Unfortunately, the most commonly used poly-A selection *via* oligo-d(T) beads was not suitable for our experiment as this protocol would only capture the most proximal 6mer motif to the 3' end. For instance, if an mRNA harbours multiple MREs (i.e multiple 6mer motifs), and all are cut, only the motif most proximal to the 3' end of the transcript will be captured *via* poly-A based purification, while the other MREs would be lost. An alternative approach is the depletion of ribosomal RNA *via* commercial ribo-depletion kits that employ RNase H mediated cleavage of contaminating RNAs. However, as of present, such kits are not available for *Drosophila* ribosomal RNAs and custom adaptations have been shown to be ineffective. Nevertheless, we reasoned that as long as the analysis is restricted to reads starting with the desired 6mer motif, such off-targets should not affect the result other than by absorbing sequencing reads. As such, we next sought alternative strategies to remove these contaminating RNA species from our library preparation and improve the sequencing bandwidth for target RNAs, enabling statistically meaningful calculations of TRACE scores.

4.3.10 DASH removes contaminating RNA

Analysis of NGS data revealed that the majority of reads were assigned to rRNA. This significantly reduced the sequencing depth for our genes of interest. Ribo-depletion *via* commercial kits or poly(A) selection was not possible for technical reasons (see above). As a consequence, to open up sequencing space for our target RNAs, we resorted to a recently published protocol termed depletion of abundant sequences by hybridisation (DASH) (Gu et al., 2016).

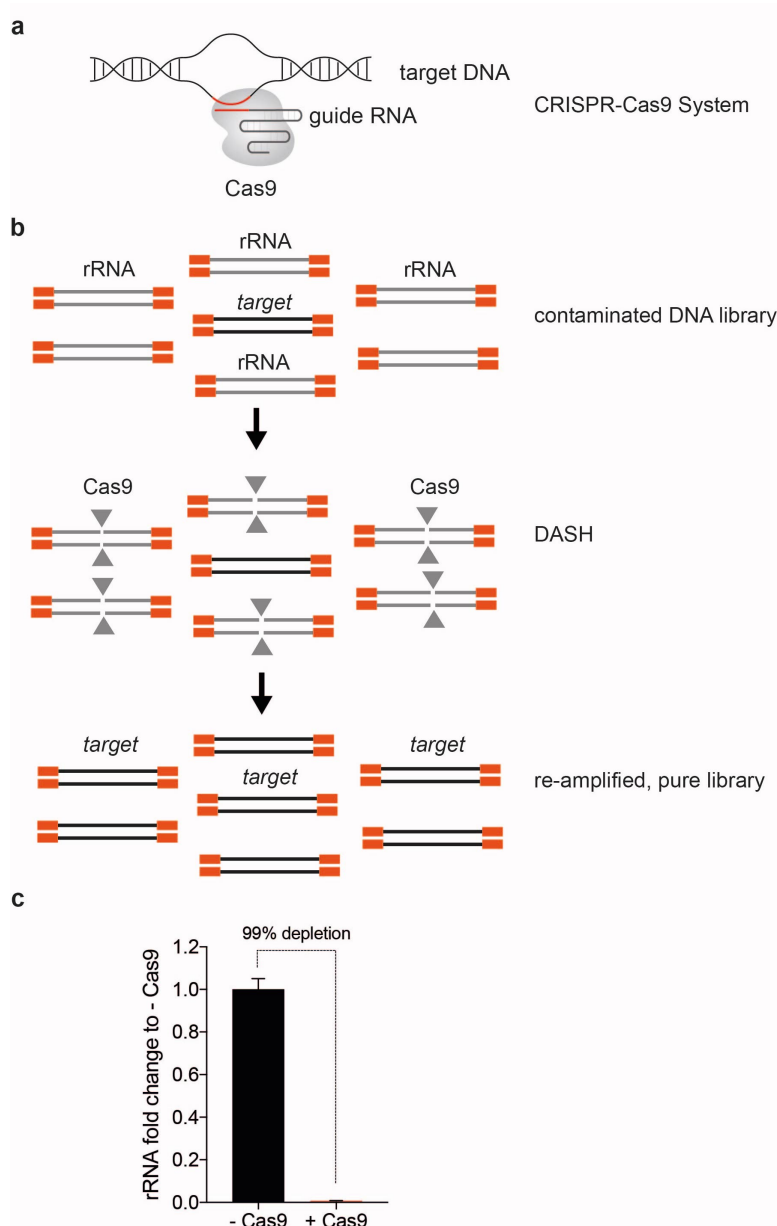


Figure 4.15 | DASH removes contaminating rRNA. **a**, DASH relies on CRISPR-Cas9 mediated cleavage of dsRNA *in vitro*. **b**, schematic protocol overview; undesired sequences

(e.g. rRNA) are abundant in the library preparation and outnumber RNAs of interest. Guide RNAs can be multiplexed for targeted *in vitro* Cas9 cleavage of undesired sequences. Post-treatment, the library is re-amplified using P5/P7 primers (against red barcode region) resulting in a higher proportion of targets of interest. **c**, rRNA is depleted in the +Cas9 sample, as determined by qPCR of a single species (chrXHet:194732-194952; 18S ribosomal RNA pseudogene:CR41602). Error bars = standard deviation.

Conceptually, this strategy uses Cas9-mediated cleavage to remove undesired dsDNA sequences present in an NGS library (Fig. 4.15a,b). This bears the considerable advantage that an already prepared library can be acted upon, obviating an expensive second *de novo* preparation. Furthermore, performing this assay *in vitro* permits the efficient multiplexing of CRISPR single guide RNAs (sgRNAs) with apparently no significant off-target effects. Hence, multiple undesired sequences can be eliminated simultaneously. To test this approach, we created several sgRNAs against the top 6 contaminating ribosomal RNA loci (Fig. 4.13b). Briefly, sgRNAs were designed to span contaminating sequences with gaps of approximately 50bp, resulting in ~ 1-3 sgRNAs per contaminating RNA locus. We pooled the libraries from all conditions (3x WT-ASO, 3x WT+ASO, 3x KO-ASO, 3x KO+ASO) into a single reaction, and added all 18 designed sgRNAs together with recombinant Cas9 protein. An important consideration when assessing the performance of the DASH strategy in removing contaminating dsDNAs from the library preparation, was that the transcript levels corresponding to our candidate genes and the *egfp* IVT control are not affected. Hence, following the execution of the DASH protocol we re-amplified the pooled library and detected by qRT-PCR a very efficient and specific depletion of rRNA (99%) without any apparent decline in *crok* or *egfp* transcript levels (Fig. 4.15c, not shown). From this data we concluded that DASH-mediated depletion is capable of eliminating the most abundant RNA species from our library preparation which in turn should liberate a considerable amount sequencing space. However, to maximise our chances of generating a

sufficient dynamic range in which to calculate TRACE scores we also upgraded our platform to the NextSeq benchtop sequencer which is capable of producing ~25-times more reads than then MiSeq (400-500m reads).

4.3.11 Post-DASH sequencing on NextSeq delivers high number of target reads

We re-sequenced our library on a NextSeq benchtop sequencer and noted considerable improvements in read depth. We obtained 329,019,498 reads, a 25-fold increase compared to 12,911,713 produced by the MiSeq run. Most of the predicted MRE loci had previously lacked sufficient read depth in the MiSeq experiment (Fig. 4.13). We wondered whether the substantial increase in read counts would now be reflected in higher coverage over the MREs of interest. We revisited all 12 loci and detected substantial coverage spanning the 6mer motif in all cases. To illustrate this, we zoomed in on three MREs, including two active ones (*crok* and CG6583) (Fig. 4.16a) and one inactive one (*yrt*) (Fig. 4.16b). For example, the *wildtype* + ASO triplicate now exhibited a mean read count over the 6mer of 53.3 for *crok* (c.f. 0.6 in MiSeq), of 82 for CG6583 (c.f. 1.3 in MiSeq) and 129.7 for *yrt* (c.f. in MiSeq). From this data we concluded that sufficient read depth has been attained to meet statistical requirements to develop an unbiased model of transcriptome-wide miR-184 binding.

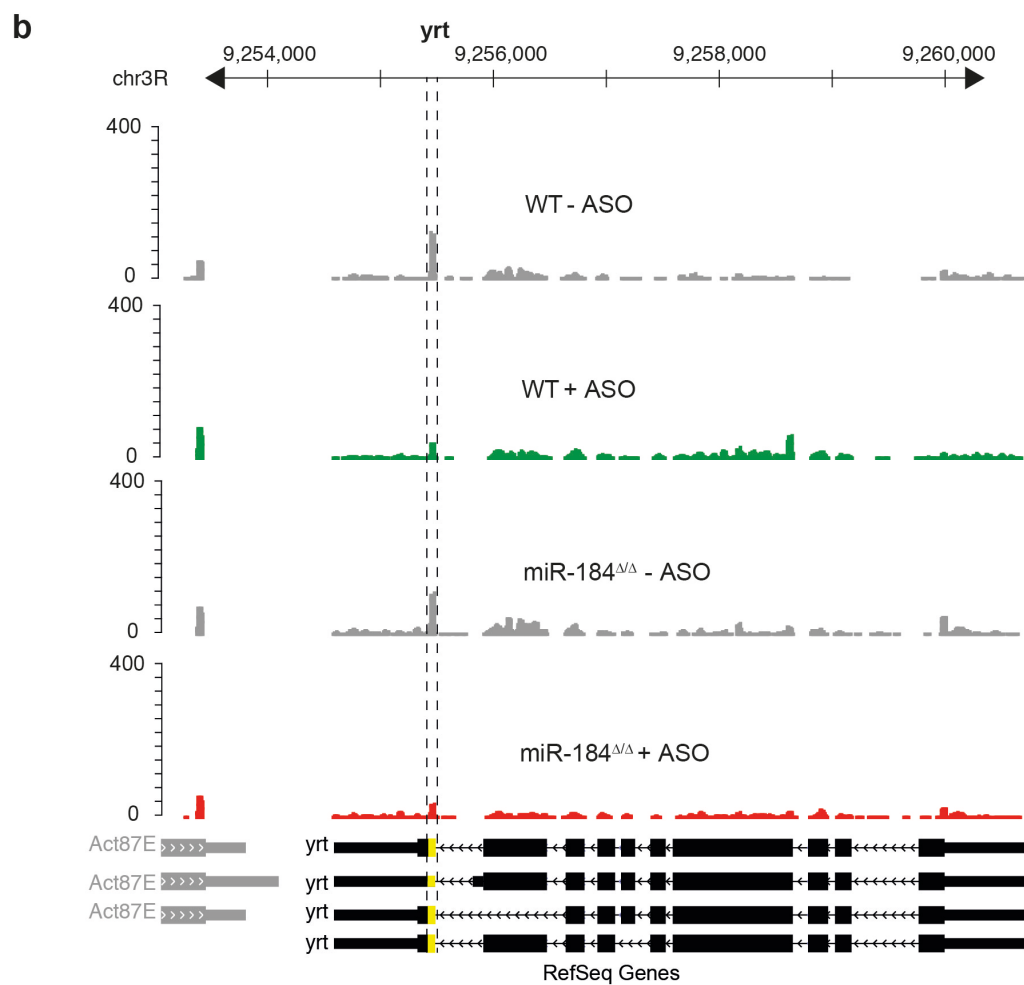
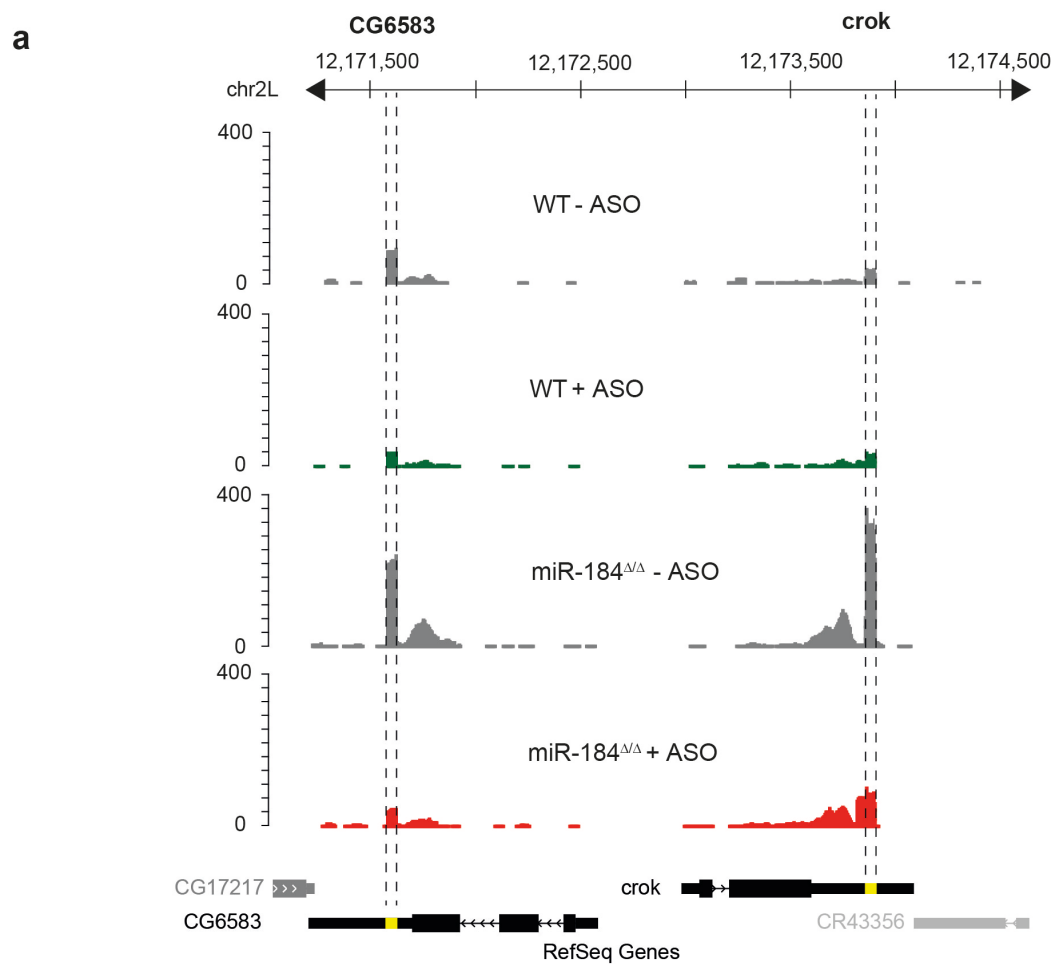


Figure 4.16 | Post-DASH NEXTseq sequencing generates high coverage over target MREs and reflects significant changes in RPS for active MREs. **a**, to visualise read counts over target MREs we represented two active MRE loci (*crok* and CG6583). Consistent with miRISC loss at these MREs both *crok* and CG6583 reveal a stronger depletion in the KO (red) (+ASO vs -ASO) than in the WT (green). **b**, *yrt* harbours an inactive MRE in which the visual difference between WT and KO depletion is not detectable. Raw reads were represented in UCSC genome browser. A single replicate is shown for each condition. Yellow box indicates location of MRE.

4.3.12 Development of quantitative TRACEseq computational pipeline (logistic regression model)

Following the successful ablation of rRNA as well as the significant improvement of read depth, we sought to develop an unbiased quantitative model to assess transcriptome-wide miR-184 binding at the 6mer seed motif. In order to produce such an unbiased prediction based off the NEXTseq sequencing data, we built a logistic regression general linear model (GLM). To train this model we defined a set of true positive miR-184 target sites (see materials and methods, Fig. 2.1). In brief, we repurposed previously generated multiple dosage miRNA overexpression and corresponding mRNA transcriptome data, as well as 17 *bona fide* MREs identified by genome-engineering based evaluation of RNA regulatory element activity (GenERA) (Qianxin Wu, *manuscript under consideration*). This resulted in 53 true positives and 1,383 true negatives per experiment (biological triplicates) with which we trained our logistic regression model. A random subset of 75% was chosen as a training set while the remaining 25% were retained as a validation set. We built the logistic regression model using the following predictors *i*) GFP control normalised read count, *ii*) relative protection score for both WT and KO and *iii*) the TRACE score, which were defined as follows:

$$GFP \text{ normalised read count (GNRC)} = \frac{\text{raw read count motif}_i}{GFP \text{ read count}_i}$$

$$Relative \text{ Protection Score (RPS)} = \frac{GNRC_{+ASO}}{GNRC_{-ASO}}$$

$$TRACE \text{ Score} = \frac{RPS_{WT}}{RPS_{KO}}$$

where i = experimental condition.

We assessed the performance of our classifier through receiver operating characteristic (ROC) curves (Fig. 4.17). The true positive rate (TPR) was plotted against the false positive rate (FPR) at various thresholds. This resulted in an area under the curve (AUC) of 0.68, which is superior to chance detection of 0.5 (Fig. 4.17a).

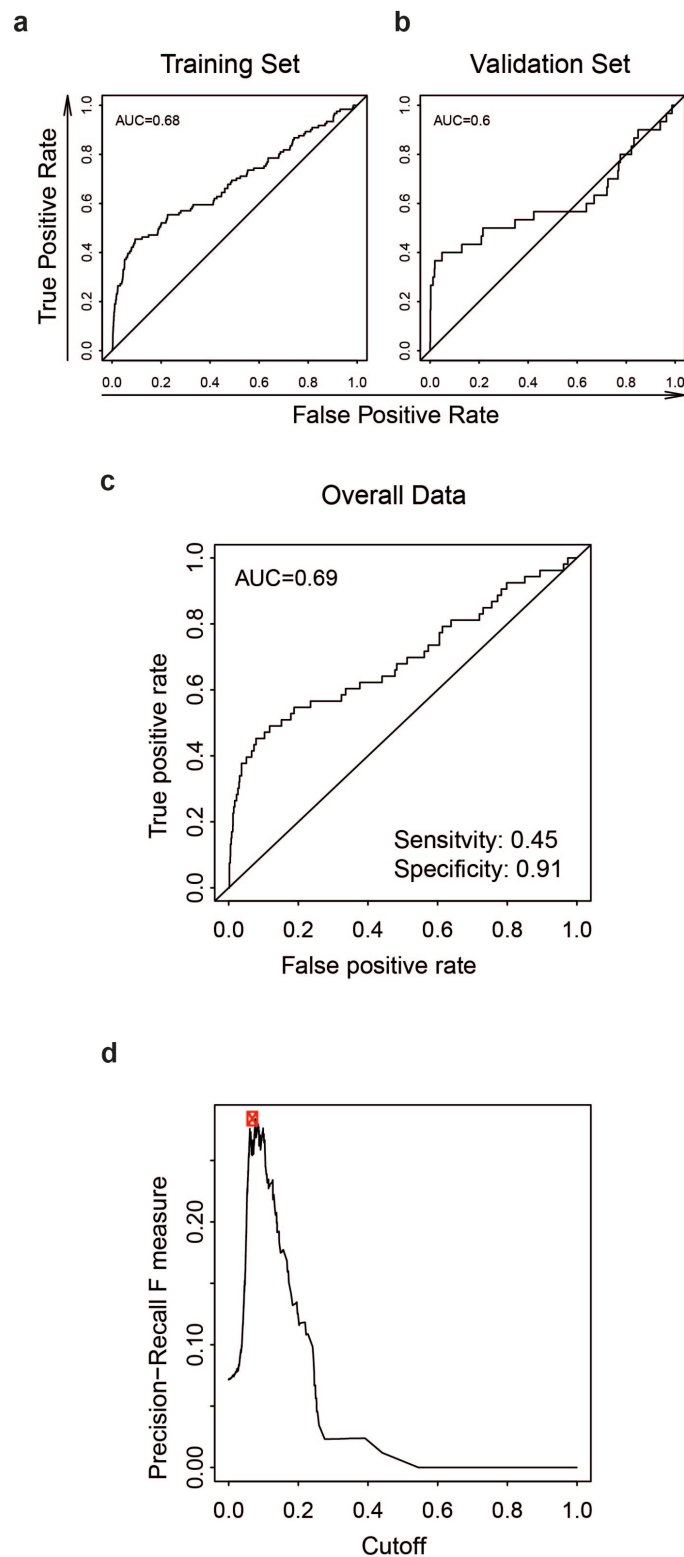


Figure 4.17 | Receiver-operating characteristics of TRACeseq logistic regression model. **a**, the GLM model was trained on 75% randomly chosen training data and revealed an AUC of 0.68. **b**, validating the model on the remaining 25% validation data revealed an AUC of 0.6. **c**, applying the logistics model to all data revealed an AUC of 0.69 with sensitivity of 0.45 and specificity of 0.91. **d**, to establish the ideal cut-off above which to call a prediction positive, the F measure was calculated and revealed a maximum at 0.076. AUC = Area under the curve.

We then tested the GLM using the remaining 25% of the data (validation set) and observed a similar overall ability to correctly discriminate between positive and negative targets (AUC = 0.6; Fig. 4.17b). To minimise technical error we next used the mean read counts of triplicates. Applying the GLM to the entire dataset still displayed an AUC of 0.69 (Fig. 4.17c), and thus retained discriminatory power beyond chance (AUC = 0.5). On this dataset, the GLM performed with a sensitivity of 0.45 and a specificity of 0.91. Although the sensitivity, defined as the rate of correctly identified true positives was sub-optimal (TPR = 0.45), the specificity, defined as the rate of correctly identified true negatives was markedly higher (true negative rate = 0.91).

To optimise the model, we evaluated its performance at various prediction thresholds using the F measure metric. The F measure represents the harmonic mean of precision and recall, defined as the rate of retrieved positives among all predicted and all true positives, respectively. As the maximum F measure was reached at cut-off 0.076, we defined all genes with prediction probabilities above this threshold as miR-184 targets, ultimately qualifying 48 MREs as *hits* and 1,378 as *non-hits*.

4.3.13 TRACEseq reveals enrichment of 8mer seed match in true targets and abundant MREs in UTRs

We next carried out a motif analysis of the two nucleotides immediately adjacent to the 6mer seed sequence (position 1 and 8) (Fig. 4.18a,b). Underscoring the predictive power of TRACEseq, these matched the extended miR-184 sequence. Amongst the *hits*, at position 1 we detected a significant enrichment of thymidine (T), with approximately 40% occurrence, which was significant compared to the expected equal occurrence of 25% (χ^2 test $p=0.04$) (Fig. 4.18a).

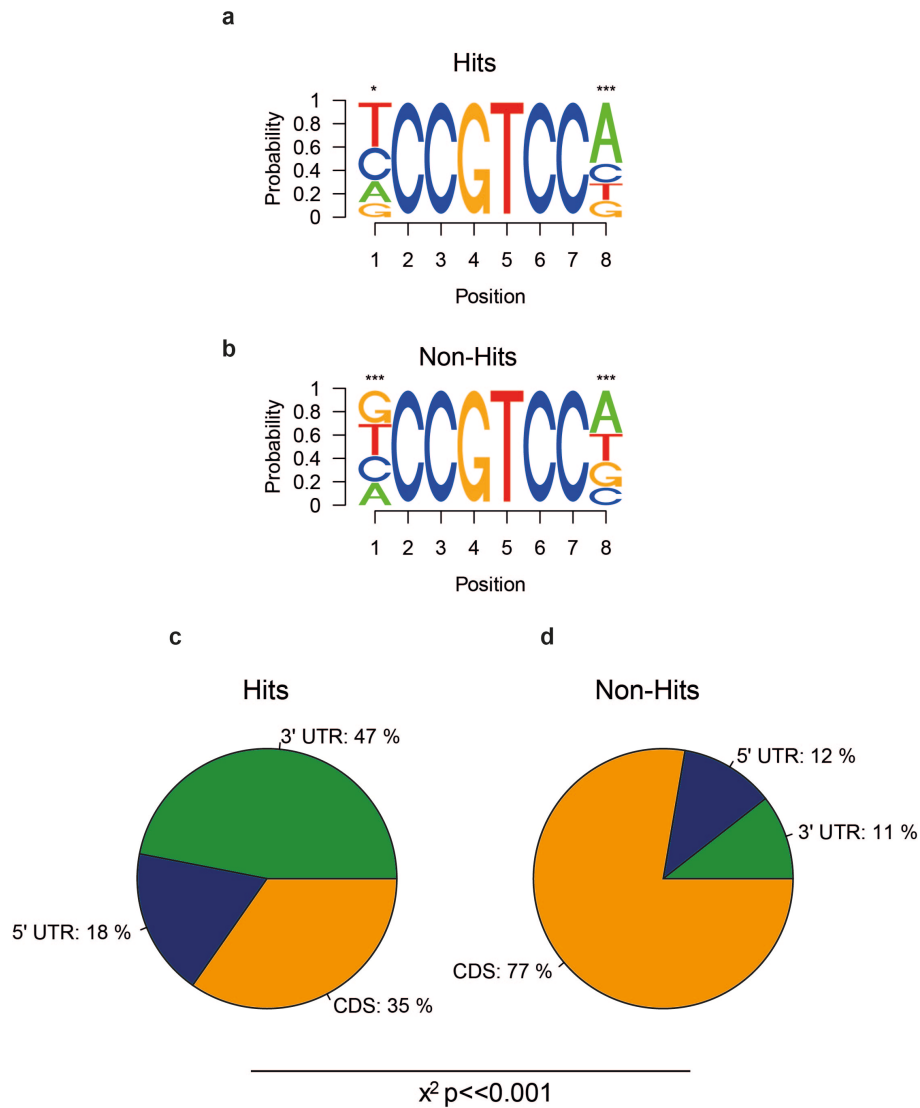


Figure 4.18 | Extended 6mer motif analysis and gene location distribution of hits. **a**, a motif analysis of the hits revealed a preference for T at position 1 and A at position 8, which reflects 8mer sites are the most abundant sites discovered by TRACEseq. **b**, the non-hits display a relatively even distribution at position 1 and 8 with a subtle but significant enrichment of G at position 1. **c**, the distribution of hits (predicted MREs) revealed that nearly half are found in the 3' UTR (47%), 35% are found in the CDS and 18% are found in the 5'UTR. This suggests a specific enrichment of target sites in the 3'UTR as in **d**, conversely, the non-hits are distributed according to the relative lengths of gene structures.

At position 8, the analysis revealed adenosine (A) as the predominant nucleotide, matching miR-184 position 1 (χ^2 test $p < 0.01$) which is consistent with previous

studies (Grimson et al., 2007). Conversely, in the *non-hits*, nucleotide probability at position 1 and 8 were more evenly distributed (Fig. 4.18b). However, guanosine (G) at position 1 and A at position 8 were also slightly enriched. The significance of this phenomenon is most likely due to a much higher sample size in the *non-hits* (n = 1,378) compared to the *hits* (n = 48).

We next analysed the site distribution of 6mers within gene structures (e.g. 3'UTR) in *hits* against *non-hits* (Fig. 4.18c,d). The *non-hits* distribution revealed 12% occurrence of 6mer motifs in the 5' UTR, 78% occurrence in the CDS and 11% in the 3'UTR (Fig. 4.18d). This matched closely the expected distribution of random 6mer motifs based on the reported median lengths of gene structures in *Drosophila*. These are 123bp for 5'UTR, 1242bp for CDS and 246bp for 3'UTR or 7.6%, 77.1% and 15.3% in relative percentages (Maroni, 1994). Notably, in the *hits*, this distribution changed dramatically to 18% 5' UTR, 35% CDS and 47% 3'UTR (Fig. 4.18d). When normalising these distributions for the median length of each gene structure, the enrichment of UTRs over CDS in the *hits* is even more pronounced: 5.2-fold for the 5'UTR and 6.8-fold for the 3'UTR. The *hits* represented a significantly different relative distribution from the *non-hits* (χ^2 test $p < 0.01$). This data is consistent with the observation that a majority of MREs are found in the 3'UTR of genes (Bartel, 2009), and further underscores the capability of TRACEseq to identify *bona fide* functional MREs.

4.3.14 TRACEseq uncovers previously validated and novel MREs

We next plotted all *hits* predicted by TRACEseq (denoted with gene name and location e.g. 3' UTR) against their respective Spearman's coefficients (generated by an anti-correlation analysis against multiple dosage miR-184 bins) (Fig. 4.19, see materials and methods).

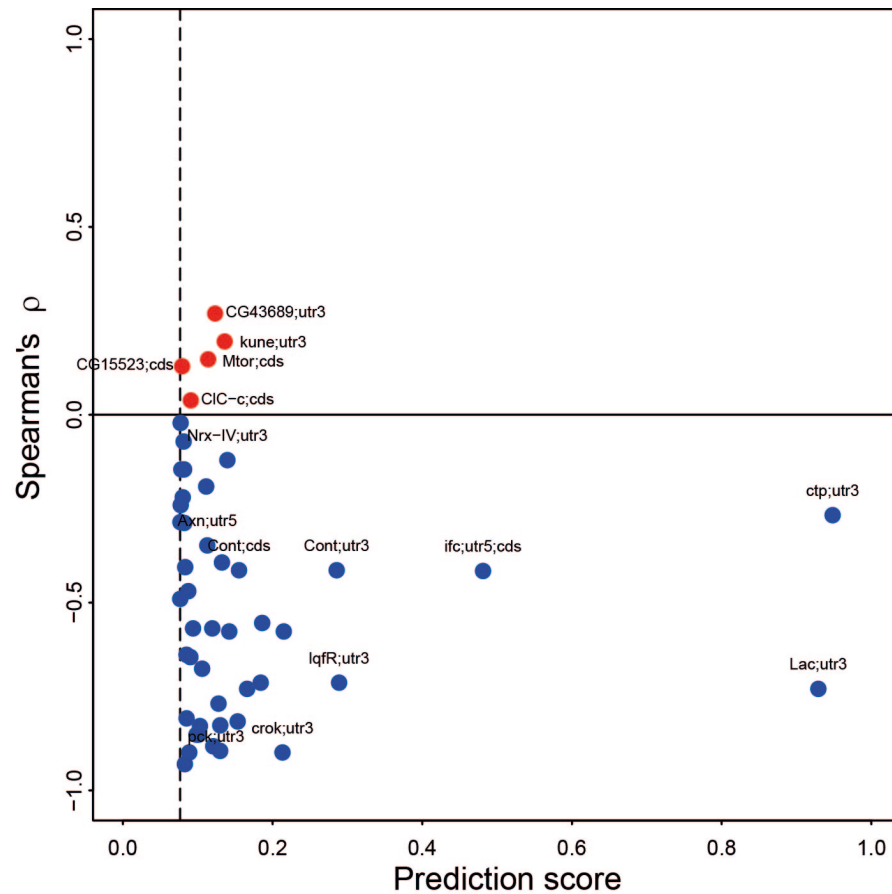


Figure 4.19 | Logistic regression predicted 48 MREs above significance threshold.

Annotation is gene ID and gene location (e.g. 3' UTR). Dotted line represents threshold above which hits were called. Full line represents minimum Spearman's ρ . Red dots represent MREs that are not anti-correlating with miR-184 but perhaps are regulated through translational inhibition rather than mRNA destabilisation.

The majority of *hits* clustered closely to the threshold line at 0.076. Previously validated strong MREs residing in the *crok*, *cont*, *lqfR*, and *Nrx-IV* were uncovered by TRACEseq. Importantly, the model uncovers several novel MREs that have hitherto not been evaluated (e.g. *kune*). As expected, the majority (90%) of GLM-called putative MREs displayed a negative Spearman's ρ , indicating mRNA degradation in response to increased miR-184 expression. Interestingly, five 6mer motifs (10%) displayed a positive Spearman's coefficient, suggesting correlation between miR-184 and regulation independent of mRNA destabilisation.

4.4 Discussion

4.4.1 Quantification of MRE accessibility by multiplex TRACE assay with grouped ASOs

The predicted miR-184 primary target network in *Drosophila* S2R+ cells comprises approximately 90 conserved MREs. The first iteration of the TRACE assay required one unique ASO per MRE and only one MRE could be assessed per reaction. We reasoned that RNase H and co-factors are present in excess in a single reaction, thus permitting the interrogation of multiple MREs in parallel. We developed an algorithm for the rational grouping of ASOs into batches with minimal inter-amplicon crosstalk. We generated TRACE scores using this approach and compared the values to individual *solo* TRACE scores derived from single MRE assays. The results revealed a linear correlation between the two methods supporting the feasibility of multiplex MRE accessibility analysis. However, we noticed a loss of resolution between the inactive MREs and those showing modest activity. Interestingly, we also registered a reduction in absolute cutting efficiency, which is most likely caused by the dilution of RNase H from a single 25mer to about ten ASOs. Although the amount of RNase H could be easily increased, in this instance the dynamic range was sufficient to carry out comparative studies between different conditions.

Reducing off-target events is critical to avoid false positive results (e.g. ASO cuts within amplicon but not over MRE) but also key to ensuring the ASO concentration is not diluted away from the on-target, which would incur a loss of dynamic range (Worley-Morse and Gunsch, 2015). Nucleic acid hybridisation is a highly dynamic process that may involve frequent off-target binding (Woolf et al., 1992). Assessing the cutting of both perfect and imperfectly matched ASOs, Woolf *et al.* determined that off-target cleavage can be engendered from as little as ten consecutive base pairs (Woolf et al., 1992). However, this result was derived from studies carried out in

anurans (*Xenopus* oocyte) cultured at lower ambient temperature (22°C) compared to our experiments which were performed at 37°C. A temperature difference of 15°C is considerable and we reasoned that the increased temperature should reduce off-target base pairing and subsequent cutting. A definitive answer could be provided by Northern blotting of a target RNA in the presence of all other group ASOs, which would clarify the extent of inter-target RNA off-target cutting.

We assumed that any residual off-targets occurring in-between amplicons would most likely involve nucleotides adjacent to the miRNA seed region, since this 6mer motif is integral to all ASOs. However, provided that this is the intended target site we did not consider it significant whether unprotected MREs are hybridised to the cognate 25mer or another ASO through imperfect targeting. Conversely, if the MRE is bound and protected, no ASO should be able to cut it. In either case, we presumed that the cognate ASO outcompetes any imperfectly bound ASOs at any given site by virtue of superior binding kinetics.

A possible caveat of this approach is that off-target analysis was restricted only to the amplicon length. If off-target cutting occurred immediately outside the amplicon sequence, it could conceivably result in transcript degradation and consequent destruction of the primer binding sites used for qPCR analysis. However, as discussed in the previous chapter, these assays were performed under denaturing conditions and in the presence of broad-spectrum RNase inhibitors, which should, in principle, prevent degradation of the target RNA. Furthermore, we concluded that off-target cutting poses a more substantial thread for dynamic *in vivo* systems, which may react *via* compensation (e.g. up-regulate an RNA following its inhibition), and thus distort the results. This risk however, is mitigated by carrying out the analysis on material extracted from fixed (crosslinked) cells.

It has been shown that RNase H can tolerate internal mismatches between the ASO and the target RNA (Woolf et al., 1992). The algorithm presented here did not take this into consideration and only screened for consecutive complementarity. To exercise additional caution the algorithm could be updated for future groupings to accommodate such various mismatch scenarios.

4.4.2 Quantification of MRE accessibility by multiplex TRACE with single ASO

Both individual TRACE assays and multiplex TRACE with grouped ASOs used fully complementary 25mer ASO in a 'one-ASO:one-MRE' modus. Based on the inherent presence of the 6mer motif CCGTCC in the miR-184 predicted target network we designed a strategy to efficiently target this motif in a combinatorial manner. To this end, we formulated a semi-random 10mer ASO of the identity NN-GGCAGG-NN, with the central 6 nucleotides complementary to the miR-184 MRE 6mer seed sequence. At physiological conditions (37°C and low salt), 12bp have been defined as the minimum length for stable ASO:RNA duplex formation (Khan and Coulson, 1993). Consistently, we observed poor RNase H activity when using a short 6mer ASO. However, while the majority of antisense approaches have employed 12-25nt long ASOs, 7mer or 8mers have also been shown to induce potent and selective downregulation of target RNAs (Wagner et al., 1996). Furthermore, RNase H protection analysis using a fully randomised 10mer ASO library has been used to identify accessible regions in Hepatitis C Virus RNA, suggesting that shorter ASOs are useful agents for probing RNA sequences (Lima et al., 1997). This supports the hypothesis that under certain conditions (e.g. high salt) relatively short ASOs (7-10mers) are capable of specific target hybridisation.

In fact, the subject of ASO length has remained somewhat controversial and it frequently results in a trade-off between genomic specificity and off-target prevalence. Statistical calculations suggest that ASOs require a minimum length of

12-15nt to map uniquely to the genome (Simon, 2013). However, the longer the ASO the more consecutive oligomers become available for transient duplex formation at off-target sites (Herschlag, 1991). Furthermore, extended ASO may form secondary structures that could poise selected sub-oligomers for off-target binding. Based on this framework we aimed to create a minimum length ASO that displays sufficient cleavage specificity. Hence, although a 12mer elicited marginally more target cleavage we decided to proceed with a 10mer. We designed a 6mer core sequence flanked by random nucleotides (NN: 25% A, 25% C, 25% G, or 25% T) on either side, resulting in a pool of 10mers that meets optimal duplex stability requirements.

Evaluating the *multiplexed* TRACE assay on twelve predicted MREs we were able to derive TRACE scores closely resembling *solo* cutting with 25mers. Similar to the *grouped* approach we noted a decrease in resolution and found some of the lower activity MREs did not cluster according to their class. We compared *multiplexed* TRACE scores against MRE scores and noted a positive, statistically significant correlation. Although the correlation parameters were lower than those observed with the *grouped* approach, we concluded that multiplexed cutting with a single ASO offered a massive scalability advantage as it could be applied to large, transcriptome-wide MRE networks.

The catalytic activity of RNase H correlates with the stability of the target RNA-ASO heteroduplex. To further improve ASO stability as well as the affinity for target RNAs a host of ASO modifications have been developed (Crooke et al., 1995; Kurreck et al., 2002; Wahlestedt et al., 2000). For example, in 'gapmers' the central unmodified DNA nucleotides are flanked on each side by 2'-O-methyl ribonucleotides or, more recently, locked nucleic acids (LNA). Such chimeras improve the affinity for RNA targets without compromising RNase H activity, as long as the DNA core of the ASO is of sufficient length. For instance, 2'-O-methyl gapmers require only a 6 DNA

nucleotides core to allow efficient RNase H recruitment (Kurreck et al., 2002). However, the same 6mer core only leads to moderate RNase H activity when flanked by LNAs, which work most efficiently if the center is extended to 7 or 8 DNA nucleotides. In general, these modifications confer remarkable affinity improvements, increasing the melting temperature by $\sim 1^{\circ}\text{C}$ per 2'-O-methyl nucleotide and by up to 9.5°C per LNA nucleotide. Perhaps the most distinct advantage of these strategies is that all cleavage is restricted to the 6mer central region only, as RNA base pairing to flanking modified nucleotides is not subject to RNase H cleavage. We envision exploring these implementations in future *multiplex* TRACE iterations.

4.4.3 Seed-targeted RT initiation enables high-throughput NGS-based TRACE analysis

We reasoned that the 6mer seed motif found in the entire predicted miR-184 MRE network offers an opportunity to initiate cDNA synthesis at precisely this position, a strategy reminiscent of primer extension. The purpose of this approach was to generate sufficient number of reads over MREs of interest to enable statistical TRACE score calculations across the entire network. Following iterative optimisation of both primer concentration and hybridisation conditions, we successfully generated a targeted cDNA library specifically from uncut MREs. Read mapping to candidate MREs revealed a significant global enrichment of the 6mer miR-184 MRE seed motif. However, this initial sequencing run lacked sufficient depth to perform meaningful TRACE score calculations. The majority of read counts could be traced to structural RNA loci belonging to ribosomal RNA.

This was surprising, as neither the abundant rRNA sequences found in the library preparation nor *rp49* contained the miR-184 6mer MRE motif, yet they were found in the cDNA preparation (*rp49* was measured before addition of a GSP). This finding was intriguing since it implies that cDNA was synthesised from RNAs that lack

complementary sites to the RT primer. Two explanations for this observation are plausible. First, it has been observed that cDNA can be generated without the addition of an exogenous primer (Sullivan et al., 1980) (Brizzard and De Kloet, 1983), suggesting that self-complementary RNA hairpin structures can 'prime' cDNA synthesis. Interestingly, Sullivan *et al.* described this phenomenon in 25S ribosomal RNA (Sullivan et al., 1980). Although this finding was confirmed by Brizzard *et al.*, it could not be extended to 18S rRNA (Brizzard and De Kloet, 1983). Surprisingly, the majority of top contaminating RNA species in our experiment were 18S rRNA.

The second explanation is based on imperfect or wobble base-pairing. A study investigating the minimum primer length required to initiate reverse transcription by *Tetrahymena* telomerase (a well studied example of a reverse transcriptase) showed that *in vitro* this enzyme can efficiently initiate cDNA synthesis from as little as three DNA base pairs (Baran et al., 2002). Supporting this notion, mispriming of hexamers has been observed across many RNAseq experiments where errors were frequently noted in the first 7 bases of any given read (Hansen et al., 2010; van Gurp et al., 2013). We screened the top 6 contaminating RNA species and found only three 5mer matches for our RT primer, all occurring in the sense direction suggesting that these could not function as priming sites. If cDNA could be initiated off a tetramer match, we would expect to see reads starting with at least four bases of the RT primer. However, most reads in our library did not begin with ACGG. This leaves one last option, namely off-target initiation *via* one of the three GSP intended to quantify *egfp*, *rp49* and *hsp83* for normalisation purposes. We searched the top six contaminating rRNAs against complementarity towards the GSP but found only a single match for the *rp49* GSP which was complementary to the 18S ribosomal RNA pseudogene:CR41602 (chrXHet:195,118-195,438). However, reads mapped broadly to this region and did not appear to start at a precise location. In summary, it remains

difficult to conclude as to how exactly these reads emerged. In our opinion, the most likely explanation is provided by promiscuous cDNA initiation, which does not require a perfect 6mer match. Fortunately, computational tools permit focussing on the 6mer motif of choice and discarding all reads that do not begin with or contain this sequence, so long as sufficient read depth is achieved during the NGS reaction.

In standard RNAseq experiments the removal of rRNA is common practice to generate sequencing space. This is predominantly achieved either *via* poly-A selection or rRNA depletion virtue of RNase H digestion (Benes et al., 2011). Although a popular choice, in our case poly-A selection would only capture MREs most proximal to the poly(A) tail. Thus, it would be impossible to analyse more than one 6mer MRE motif per transcript. As multiple MREs with an identical seed frequently reside in the same mRNA we refrained from using this approach. Unfortunately, the commonly used alternative, commercial rRNA depletion kits have not been developed for *Drosophila* rRNAs. A different approach is provided by so-called terminator blocking oligos which prevent reverse transcription off rRNA (Wickersheim and Blumenstiel, 2013). However, this has only been applied to short 30nt 2S RNA, to remove this species from small RNA library preparation. The most recent strategy to deplete undesired abundant sequences was developed by the DeRisi lab and relies on dsDNA removal by *in vitro* Cas9 digestion (Gu et al., 2016). This approach has the distinct advantage of being applied directly to an existing NGS library, without requiring *de novo* preparation. Adopting this innovative approach resulted in a significant depletion of a ribosomal RNA candidate by 99% without apparently affecting the concentration of our genes of interest (*crok* or the control *egfp*). As a second provision, we upgraded our sequencing platform from MiSeq to NextSeq, engendering ~25 times higher read depth.

4.4.4 Development of quantitative TRACEseq computational pipeline (logistic regression model)

We developed a TRACEseq computational pipeline for the identification of transcriptome-wide miRNA binding events. The GLM displayed a sensitivity of 0.45, suggesting that the rate with which true positives were identified was only moderate. The specificity, defined as the rate of correctly identified true negatives, was markedly higher at 0.91. This suggested that the strength of TRACEseq lies in a low false negative rate; positive MREs are predicted at high confidence. However, this is most likely a conservative estimate and it is probable that additional true positives are amongst the training data that have not been assigned as such. As a consequence, the performance of the model is probably underestimated.

TRACEseq predicted 48 putative miR-184 MREs (*hits*). We performed an enrichment analysis of the 6mer motif with its two immediately adjacent neighbours. Interestingly, in the *hits* we noted significant enrichment of T at position 1 and A at position 8. These nucleotides are complementary to position 8 and 1 of the miR-184, respectively. This suggested that 8mer seed matches are the predominant target sites identified by TRACEseq. This is plausible as 8mer target sites confer most potent repression upon the mRNAs they are embedded in (Agarwal et al., 2015; Bartel, 2009; Lewis et al., 2005), which is most likely reflected in the robust assembly of the miRISC at those sites.

We next compared the relative location and distribution of 6mer motifs within gene structures (e.g. CDS vs. 3'UTR) in *hits* and *non-hits*. The distribution in *non-hits* revealed a preponderance of 6mer motifs in the CDS, which is expected if 6mer distribution is random across the genome, as CDS are on average significantly longer than UTRs (Maroni, 1994). Importantly, in the *hits*, this distribution changed towards a preference for the 3'UTR. The results of TRACEseq are thus in full

agreement with the current notion that MREs which are most frequently located in 3'UTR regions (Bartel, 2009).

A significant number of miR-TRACEseq predicted targets (15/48) have been orthogonally validated to be regulated by miR-184 (Qianxin Wu, *manuscript under consideration*). Additionally, our model predicted many novel targets. As miRNAs have been heralded as master regulators of genes participating in common biological processes, we sought to globally characterise these in terms of biological function. Gene Ontology (GO) enrichment analysis revealed these genes converge at the functional level and are implicated in nervous system development. Specifically, they are associated with *glial cell development* (GO:0021782, adjusted $p = 9.04 \times 10^{-6}$), *glial cell differentiation* (GO:0010001; adjusted $p = 3.36 \times 10^{-5}$) and *nerve maturation* (GO:0021682; adjusted $p = 0.0001$). More generally, predicted targets are enriched for cell-cycle components such as the *mitotic spindle* (GO:0072686; adjusted $p = 0.0166$) and the *microtubule organising centre* (GO:0060856; adjusted $p = 0.0208$). These enrichments are consistent with the presumed role of miR-184 during nervous cell development, as suggested by its spatiotemporal expression pattern (Li et al., 2011) and the neuromuscular junction and bouton abnormalities observed in miR-184 *Drosophila* null-mutants (Peng et al., 2015). For example the human orthologue (*CNTNAP2*) of the *Nrx-IV* gene, encoding a neurexin required for the formation of glutamatergic neuromuscular junctions, is a critical protein during synaptic formation and has been associated with several neurodevelopmental disorders, including mental retardation and autism (Betancur et al., 2009).

90% of all *hits* revealed negative Spearman coefficient. This is consistent with the paradigm that miRNAs predominantly regulate target gene expression through mRNA destabilisation, which is typically analysed by RNAseq (Guo et al., 2010). However, a caveat of all RNA-based profiling approaches is that mRNA

destabilisation is an essential requirement for differential detection. Conversely, any miRNA::target interactions mediating translational inhibition is masked. About 10% of the *hits* identified by TRACEseq, displayed a modest positive correlation with miR-184 (Spearman's $\rho > 0$). *Hits* that barely correlated ($\rho \sim 0$), for example *CIC*, are perhaps candidates for translational inhibition, which occurs far less frequently than mRNA destabilisation. The *hits* displaying stronger positive correlation suggest an mRNA stabilising effect of miR-184 binding. While miRNA-mediated post-transcriptional upregulation has been reported in a few number of cases (Vasudevan et al., 2007; Ørom et al., 2008), this mode acts through translation activation rather than mRNA stabilisation and hence fails to explain this observation (Vasudevan, 2012). Uncovering the functional output of these few miRNA binding events will entail mutagenesis of the MREs in question, followed by a tandem readout of both RNA and protein level. However, if these turn out to be functional, this would suggest TRACEseq is capable of uncovering miRNA-mediated translational repression events, which had hitherto been limited to proteomics coupled readouts, ribosome profiling, or fluorescent reporters.

4.4.5 Comparison to other high-throughput *cis/trans* target identification techniques

The majority of existing transcriptome-wide RBP profiles, including those of miRISC ternary complexes, stemmed from RIP or CLIP approaches (Cambronne et al., 2012; Chi et al., 2009; Meier et al., 2013; Zhang et al., 2007). Arguably the workhorse for modern day miRNA::target identification studies, the broad implementation of AGO-HITS-CLIP has been limited by its complex experimental and computational framework (Chi et al., 2009). CLIP and PAR-CLIP approaches uncover miRISC binding sites on target RNAs with high-resolution, however, identifying the corresponding miRNA that guides the miRISC complex to this very site has to be

inferred computationally (Hausser and Zavolan, 2014). CLASH and iPAR-CLIP circumvent this problem through intermolecular *in situ* ligation of miRNAs and their targets. However, to date, this is an inefficient process as only 2% of reads are miRNA-mRNA chimeras (Grosswendt et al., 2014; Helwak et al., 2013).

Furthermore, immunoprecipitation is presently limited to AGO2, which is only one of four AGOs in humans (Su et al., 2009). Moreover, this protocol has been restricted to certain tissues and organisms. For instance, the application of AGO-HITS-CLIP to *Drosophila* miRNA studies has so far remained unsuccessful, presumably due to the lack of a suitable Ago antibody. Unlike AGO-CLIP, TRACE does not provide a global view of miRNA::target binding but rather reflects the miRISC occupancy of a candidate miRNA at all its primary endogenous targets. The present use of 6mers implies a confinement to miRNA seeds which means only canonical (seed) target sites are investigated in this approach. However, this obviates the need for deconvoluting miRNA::target binding events by complex computational approaches, which have caused difficulty and controversy. In general, the interpretation of TRACE data is relatively straightforward through the simple calculation of a TRACE score.

Offering a simple high-throughput interaction detection method, single miRNA or single mRNA capture approaches with subsequent sequencing of their associated mRNAs or miRNAs have been developed. These are of particular utility if a candidate miRNA or a candidate already exists. However, if these rely on artificial tethers, abundant concerns have been raised with regards to their physiological relevance (Cloonan, 2015). For example, mRNA bait approaches such as MS2-TRAP require the insertion of exogenous RNA elements (e.g. 6-12 copies of 58 base pair MS2 loops) into the 3' UTR (Ngan K Voa et al., 2010; Yoon et al., 2012). Although these elements can be designed to avoid putative MREs it often remains unclear how this affects higher order RNA structures *in vivo* and thus influences miRNA binding.

It has been more difficult to identify the target sites of RBPs. One approach that engenders a 'footprint' of RBPs on their target RNA is PAR-CLIP. Castello *et al.* have captured all polyadenylated transcripts and successfully used this approach to map RBPs to target RNAs with nucleotide resolution (Castello *et al.*, 2012). However, a general limitation of PAR-CLIP is the obligatory incorporation of modified nucleosides into nascent RNA. While this is feasible in cell culture systems and has been successfully used in *C. elegans*, it has remained difficult to introduce chemical analogues into other animal model organisms (e.g. flies). TRACE circumvents this caveat *via* analysis of protein occupancy of endogenous RNA *in situ*, obviating the need for modifying the target RNA or the RBP.

Perhaps its greatest advantage, TRACE works directly on native, endogenous RNA without the need for any external interaction or transfection of artificial constructs. Given their cardinal role as homing sites for effector proteins, several high-throughput technologies for the analysis of *cis* regulatory elements have recently been developed. However, the vast majority of high-throughput implementations assess *cis* elements under non-physiological circumstances and outside their native context. For instance, Wolter and colleagues recently presented luminescent identification of functional elements in 3' UTRs (3'LIFE) (Wolter *et al.*, 2015). This setup is an upscaled version on the classical sensor assay: UTR sequences of interest are fused to luciferase reporters and their regulatory capacity is read out as a function of reporter translation. In spite of its convenient readout, this assay analyses regulatory potential under entirely artificial circumstances.

A higher throughput technique was developed by Zhao *et al.* and termed massively parallel functional annotation of sequences from 3'UTRs (fast-UTR) (Zhao *et al.*, 2014). In this implementation the authors generated pools of 200nt oligomers that recapitulated a wide spectrum of 3'UTR sequences. The regulatory effect conferred

by each sequence was assessed by deep sequencing and paired analysis of cRNA/gDNA, following doxycycline-mediated transcription arrest. A similar alternative that engenders near nucleotide resolution was recently proposed by Wissink *et al.* (Wissink *et al.*, 2016). The authors developed an artificial sensor system in which short sequences can be tested for their regulatory potential through their insertion into a fluorescent reporter. Fluorescence activated cell sorting (FACS) separates cells according to their GFP level; cells with a high signal contain positive regulating sequences, cells with a low signal harbour repressive sequences. High-throughput sequencing of each population reveals the identity of the regulatory sequences. The most recent and arguably most powerful *cis* regulatory RNA discovery platform was just presented by the Giraldez lab (Yartseva *et al.*, 2016). Circumventing the need for *de novo* oligomer synthesis, the technology termed RNA-element selection assay (RESA) relies on sequence amplification followed by random fragmentation of large regulatory regions ('bashing'). Reporter libraries carrying overlapping RNA sub-regions are then transduced into cells or zebrafish embryos and differential selection (destabilisation vs. stabilisation) is analysed by RNAseq.

Albeit powerful and permitting the investigation of large regions, these synthetic constructs ignore critical contextual features such as RBP occupancy, RNA secondary structure or alternative poly-adenylation that have been shown cardinal for regulation of miRNA targeting (Kertesz *et al.*, 2007; Nam *et al.*, 2014). Although these sensor assays are carried out *in vivo*, the putative *cis* elements only show capability of regulating artificial reporters, which in all cases will require orthogonal validation. Furthermore, the use of exogenously transfected constructs prohibits the investigation of elements of interest under truly physiological homeostasis.

Perhaps the most elusive question in the field of miRNA research is how miRNAs choose their targets. Besides site intrinsic factors (e.g. site type), two additional

layers of control have been proposed: RNA secondary structure (RSS) and binding of RBPs in proximity (or overlapping) of MREs (Kertesz et al., 2007; Kundu et al., 2012). In its current form, the TRACE technology cannot distinguish between these two mechanisms. However, we reasoned that TRACE data could be readily integrated with high-throughput RSS approaches to dissect the contribution of RSS *versus* RBP masking (Wan et al., 2015).

Finally, in addition to studying miRNAs or RBPs, the genome-wide TRACE studies could provide a powerful platform to assess the consequences of off-target hybridisation in therapeutic approaches. Antisense therapy *via* RNAi, DNA oligomers as well as steric-blocking oligonucleotides have received massive attention over the past years due to their therapeutic potential (Kole et al., 2012). High-throughput TRACE data contains ample information on transcriptome-wide oligonucleotide binding kinetics, rendering this approach an attractive source for therapeutic off-target analysis.

4.4.6 Final Conclusions

Efficient high-throughput miRNA::target identification technologies remain in high demand. In this chapter we have pipelined miR-TRACEseq, the first transcriptome-wide miRNA::target identification platform that relies on RNase H protection analysis of *in vivo* crosslinked miRNA targets. miR-TRACEseq enables the parallel analysis of multiple MREs or an entire miRNA target network under endogenous conditions. The application prospect of investigating the genome-wide accessibility of defined *cis* motifs extend well beyond miRNA::target identification.

Chapter 5: Towards Characterisation of the *bicoid* short RNA Interactome

5.1 Introduction

The establishment of axial polarity in many animals ranging from insects to mammals relies on the deposition of maternal mRNA into the developing oocyte (Jansen, 2001; Ma et al., 2016; Slack, 2014). In *Drosophila* for example, both the anterioposterior (AP) and the dorsoventral (DV) axes are specified during oogenesis by the asymmetric localisation of mRNAs encoded by the genes *bicoid* (*bcd*) and *gurken* (*grk*). These mRNAs undergo extensive post-transcriptional regulation in time and space before the protein products fulfil their roles in development. Interestingly, although *gurken* mRNA is actively translated throughout oogenesis, *bcd* mRNA appears to remain translationally quiescent until egg activation (Berleth et al., 1988; Driever and Nüsslein-Volhard, 1988). However, early transcription of *bcd* mRNA in the nurse cells and transport to the anterior end of the developing oocyte constitutes a key determinant for the establishment of the embryonic AP axis (Driever and Nüsslein-Volhard, 1988). The localisation of *bcd* mRNA was shown to be dependent on *cis*-regulatory elements in its 3'UTR as well as direct or indirect interactions with the proteins Swallow, Exuperantia and Staufén (Ferrandon et al., 1994; Riechmann and Ephrussi, 2004; Schnorrer et al., 2000; Weil et al., 2010). As a result of its accumulation at the anterior margin, the Bicoid protein functions in early embryos as a morphogen acting both as a transcription factor activating anterior patterning genes (e.g. *hunchback*) as well as a post-transcriptional factor repressing *caudal* mRNA (Dubnau and Struhl, 1996; Struhl et al., 1989). While *bcd* mRNA localisation and protein function have been extensively studied, the mechanisms underlying its translational repression prior to egg activation are far less clear. Although the effect

of premature, ectopic Bicoid expression in the ovary has never been observed, it is plausible that its consequence would be detrimental. Underscoring the robustness of *bcd* mRNA silencing, introduction of eight extra copies of the *bcd* gene into female flies did not result in protein expression in the ovary (Driever et al., 1990).

Previous studies suggested that the regulation of poly-A tail length represents one prominent post-transcriptional mechanism controlling the fate of mRNAs during oogenesis (Richter, 1991; Vassalli et al., 1989). Indeed the elongation of the poly-(A) tail of *bcd* mRNA following egg laying appears to be necessary for its productive translation (Eichhorn et al., 2016; Sallés et al., 1994). However, while this process was shown to be dependent on the polyadenylase Wispy and the anaphase-promoting complex activator Cortex, the temporal specificity of this regulation remains controversial (Benoit et al., 2008; Lieberfarb et al., 1996). Thus, although Wispy is expressed and promotes polyadenylation of translationally active mRNAs during oocyte maturation, *bcd* mRNA remains incompetent for translation (Benoit et al., 2008). This observation suggests the existence of a *trans* acting factor antagonising the action of Wispy in a *bcd*-specific manner.

Eukaryotic cells contain specialised structures termed processing bodies (P-bodies), which appear to function as central hubs for mRNA metabolism including RNA destabilisation (Chen and Shyu, 2013). Interestingly, a recent ultra-structural analysis provided evidence that in the oocyte, *bcd* mRNA is enriched six-fold in the centre of P-bodies (Weil et al., 2012). In contrast, actively translated *gurken* mRNA is found at the periphery of such structures (Weil et al., 2012). A major constituent of P-bodies is represented by microRNA (miRNA) machinery core components (Sen and Blau, 2005). The overt consequence of miRNA action is post-transcriptional silencing *via* mRNA destabilisation or translational inhibition. This activity is frequently mediated by deadenylation of target mRNAs as a consequence of recruitment of deadenylases

belonging to the CCR4-NOT complex (Eulalio et al., 2009). Furthermore, it was recently reported that another family of short non-coding RNAs, piwi-interacting RNAs (piRNAs), can also deadenylate maternal transcripts via this pathway in *Drosophila* (Rouget et al., 2010). In contrast to miRNAs, piRNAs tend to display longer regions of complementarity with their targets, which are primarily represented by transposable genetic elements. In the aforementioned study, piRNAs originating from the *roo* and *412* transposons were reported to promote deadenylation and subsequent decay of *nanos*, another maternally encoded mRNA required for patterning of posterior structures (Rouget et al., 2010). It is therefore plausible that *bcd* mRNA post-transcriptional silencing during oogenesis, and the subsequent transition to a translationally competent state at the onset of embryogenesis, is mediated in part by short regulatory RNAs. Supporting this hypothesis, it was reported that clearance of *bcd* mRNA requires the activity of Pumilio, an RNA binding protein (RBP) recently shown to facilitate targeting of p27-3'UTR by miR-221/miR-222 (Gamberi et al., 2002; Kedde et al., 2010).

Several studies have confirmed the existence of miRNAs in the *Drosophila* germline and their capacity to actively repress protein translation (Neumüller et al., 2008; Reich et al., 2009). One of the studies detected 43 miRNAs by qPCR in the germline (Neumüller et al., 2008). In mid-to-late stage oocytes, the Gaul lab reported the first germline-specific phenotype in *Drosophila* miR-184 null mutant animals (Iovino et al., 2009). miR-184 loss of function resulted in defects in differentiation of germ stem cells, DV patterning, and AP patterning, processes mediated through regulating the expression of *saxophone*, *k10*, and *tramtrack69*, respectively. Reich and colleagues further demonstrated the activity of miRNA-mediated translational repression in the developing oocyte using a GFP sensor that had been fitted with multiple miR-312 binding sites (Reich et al., 2009).

The involvement of miRNAs in *bcd* silencing could in theory be tested by global miRNA pathway mutagenesis (e.g. *dicer* *-/-* or *ago* *-/-*) followed by antibody staining against Bicoid protein in the oocyte. However, this approach is hindered by the lethality or at least insufficient development of egg chambers derived from most of miRNA pathway mutants (Hatfield et al., 2005; Yang et al., 2007). For instance, *ago1* mutagenesis leads to loss of germ line stem cells (GSC) (Yang et al., 2007). *Dcr-1* absence results in GSC maintenance and self-renewal defects (Jin and Xie, 2007). Clonal analysis of ovaries from female flies lacking *ago1* revealed that a majority of mutant egg chambers do not develop an oocyte and contain only 8 nurse cells (Azzam et al., 2012). These results were phenocopied by epistatic miRNA pathway mutants, such as *dicer-1*, *pasha* and *drosha* which underpinned defects in germline cell division and oocyte determination (Azzam et al., 2012). In summary, these modified genetic backgrounds result in abnormal oocyte development, which fails to attain the stage required for assessing ectopic Bicoid expression by immunostaining. As a consequence, screening for *trans* acting factors *via* global disruption of miRNA biogenesis remains challenging.

A reciprocal approach could entail identification of regulatory factors through mutagenising parts of the *bcd* sequence in *cis*. While this approach was successful in defining *cis* elements required for proper *bcd* localisation, experiments directed towards understanding the regulation of Bicoid protein expression have remained elusive. Driever *et al.* noticed several potential upstream ORFs (uORF) that had elsewhere been implicated in translational repression (Driever et al., 1990; Kozak, 1984). Replacing the 5' UTR of *bcd* with *Xenopus* β globin 5' UTR, which lacks uORFs and secondary structure, resulted in a massively enhanced *in vitro* translation efficiency. Disappointingly however, transgenic delivery of this construct did not lead to ectopic translation of Bicoid in the ovary (Driever et al., 1990; Wharton, 1992).

Likewise, substituting the 3' UTR of *bcd* with the *alphatub84B* 3' UTR revealed no relief of translational repression (Wharton, 1992). This suggests that *bcd* translational control is more complex than originally anticipated and could entail the interplay of redundant mechanisms. Alternatively, it is also possible that key *cis* regulatory elements reside in the ORF rather than UTRs.

The relatively stereotypical rules governing short non-coding RNA binding enable prediction of putative targets using bioinformatics algorithms. In conjunction with genome-wide profiling data, such predictions can guide candidate-based functional studies. To this end, we sought to explore the miRNA and piRNA expression profiles of *Drosophila* ovary tissues and match these to putative target sites encoded in the *bcd* gene. Of note, although a useful resource for guiding candidate-based miRNA studies, *in silico* prediction algorithms have high false positive rates (up to 63%) (Rajewsky, 2006). Therefore, all candidates need to be experimentally validated and this strategy is only valuable when performed in conjunction with a binding assay.

To explore putative regulatory *cis* elements, including several predicted MREs, that may serve as binding sites for repressors (e.g. miRNAs) we sought to map the *bcd* 3' UTR accessibility before and after the oocyte-to-embryo transition (OET). To this end, we performed TRACE (see chapters 3 and 4) to determine RNase H sensitivity-based accessibility scores for a total of 38 consecutive regions and subsequently compared *bcd* 3'UTR accessibility in oocytes *versus* embryos. As this approach is specifically directed at uncovering differentially occupied *cis* regions, I also considered a complementary approach to identify potential *trans* factors.

I reasoned that the most adequate approach for an unbiased identification of non-coding RNA *trans* factors would be a *bcd* mRNA pulldown followed by sequencing of the co-purified short RNA interactome. Several groups have previously performed *bcd* pulldowns from ovary lysates, with varying degrees of success. For example,

through immobilisation of BLE1, Macdonald and colleagues were able to identify Exl as an interacting partner (Macdonald et al., 1995). A similar RNA chromatography approach employing both the IV and V stem loops of the *bcd* 3'UTR co-purified several putative *trans* interacting factors, including Swa and Modelo (Mod) (Arn et al., 2003). Albeit compelling at the time, it has become apparent that this approach failed to discern direct from indirect interacting partners. For instance, a more recent study applying advanced light microscopy imaging revealed that Swa does not directly interact with *bcd* and is in fact not part of the same RNP (Weil et al., 2010). This warrants special care when trying to isolate *bcd* mRNA and its putative regulatory factors.

To design an effective strategy for biochemical isolation of *bcd* interacting RNAs, I evaluated the suitability of several recently developed RNA isolation procedures. Latest advances in the long non-coding RNA (lncRNA) field relied on crosslink-based pulldowns which provided invaluable information regarding the genomic target sites as well as the RNA and protein interactomes of lncRNAs (Chu et al., 2011; Kretz et al., 2012; Simon et al., 2011; Vance et al., 2014). Two related experimental protocols underlie these new technologies. The Chang lab developed an approach termed Chromatin Isolation by RNA Purification (ChIRP) which relies on *in vivo* chemical (formaldehyde) crosslinking of lncRNAs:protein adducts to chromatin, followed by hybridisation of the lncRNA transcript with dozens of tiled biotinylated oligonucleotides and subsequent streptavidin capture (Chu et al., 2011). A complementary approach called Capture Hybridisation Analysis of RNA Targets (CHART) has been independently developed in the Kingston laboratory and also relies on antisense mediated capture of target RNAs (Simon, 2013; Simon et al., 2011; 2013). The key difference between ChIRP and CHART is their use of oligonucleotides. ChIRP employs hundreds of oligonucleotides with the aim to target

any accessible region (devoid of secondary structure and RBPs). In contrast, CHART employs probe design guided by RNase H sensitivity maps (c.f. TRACE chapter 3) and uses only select-few antisense oligonucleotides against accessible regions. While the authors of the ChIRP method argued that using a limited number of probes yielded low transcript recovery, we considered intelligent probe design favourable for our particular purpose. We reasoned that in order to avoid targeting probes against putative *cis* elements (e.g. predicted MREs) and thereby isolating only unbound, non-crosslinked *bcd* transcripts (as *trans* factor and probe binding is mutually exclusive at the same location), intelligent probe design is preferable.

A distinct approach termed Interactome Capture was reported by Castello and colleagues (Castello et al., 2012; 2013b). Employing oligo-d(T) beads the authors enriched for polyadenylated RNAs from UV crosslinked lysates and subjected the co-purified protein fraction to mass spectrometry analysis. This approach provided for the first time a comprehensive overview of RBPs in HeLa cells and has recently also incorporated the unbiased discovery of corresponding RNA binding domains (RBD) (Castello et al., 2012; 2016). The same protocol has also been successfully applied to *Drosophila* embryos where distinct RNA binding proteomes were analysed across two developmental stages (Sysoev et al., 2016). UV light offers superior crosslinking specificity over formaldehyde as only very close proximity (i.e. direct) RNA-protein interactions are captured (Pashev et al., 1991). Since protein-protein interactions are not stabilised, this technology offers arguably the ‘cleanest’ representation of *in vivo* RBP-RNA interactions. While biotinylated oligonucleotides have been successfully used in various experimental settings, the delivery of streptavidin coated capture agents has the caveat of binding with high affinity to not just the intended oligonucleotides but also to any biotinylated protein in the cell. Another technical advantage provided by Interactome Capture is its use of streptavidin-free, covalently-

coupled oligo-d(T) beads. Taking advantage of this system, we designed modified oligonucleotides that can be covalently coupled to magnetic beads, circumventing the need for the commonly used biotin-streptavidin chemistry. In summary, to enable specific isolation of the *bcd* RNA interactome, we developed a protocol which combined intelligent probe design and an adaption of Castello's Interactome Capture.

In this chapter, I present a new integrated approach towards identifying miRNAs and their putative MREs in the *bcd* transcript. To this end, I deployed Nanostring profiling of ovary expressed miRNAs, exploratory TRACE mapping across the *bcd* 3' UTR and small RNA interactome capture. This data provides the first link between *bcd* post-transcriptional regulation and short regulatory RNAs.

5.2 Chapter Aims

- Analysis of small RNA profiles in ovaries and *in silico* screen for potential interactions with *bcd* mRNA
- Mapping the *bcd* 3' UTR for regulatory *cis* elements by TRACE across the oocyte-to-embryo transition
- Development of highly specific enrichment of *bcd* mRNA for co-purification of *trans* interacting RNAs (small RNA Interactome Capture)
- Preliminary functional characterisation of *bcd cis* regulatory elements

5.3 Results

5.3.1 *Drosophila* ovaries express miRNAs and piRNAs predicted to target *bcd*

To establish the *Drosophila* miRNA complement during oogenesis, we used the Nanostring nCounter technology, a powerful hybridisation-based platform for the specific and sensitive, simultaneous detection of up to 800 target miRNAs without amplification (Geiss et al., 2008). We dissected wild type Oregon R ovaries and separated egg chambers into four distinct developmental stages, to generate a temporal profile of miRNAs expressed at: *i*) germarium-stage 8 (pre-9), *ii*) stage 9, *iii*) stage 10-13 (post-9), and *iv*) stage 14. Nanostring analysis revealed a total of 40 mature miRNAs with counts ≥ 100 in at least one of the four bins (Table 5.1). Computational prediction using the *miRanda* algorithm revealed that several highly abundant miRNAs are predicted to target *bcd* at varying locations (Table 5.1). Amongst these, we identified various seed class architectures, ranging from 6mers to more powerful 8mers.

Table 5.1: miRNAs expressed in the ovary

miRNA	Nanostring Counts				Prediction	
	Pre-9	9	Post-9	14	Seed	Location
dme-miR-9c-5p	10197	12758	8064	14776		
dme-miR-92a-3p	7355	8358	6838	6755		
dme-miR-318-3p	1404	10891	7871	8575		
dme-miR-279-3p	9810	6827	4010	3782	7mer	3'UTR
dme-miR-8-3p	15001	4730	1662	1955		
dme-miR-275-3p	4480	5182	2079	2322		
dme-miR-92b-3p	3075	3853	3796	2707		
dme-miR-989-3p	9444	2998	285	163	8mer/7mer	5'UTR/3'UTR
dme-miR-306-5p	4111	2815	2833	2175		
dme-miR-304-5p	5079	2501	1393	2192		
dme-miR-13b-3p	2568	2128	1526	2718		
dme-miR-994-5p	497	3564	2008	2728		
dme-miR-184-3p	4518	1351	1066	980		
dme-miR-311-3p	787	2209	2601	1490		
dme-miR-310-3p	696	1954	1176	1465		
dme-miR-276a-3p	1589	873	646	1349		
dme-miR-995-3p	1137	841	743	914		
dme-miR-996-3p	1651	686	618	502	7mer	3'UTR
dme-miR-986-5p	346	635	529	1837		
dme-miR-11-3p	1074	627	520	1042		
dme-miR-9b-5p	679	876	491	595		
dme-miR-12-5p	1562	437	317	213	6mer+G:U	3'UTR
dme-miR-312-3p	239	660	845	526		
dme-miR-283-5p	1325	250	166	217	7mer	3'UTR
dme-miR-125-5p	682	331	414	351		
dme-miR-1010-3p	174	277	240	572	7mer	CDS
dme-miR-277-3p	235	50	244	509		
dme-miR-100-5p	281	216	192	244		
dme-miR-999-3p	234	238	204	222	8mer/6mer+G:U	3'UTR
dme-miR-1012-3p	137	264	218	272		
dme-let-7-5p	279	145	129	278		
dme-bantam-3p	270	173	135	161		
dme-miR-313-3p	47	226	238	200		
dme-miR-970-3p	139	159	178	139		
dme-miR-2b-3p	196	122	159	108		
dme-miR-988-3p	152	127	110	96		
dme-miR-2a-3p	215	74	73	122		
dme-miR-34-5p	228	21	20	85		
dme-miR-1003-3p	89	71	73	115		
dme-miR-1-3p	191	3	19	8	8mer+G:U	3'UTR

Nanostring analysis of four miRNA complements isolated from dissected *Drosophila* ovaries. Only miRNAs with counts ≥ 100 in at least one bin are displayed. Full length bcd mRNA was aligned against all expressed miRNAs using the miRanda algorithm. Several miRNAs are predicted to target bcd (bold) with varying seed architectures and locations across the bcd transcript.

A comprehensive analysis of the piRNA complement during *Drosophila* oogenesis and embryogenesis has been previously reported (Brennecke et al., 2008). To identify piRNAs potentially targeting *bcd*, we blasted all sequences from the aforementioned dataset against the *bcd* mRNA using the NCBI-BLAST sequence alignment algorithm as previously reported (Rouget et al., 2010). This analysis revealed two regions of interest in the *bcd* ORF at positions 1078-1091 and 1616-1629, where piRNAs with at least 13 consecutive nucleotide complementarity occurred at high frequency (168 and 182 respectively) (Fig. 5.1a). Consistent with a putative silencing function in ovaries and subsequent relief after the oocyte-to-embryo transition (OET), when *bcd* gets translated, these piRNA clusters appeared to be downregulated in early embryos (Fig. 5.1a, green). A more careful analysis revealed that these piRNAs mapped to the *roo* or *stalker* transposable elements, respectively. The composition of the cluster originating from *stalker* is presented in Figure 5.1b.

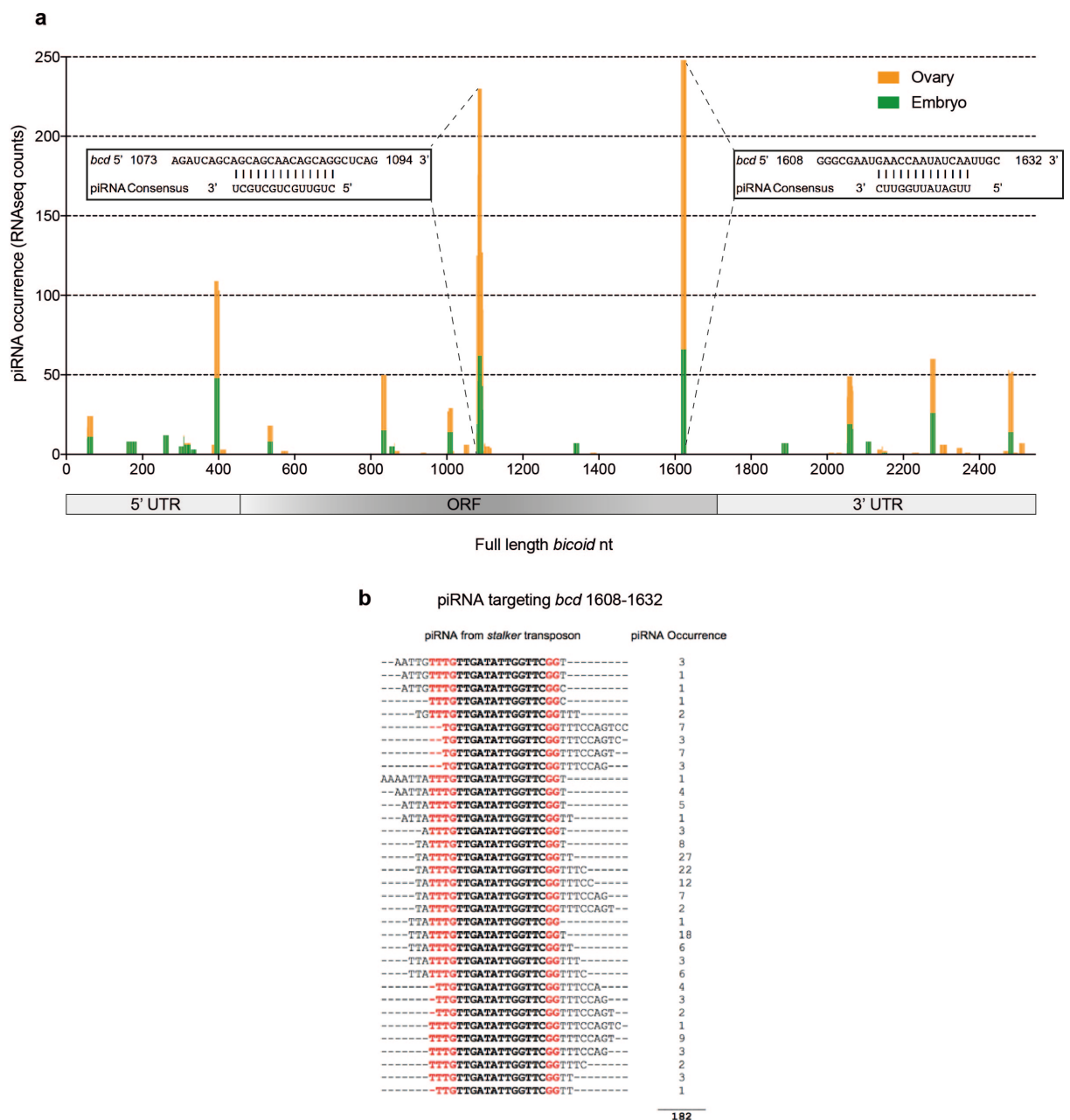


Figure 5.1 | *bicoid* carries putative target sites for piRNAs. **a**, BLAST output showing complementary piRNA occurrence against the entire *bicoid* mRNA length in ovaries (orange) and embryos (green). Regions starting at positions 1078 and 1616 indicate high occurrence of piRNAs (168 and 182) potentially capable of targeting *bicoid*. These piRNA clusters are downregulated in the embryo. (The columns are stacked.) **b**, alignment of piRNA cluster for region at 1616. Sequences represent mature piRNAs and were found to originate from the *stalker* transposon. Nucleotides complementary to *bicoid* are highlighted in bold (black) and number of reads for each species is shown. Analysis based on data published in Brennecke et al., 2008.

5.3.2 Unbiased mapping across the *bcd* 3'UTR using TRACE

Following the proof of concept of TRACE in chapters 3 and 4, we next sought to repurpose this system for the exploratory mapping of *cis* elements in the *bcd* 3' UTR. We reasoned that information on differential miRISC binding or in a wider sense any protein occupancy across two biological conditions would be highly relevant for the unbiased interrogation of *cis* regulatory elements embedded in mRNAs (Fig. 5.2).

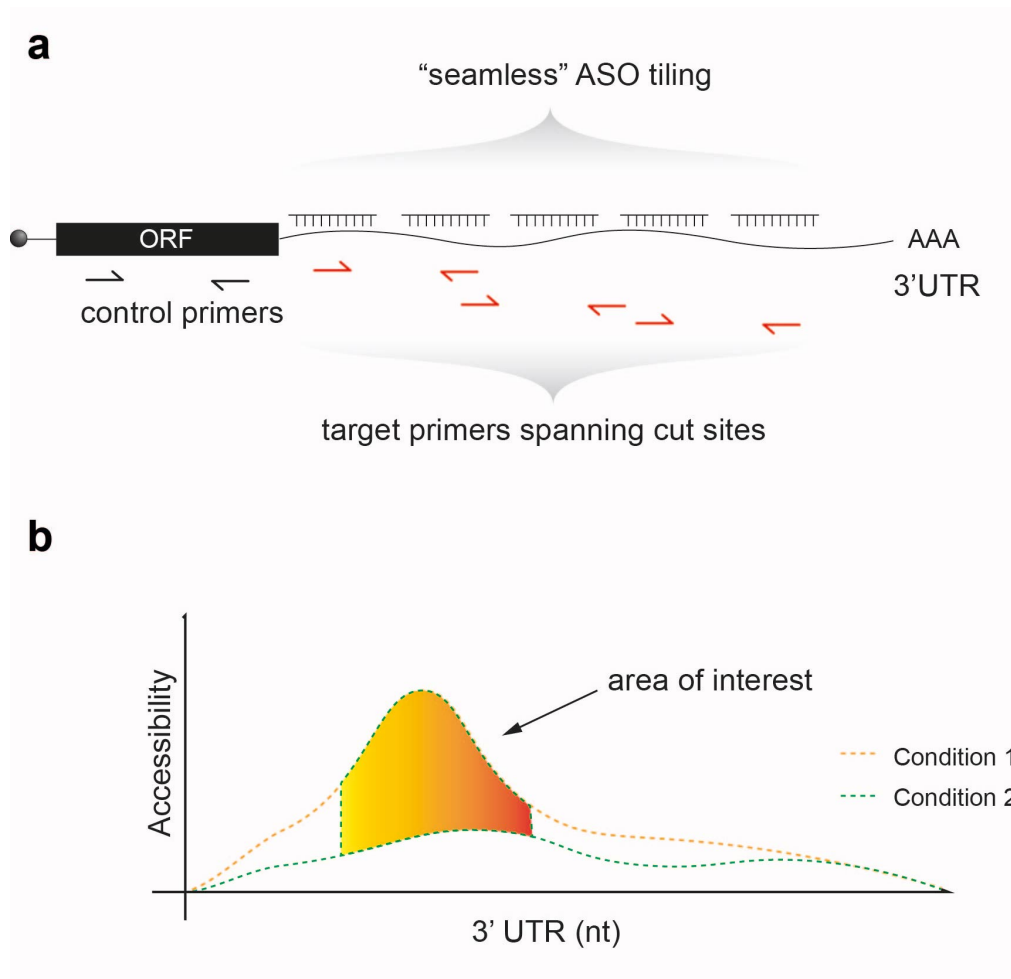


Figure 5.2 | The TRACE assay can be used to explore 3'UTR accessibilities and compare those across different biological conditions or developmental stages. a, Any 3'UTR can be tiled with 25mer ASOs and its accessibility measured with corresponding primers to generate maps over a larger RNA region. **b,** The direct comparison of accessibilities between two different states of the RNA can unbiasedly pinpoint *cis* regulatory elements (e.g. MREs) that are preferentially occupied in one stage but not in another (orange area under the curve).

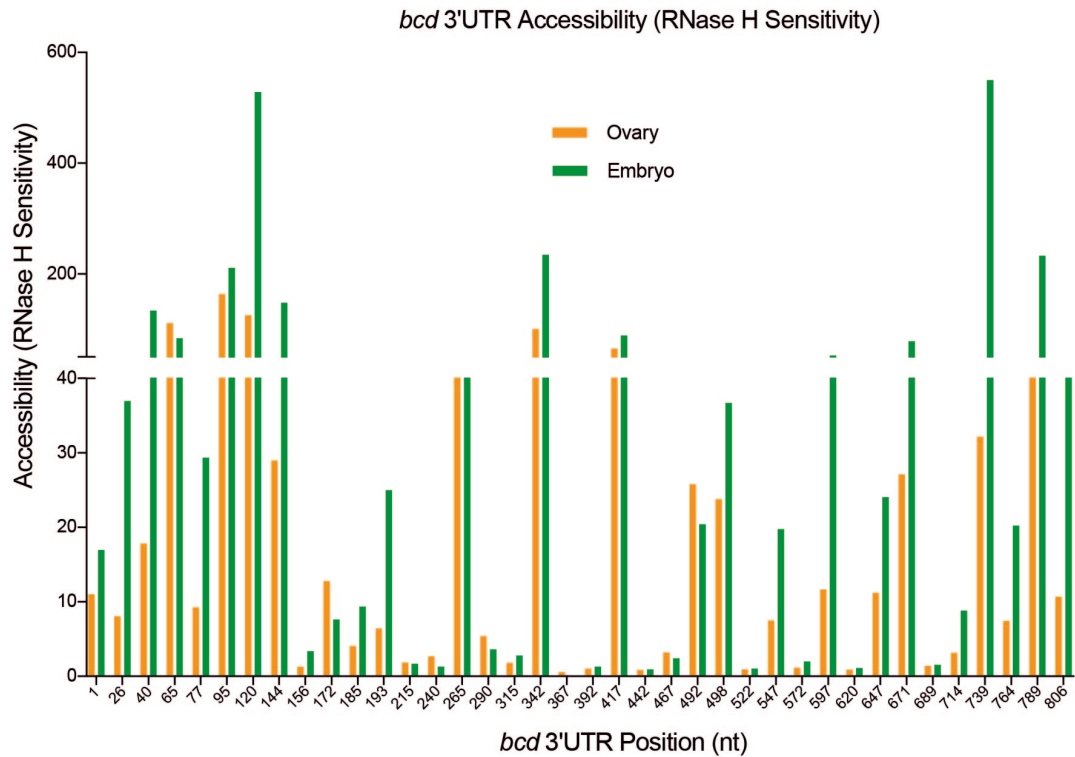
For this analysis specifically, we took advantage of the fact that *bcd* transitions between two dramatically distinct states: ovarian *bcd* RNA is highly localised and translationally repressed, whilst embryonic *bcd* is released from its anterior anchor and competent for translation. We performed TRACE on *bcd* isolated from total ovaries, and directly compared its RNase H sensitivity profile (reflecting accessibility) to its embryonic counterpart. We reasoned that ovarian *bcd* is silenced through the physical association with a repressor factor, which terminates following the OET. Thus, the RNA is expected to undergo a change in its RNase H sensitivity at the site of repressor binding, which should be detected by TRACE.

To this end we seamlessly tiled the 834bp 3'UTR of isoform E (longest of 5 isoforms) with a total of 38 25mer oligonucleotides covering 830bp, providing 99.52% coverage. For each ASO targeting the 3'UTR we calculated an RNase H sensitivity score for ovary and embryo (Fig. 5.3a) according to the following formula.

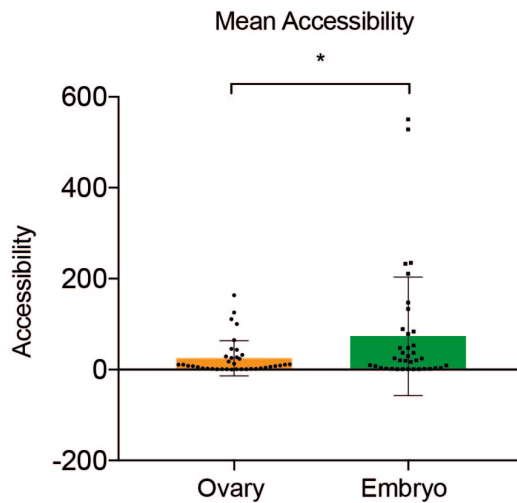
$$\text{RNase H Sensitivity} = \frac{2^{-\left(\frac{\text{Ct}_{\text{ASO}} - \text{Ct}_{\text{no ASO}}}{\text{TARGET PRIMERS}}\right)}}{2^{-\left(\frac{\text{Ct}_{\text{ASO}} - \text{Ct}_{\text{no ASO}}}{\text{CONTROL PRIMERS}}\right)}}$$

We noted that accessibility appeared to be generally higher in the embryo with a mean RNase H sensitivity of 73.41 compared to 24.67 for ovary *bcd* (Fig. 5.3b). The median of *bcd* RNase H accessibility in the embryo was 20.41, in contrast to 8.66 in the ovary (Fig. 5.3c). This robust overall increase is consistent with *bcd*'s transition from a repressed and tightly anchored state at late oogenesis to a translated and released one in early embryos (Weil et al., 2008).

a



b



c

	Ovary	Embryo
Number of values	38	37
Minimum	0.574	0.924
25% Percentile	1.831	2.603
Median	8.658	20.41
75% Percentile	27.58	81.11
Maximum	163.5	550.1
Mean	24.67	73.41
Std. Deviation	38.66	130.3
Std. Error of Mean	6.272	21.42
Lower 95% CI of mean	11.96	29.97
Upper 95% CI of mean	37.38	116.9
Sum	937.5	2716

Figure 5.3 | TRACE map of the *bicoid* 3' UTR. **a**, Individual RNase H accessibilities for each ASO in ovary (orange) and embryo (green). **b**, Embryonic *bicoid* exhibits overall larger mean accessibility. **c**, Column statistics summarizing RNase H accessibilities derived from individual ASOs show that *bicoid* is more accessible in the embryo. A single TRACE experiment was analysed in technical replicates by qPCR. Error bars = standard deviation, $P^* < 0.05$.

5.3.3 RNA accessibility reflects known functional *cis* regulatory elements in the *bcd* 3'UTR

We noted higher accessibilities in the peripheries of the UTR with a drop in accessibility around the central region approximately between nucleotide positions 156 to 714, with only occasional high accessibility spikes therein (Fig. 5.3a, 5.4). Interestingly, the peripheral regions have been reported to form the core secondary structure and harbours *cis* regulatory elements critical for adequate *bcd* anterior localisation and anchoring (Fig. 5.4a) (Brunel and Ehresmann, 2004; Macdonald and Struhl, 1988). We connected the discreet RNase H values (one per ASO) to generate an area under the curve visualising RNase H accessibilities of larger regions (Fig. 5.4b). To extract more global information from this data, we smoothened the curve by calculating new means including a varying number (2-5) of scores of neighbouring ASOs (Fig. 5.4b-d). Interestingly, several regions previously implicated in *bcd* dimerisation and localisation appeared inaccessible. The dimerisation domains with a hairpin around position 280, the key localisation domain (also known as domain V) around position 470 and a secondary localisation domain at position 580 all exhibited low RNase H sensitivities (Fig 5.4a-b).

We noticed an insular increase in *bcd* accessibility in both tissues around nucleotide 350 (Fig. 5.4b). Interestingly, *bcd* has been proposed to form a flexible, single stranded hinge at this very region, explaining the relatively isolated increase in accessibility at this stretch (Brunel and Ehresmann, 2004).

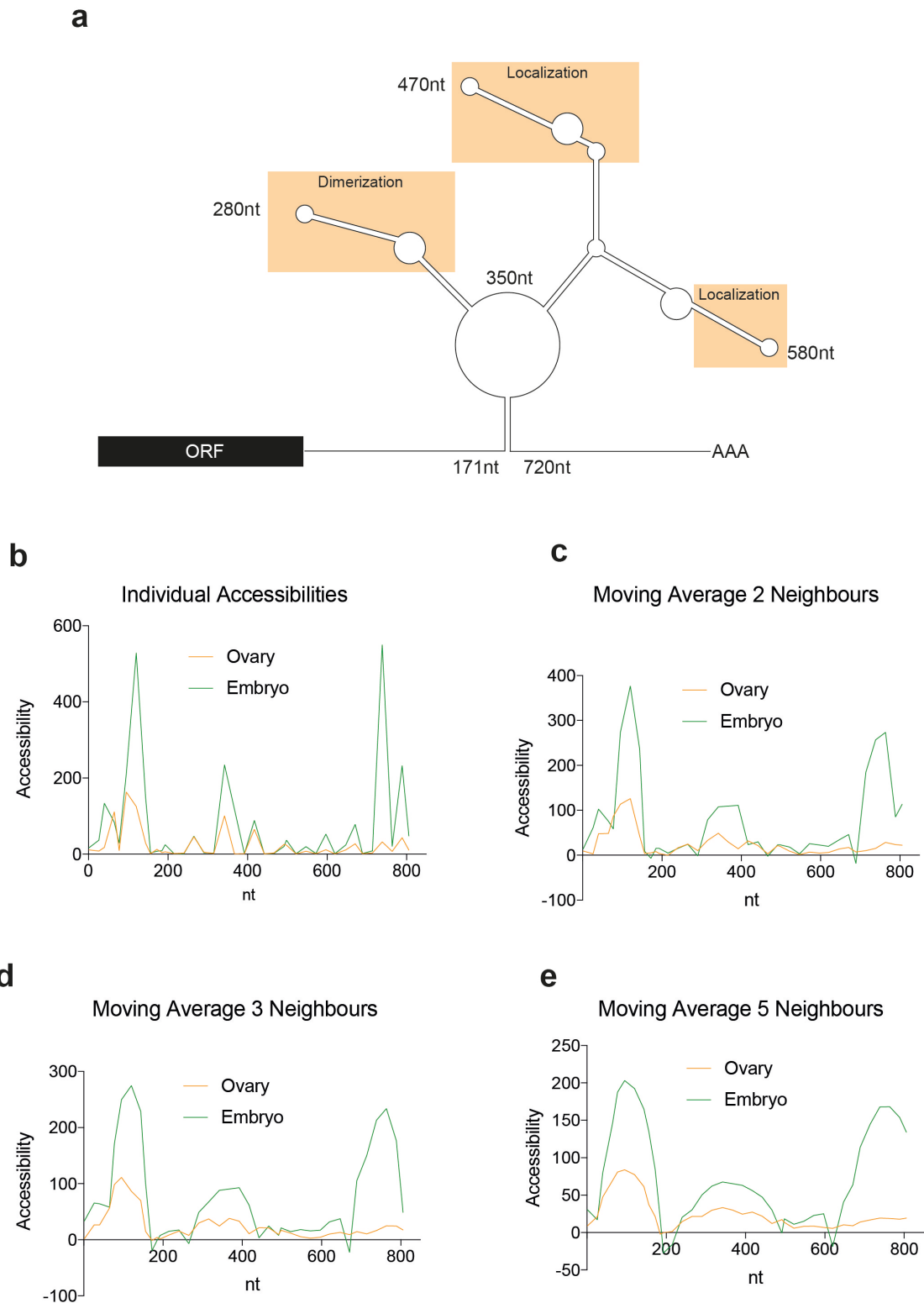


Figure 5.4 | The RNase H sensitivity of the *bicoid* 3'UTR reflects functional *cis* regulatory elements. **a**, The proposed secondary structure of the *bicoid* 3'UTR with well studied localisation domains. **b**, Connecting individual ASO RNase H sensitivities generates an accessibility area under the curve. **c-e**, Smoothing the curves by calculating moving averages with varying number of neighbours reduces possible noise.

While these global observations were consistent with the reported *bcd* 3' UTR architecture, we desired to further confirm the biological relevance of our technique by choosing to examine more closely two previously defined *cis* elements required for *bcd* localisation. If our technique was sensitive enough, we would expect to find particularly low accessibilities across these cardinal *cis* elements as they are likely bound by RNA binding proteins. However, as the RNA is less tightly bound after egg activation and fertilisation we would anticipate an increase in the accessibility in embryos at some of these regions. The Bicoid Localisation Element 1 (BLE1) defined by Macdonald *et al.* and shown to be necessary for anterior localisation has been pinpointed to position 452 to 505 (Macdonald *et al.*, 1993). We observed low accessibilities for both ovarian and embryonic *bcd* at this region with RNase H sensitivities of 10.0 and 7.9 (3 spanning ASOs averaged), respectively (Fig. 5.3). Both values are far below the mean RNase H sensitivities of 24.7 and 73.4 for ovary and embryo, respectively. Furthermore, another *cis* acting nonamer, defined by Gottlieb *et al.*, resides at position 52 to 60 (Gottlieb, 1992). The ovary RNase H sensitivity measured with an ASO spanning this site is 17.9, which sharply rose to 134.0 in the embryo (Fig. 5.3, ASO starting at nt 40). Taking these observations together, we were confident that TRACE reliably reflects protein occupancy on the *bcd* 3'UTR and returned our focus to predicted MREs residing in the *bcd* 3'UTR, possibly contributing to its translational profile.

5.3.4 The *bcd* 3'UTR harbours several differentially bound MREs across the OET

Nanostring profiling coupled with *in silico* prediction had previously revealed multiple MREs in the *bcd* 3'UTR. We next examined closely the accessibility of predicted MREs in *bcd* derived from ovaries and embryos (Fig. 5.5).

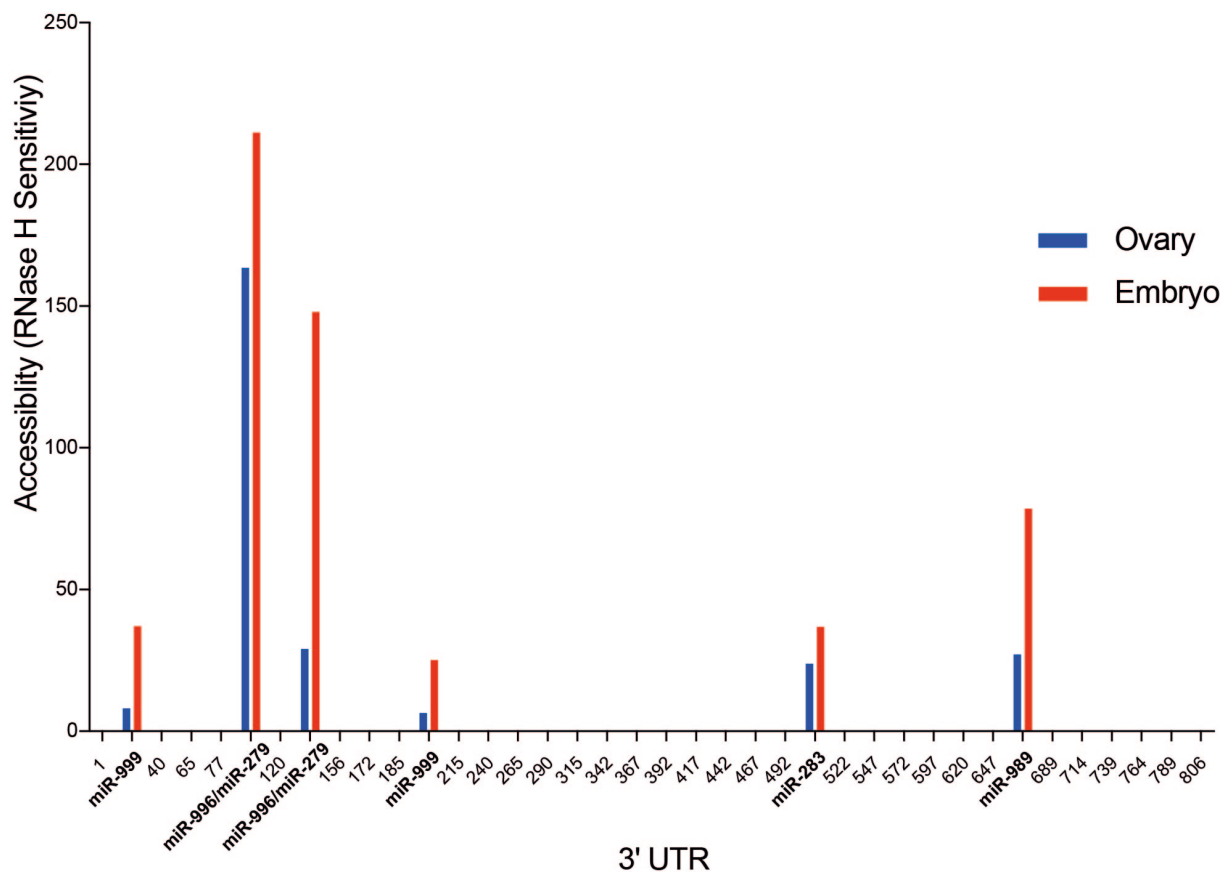


Figure 5.5 | TRACE reveals predicted MREs display increase in accessibility across OET. All MREs corresponding to the miRanda predicted and Nanostring expressed miRNAs exhibited an increase in accessibility. X-axis numbers represent nucleotide position in *bcd* 3'UTR. A single TRACE experiment was performed and analysed in technical triplicates by qPCR.

All MREs revealed an increase in their accessibility (Fig. 5.5). Particularly notable were the sharp rises in accessibility observed for both miR-999 MREs (4.5- and 3.9-fold), the second miR-996/miR-279 MRE (5.1-fold) and the miR-989 MRE (2.9-fold). Importantly, these changes exceeded the 'background' increase observed for the

majority of ASOs (oocyte median = 8, embryo median = 20; 2.5-fold) (Fig. 5.3b,c). This further underscored the possibility that miRNAs bind *bcd* in the oocyte. Nevertheless, orthogonal validation of these putative interactions is required. To this end, we sought to develop a complementary unbiased physical binding assay that discovers any small RNA *trans* factors associated to *bcd* mRNA.

5.3.5 Design and implementation of *bcd* RNA pulldown strategy

TRACE informs on the relative accessibilities of RNA regions, however, unless performed in a candidate-paired manner (e.g. miR-184 WT vs. KO), complementary approaches are required to confirm the identity of the corresponding *trans* factor. Hence, we next resorted to an antipode approach, in which we isolated *trans* interacting factors by means of a target mRNA pulldown. Defining physiological RNA and protein interactomes by high-throughput sequencing and quantitative proteomics respectively relies on highly efficient and specific capturing of the target RNA from intact cells. The workflow we developed for *bcd* interactome capture is described schematically in Figure 5.6.

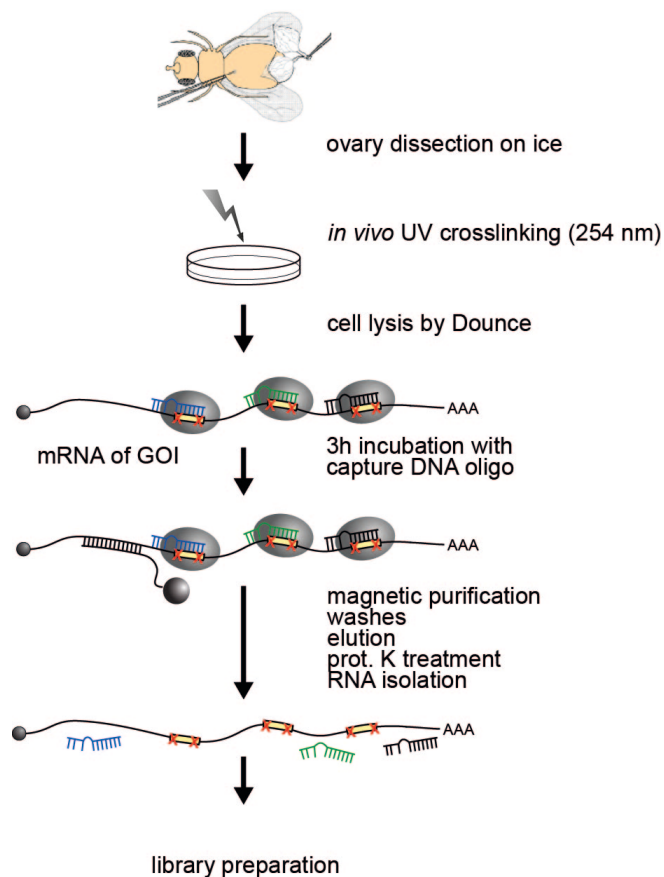


Figure 5.6 | Schematic *bcd* pulldown overview. Ovaries are dissected on ice from yeast fattened female flies. They are then irradiated at 254nm while still intact and subsequently lysed by Dounce homogenisation. The lysate is incubated for 3h with *bcd* antisense ASO capture beads, which are then magnetically purified and washed. Elution occurs at 95°C, followed by proteinase K treatment and RNA isolation. *Bcd* enrichment is measured by qPCR and the RNA undergoes library preparation.

Briefly, ovaries are dissected, dissociated into single egg chambers and UV crosslinked at 254nm. Cells are then lysed under high salt, high detergent conditions and lysates are applied to probe-coupled beads for 3h at 40°C. Hybridised beads are stringently washed in a series of buffers (high salt and detergent) to eliminate all non-crosslinked material. *Bcd* and all associated binding molecules are then eluted in a salt-free buffer. Analysis of the bound RNA interactome requires a crosslink reversal step (extensive Proteinase K treatment) to remove the protein fraction.

5.3.6 Generation of antisense probes

One of the key features influencing the efficiency of RNA pull-down assays is the affinity and specificity of the capture probes. Before the TRACE data was generated, we had previously evaluated an independent set of nine antisense DNA oligonucleotides (ASO) as candidate capture probes. These were designed to span

the 3'UTR of *bcd*, avoiding predicted miRNA binding sites and known protein recognition motifs (Fig. 5.7a). To identify and rule out potential off-target effects, all sequences were also aligned using BLAST against the *Drosophila* genome. The selected oligonucleotides were ~25 nt long with a T_m of approximately 60°C. Each probe was hybridised to formaldehyde crosslinked *bcd* mRNA in a separate cell lysate reaction (Fig. 5.7a). Samples were then incubated with RNase H followed by an extensive crosslink reversal step, and RNA extraction. cDNA was synthesised and assessed by qRT-PCR using primer pairs spanning the putative regions of hybridisation and normalised to a control primer pair outside the region of interest. The ΔC_t value between the target and control primer pairs, allows the comparison of accessibility of various regions across the target. A large ΔC_t corresponds to a highly accessible site. RNase H accessibility can then be calculated using the aforementioned formula (normalising for primer efficiency and RNA input).

This analysis revealed that the target sites covered by oligonucleotides 5 and 8 are considerably more accessible than 1, 2, 6 and 9, while 3, 4 and 7 are effectively inaccessible (Fig. 5.7b). Consistent with these results, alignment of oligonucleotides 5 and 8 to the predicted secondary structure of *bcd* 3'UTR, mapped them outside of functional domains known to interact with proteins. Providing further confidence in the accessibility of target sites 5 and 8, these probes appear to overlap with the flexible and predominantly single-stranded central hinge region (Fig. 5.7c) (Brunel and Ehresmann, 2004).

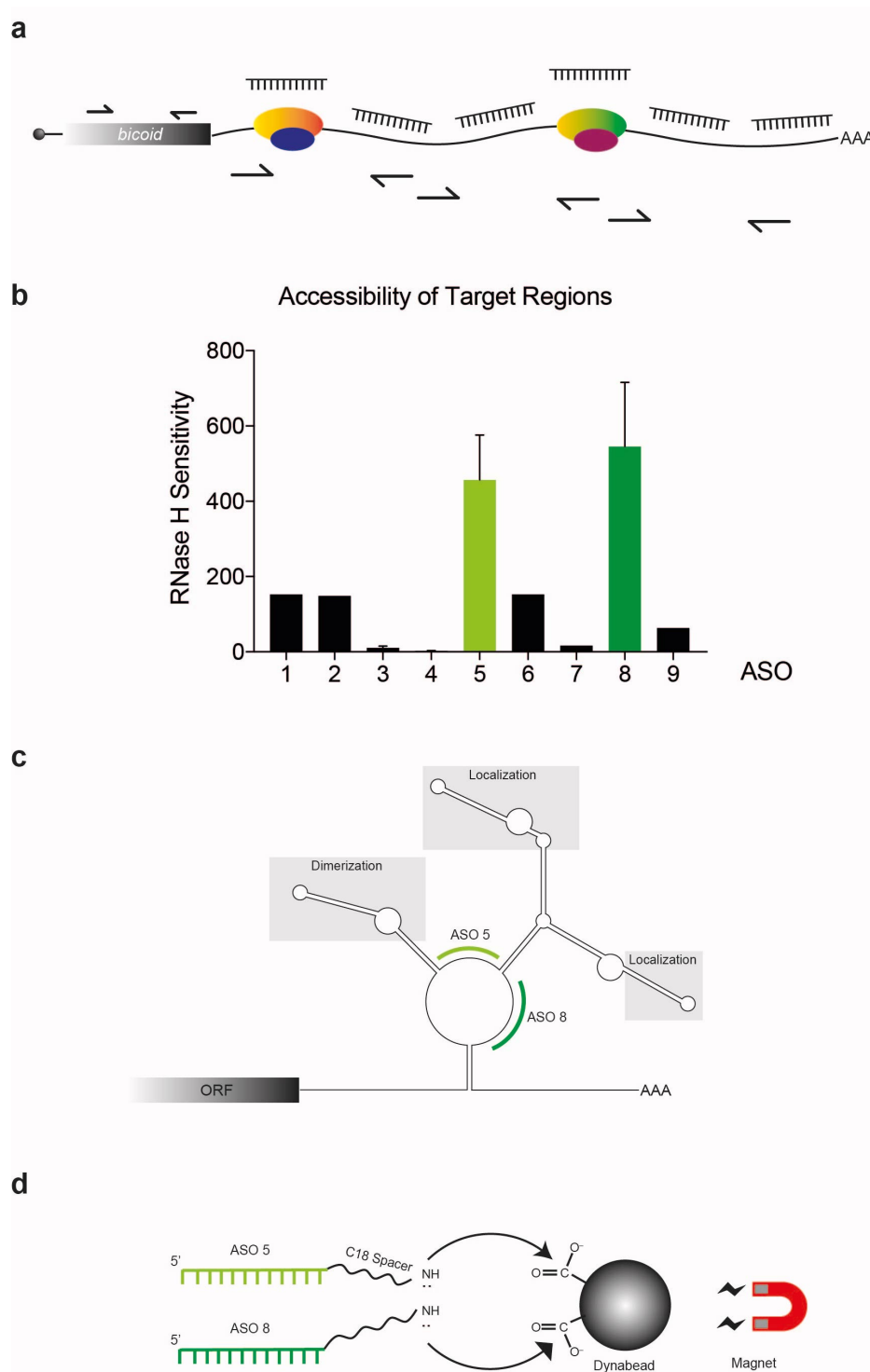


Figure 5.7 | Generation of antisense probes. **a**, a total of 9 ~25mer, 60°C T_m antisense oligonucleotides (ASO) were individually assessed for the hybridisation capability against crosslinked *bcd* (see TRACE chapter). **b**, RNase H sensitivity measurement revealed ASO 5 and 8 are particularly sensitive and **c**, also overlap with previously identified single stranded regions. **d**, ASO 5 and 8 are fitted with a C18 spacer and a 3' amino group which can be activated to form a covalent peptide bond to carboxylic acid coated dynabeads. The beads are magnetic and can thence be retrieved following lysate incubation and *bcd* capture.

Based on these results, oligonucleotides 5 and 8 were selected as pull-down probes. Probes were synthesised as previously described, using the following configuration: {5'-Oligonucleotide Sequence-3'} – C18 Spacer – Amino Modifier (Simon, 2013). To increase RNA pull-down efficiency and specificity, probes 5 and 8 were then pooled and chemically coupled to the same EDC-activated Dynabeads® My One Carboxylic Acid (Fig. 5.7d). Previous studies have shown that covalently coupled oligo-d(T) magnetic Dynabeads are suitable for total RNA-protein interactome capture (Castello et al., 2013b). This approach offers superior specificity over biotin-streptavidin affinity based technologies, although the latter have also been employed successfully in various studies (Engreitz et al., 2013; Hassan et al., 2013; Simon et al., 2011).

5.3.7 *Bcd* RNA pulldown optimisation

To assess the specificity and efficiency of the experimental strategy described above, a pilot assay was performed. The quantity and purity of *bcd* mRNA was determined by qRT-PCR relative to the ribosomal housekeeping gene *rp49*. Several iterations of the pulldown strategy led to progressive improvements in the fold-enrichment of *bcd* RNA (Fig. 5.8). The 1st iteration of this protocol followed the RNase H elution step conditions previously published as part of the CHART method (Simon, 2013) and resulted in a 97-fold enrichment of *bcd* relative to *rp49* from input to eluate (Fig. 5.8). Replacing the elution step with a heat elution at 95°C produced a moderately improved enrichment (118-fold). In the 3rd iteration of the protocol, the pulldown efficiency was improved by almost 30x to 2939-fold, by simple pre-absorption of the capture beads with an *in vitro* transcribed RNA (*Xenopus eef1a*). This suggested that naive beads are 'sticky' and benefit from a blocking step leaving only antisense oligonucleotides available for target capture. The 4th iteration was performed using a

stronger crosslinking dosage (1J.cm^{-2}) and lysis at increased DTT concentration (12.5mM) (Sysoev et al., 2016).

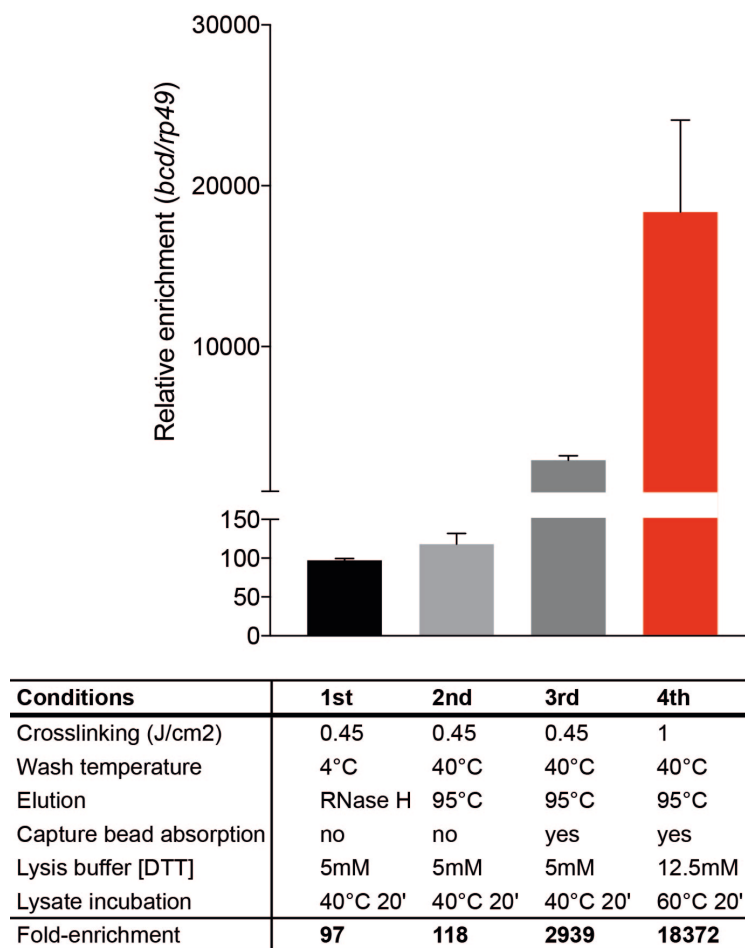


Figure 5.8 | Development of pulldown protocol. Several iterations of the pulldown protocol led to improved *bcd* enrichment results. The introduction of capture bead pre-absorption with an IVT RNA resulted in a major improvement in capture efficiency (3rd iteration). The 4th iteration included a separate lysate pre-incubation with an increased DTT concentration that engendered another upgrade of the pulldown efficiency.

Critically, this iteration incorporated a lysate pre-treatment step at an elevated temperature of 60°C for 20 min, which aimed to remove all non-crosslinked factors and clear debris (Sysoev et al., 2016). This configuration led to a considerable enrichment of *bcd* in the eluate of 18,372-fold and was chosen as the optimal protocol for the small RNA interactome capture.

5.3.8 Confirmation of *bcd* capture efficiency by ddPCR

Digital droplet PCR (ddPCR) is superior over analogue real-time PCR since it allows absolute digital quantification of transcript molecules and does not rely on $\Delta\Delta C_t$ normalisation methods (Hindson et al., 2013). In pull down assays, the eluate is deliberately depleted of reference genes, thus making quantification of target RNA by standard qPCR difficult to perform. Instead, in ddPCR the target RNA levels can be directly quantified in copies/ μ l from a reverse transcription reaction where RNA input levels are normalised across all fractions. Pull down efficiency based on the 3rd protocol iteration had previously been determined by measuring the enrichment of the *bcd* fraction in the eluate *versus* the input cell lysate (Fig. 5.9a). We observed that the input fractions contained concentrations (in copies/ μ l) of 25, 943, 186, 729 for *bcd*, *rp49*, *gapdh*, and *act5C*, respectively (Fig. 5.9a,b, *rp49* and *act5C* not shown). Expressing these values as percentages of the sum of all four genes suggested that *bcd* accounts for ~1%, whilst *rp49*, *gapdh*, and, *act5C* correspond to 50%, 10%, and 39%, respectively (Fig. 5.9c). The same analysis of the eluate fractions revealed concentrations (in copies/ μ l) of 423, 7.49, 0.625, and 2.46 for *bcd*, *rp49*, *gapdh*, and *act5C*, respectively (Fig. 5.9a,b, *rp49* and *act5C* not shown). This corresponds to a 16.92-fold enrichment of *bcd* concomitant with a massive depletion of the other mRNAs (125.9, 297.6, and 296.3-fold). Thus, in the eluate fraction, *bcd* accounts for 97% of the four mRNAs, reflecting a robust increase from ~1% in the input fraction (Fig. 5.9c). Although only the 3rd iteration of the pulldown was assessed by ddPCR, we were confident that between this data and the analog qPCR data obtained for the 4th iteration, our final pulldown protocol engendered optimal *bcd* capture efficiency.

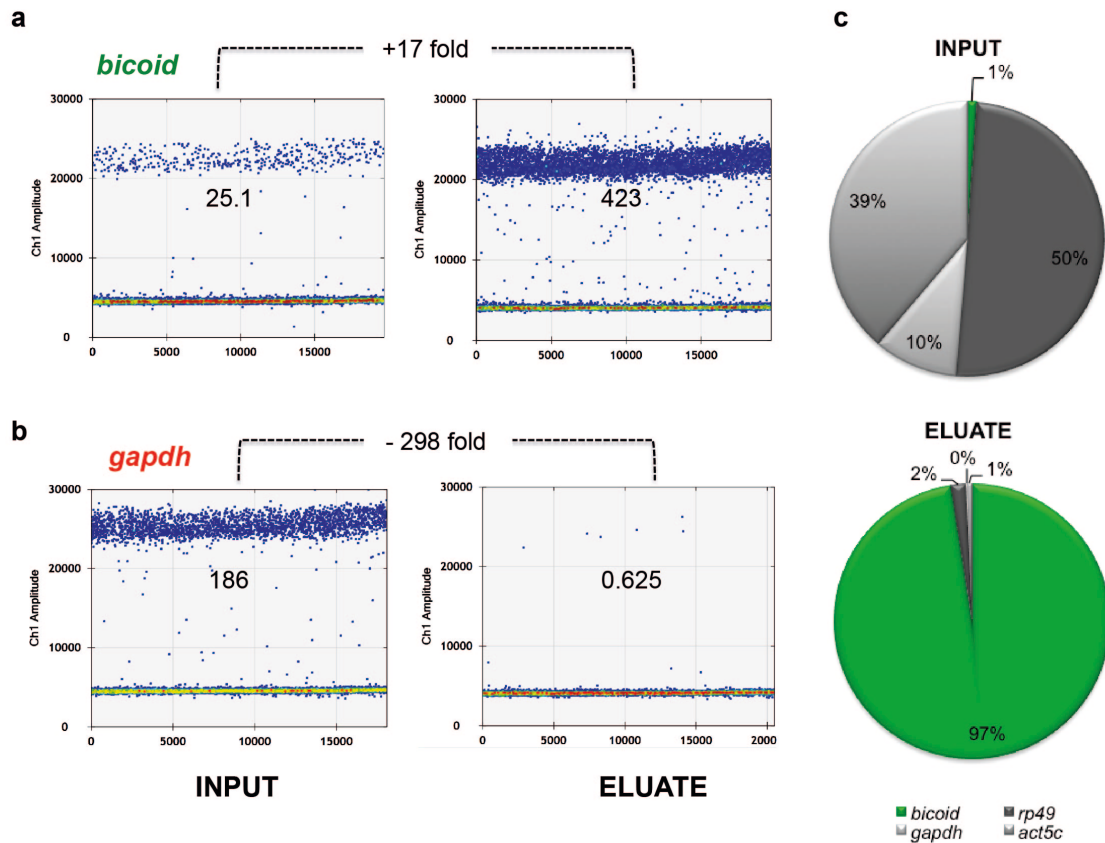


Figure 5.9 | *Bicoid* can be specifically captured from ovaries. **a**, Counts of *bcd* in input fraction vs eluate. The concentration in copies/ μ l increases from 25.1 to 423, with 900 copies/ μ l being retained on the beads. **b**, Counts of *gapdh* in input fraction vs eluate. The concentration is depleted from 186 copies/ μ l to 0.625 copies/ μ l. **c**, Pie charts representing the percentage of *bicoid* present in input vs eluate as approximated by relative amount to 3 highly abundant housekeeping genes (*rp49*, *gapdh*, *act5C*). *Bicoid* levels increase from 1% in the input to 97% in the eluate. Technical replicates were performed by ddPCR. In **a** and **b** the x-axis shows number of droplets, the y-axis signal amplitude (i.e. +/- target molecules in each droplet).

5.3.9 Implementation of a small RNA sequencing protocol

Following the establishment of high efficiency *bcd* capture, small RNAs from the eluate fraction were subjected to library preparation using the NEBNext Small RNA Sequencing kit. The recommended library input amount is 100ng-1ug total RNA. Since our goal was to isolate a single RNA species plus its interactome it was not surprising that this guideline RNA concentration was unattainable. Using the Qubit platform, we estimated our eluate RNA concentration to less than 20ng/ml. To maximise the chances of detecting putative *bcd* interacting small RNAs (such as miRNAs), we loaded the maximum RNA volume (6µl) into the library preparation. NGS library preparation includes a PCR step with a recommended amplification of 12-15 cycles. Given the low input amount we initially tried this step with 15 cycles. To assess library fragment size, we analysed the library on a Bioanalyzer DNA High Sensitivity Chip (Fig. 5.10a). Adaptor-containing miRNAs and piRNAs are expected at 143bp and 153bp, respectively.

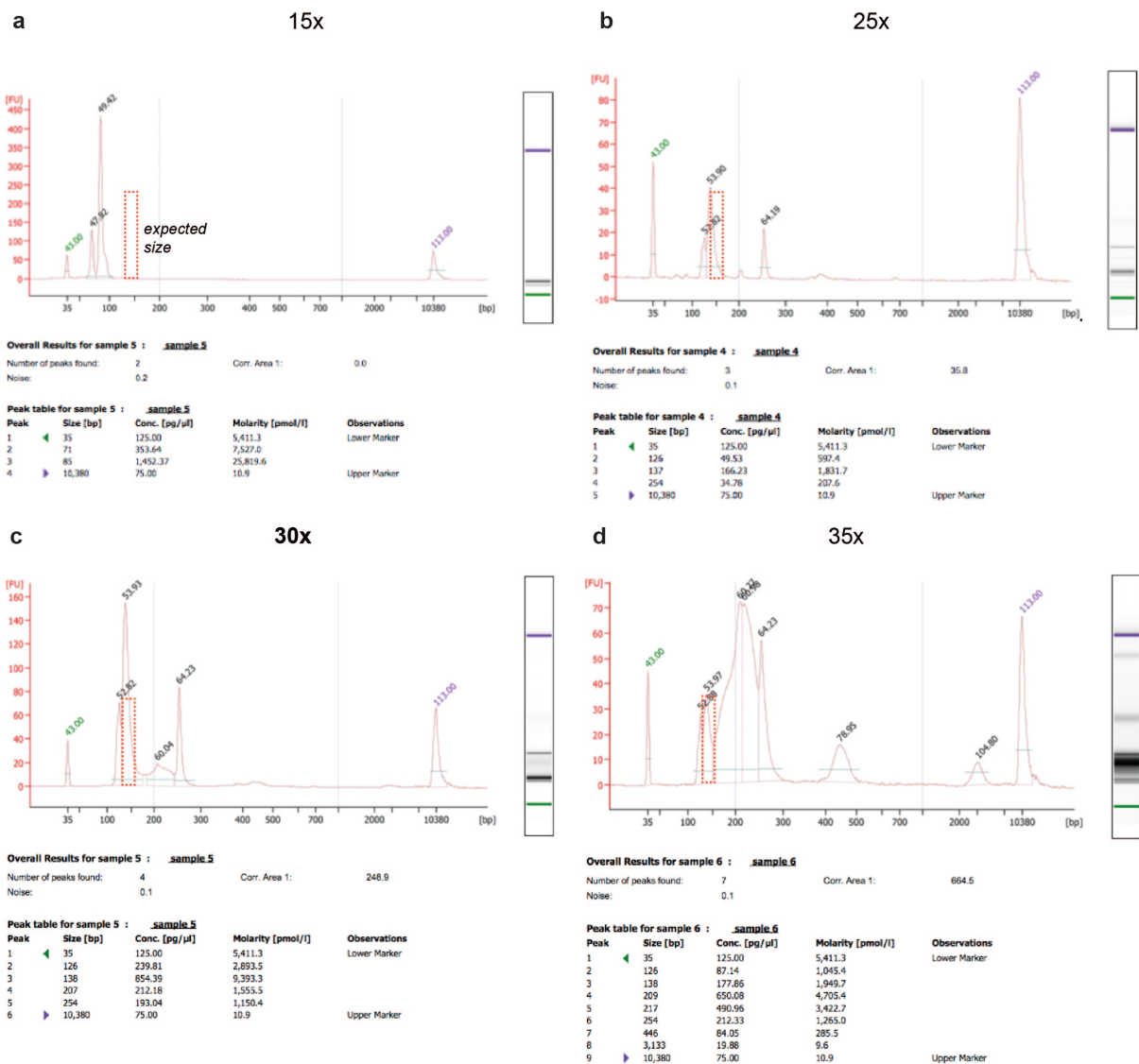


Figure 5.10 | Bioanalyzer traces of small RNA library preparation. **a**, the recommended number of 15 PCR cycles resulted in no discernable peak at the expected location of small libraries (143-153bp). Primer dimers are clearly visible at 85bp. **b**, increasing the cycle number to 25 cycles produced a peak at 137bp with a shoulder extending into the expected size window (~140-150bp). **c**, increasing the cycle number to 30 specifically increased this peak (with a concomitant increase in adaptor dimer). **d**, expanding the PCR to 35 cycles resulted in a trace where no clear library peaks could be singled out.

However, upon 15x amplification, no peak could be detected at the relevant size, although primer dimers were clearly visible at 85bp (Fig. 5.10a). This prompted us to increase the number of amplification cycles to 25x, 30x and 35x. At 25x we started to

detect adaptor dimers (127bp) but also a peak at the approximately correct size of 137bp (Fig. 5.10b). Since this appeared with a shoulder towards larger sizes we reasoned that our desired small RNA library might have emerged as part of this peak. We tested a 30-cycle amplification protocol and observed a marked increase at this putative small RNA peak (now called at 138bp) (Fig. 5.10c). However, adaptor dimers were also further amplified. Finally, 35-cycles amplification resulted in a distortion of the library with non-specific, wide areas emerging around 200bp (Fig 5.10d). Consequently, we settled on 30 cycles as optimal PCR library amplification, which offered a strong signal immediately adjacent to the peak expected to contain small RNAs.

Notably, this library still carried contaminants such as adaptor dimers, which could reduce the quality of sequencing runs. Several options exist to size select RNA libraries, including bead purification and polyacrylamide gel size selection. However, the Bioanalyzer trace revealed high levels of adaptor contamination (127bp) in immediate vicinity of our expected small RNA peak (143-153bp). Based on this observation, we decided to use a highly accurate targeted size selection system on the Pippin Prep Station (Sage Bioscience). This approach enables isolation of narrow size intervals that are otherwise impossible to resolve using conventional bead or polyacrylamide gel selection procedures. Size selecting for a 135-160bp range on the Pippin device produced a single peak at the precise window where adaptor ligated miRNAs/piRNAs are expected (Fig. 5.11b). Unfortunately, this procedure led to substantial sample loss as reflected by the fluorescence measured on the Y-axis decreasing from ~100 to 5 FU (Fig. 5.11a,b).

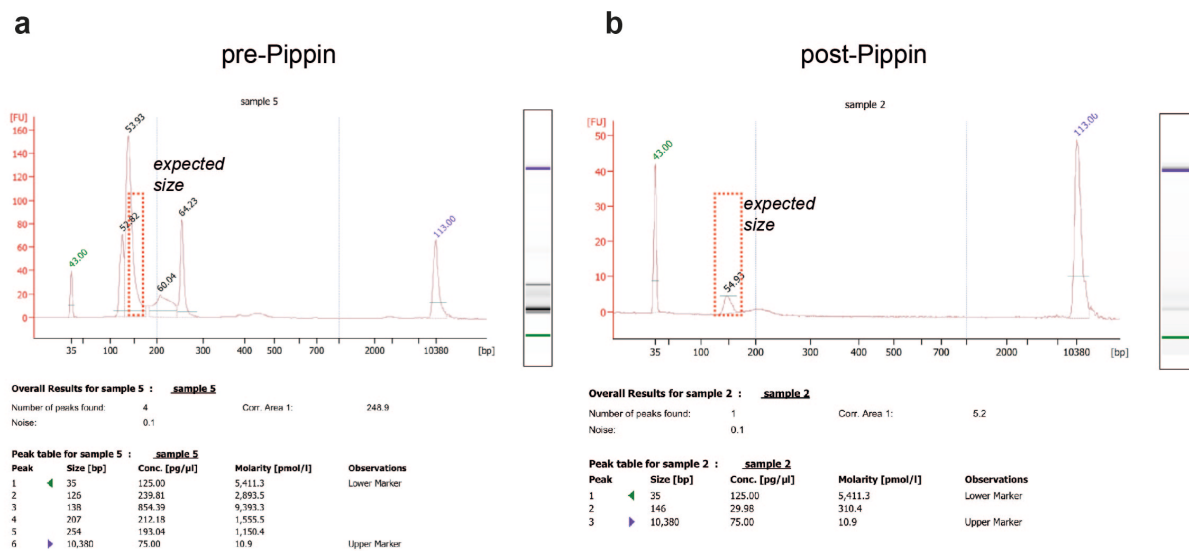


Figure 5.11 | Pippin size selection cleans up library trace. **a**, Bioanalyzer trace of the small RNA library amplified with 30 cycles revealed a shoulder adjacent to the 137bp peak that is likely to contain small RNAs (143-153bp). **b**, following Pippin size selection using a 135-160bp range a small peak at the correct size is visible.

5.3.10 Interactome capture isolates abundant miRNAs associated to *bcd*

The small RNA library generated by the *bcd* interactome capture approach was sequenced on a NextSeq benchtop sequencer. In total, 354,401 reads mapped to *Drosophila* genome. Figure 5.12 represents the top 51 gene IDs to which reads have been mapped. The top 10 ribosomal RNA species have been excluded. Although the NEBNext Small RNA Sequencing kit is not designed to produce long RNA libraries, fortuitous fragmentation or degradation of longer transcripts most likely occurred during sample processing which led to the abundant detection of long RNAs (including *bcd*) in this sequencing run (Fig. 5.12a). This may also explain why the miRNAs only represent ~1.97% of all mapped reads (Fig. 5.12b). Within the top candidate genes, ~50% of the reads mapped to protein coding genes (Fig. 5.12b) and only four hits were represented by miRNAs (Fig. 5.12a). Interestingly, despite considerable enrichment in the qPCR experiments, *bcd* itself only ranked at position

14 in the most abundant RNA species present in the pulldown (post rRNA exclusion). With regard to miRNAs, miR-989, which is predicted to target *bcd* at two locations, appears to be very abundant in the pulldown fraction. Intriguingly, miR-184, which is abundantly expressed in the ovary but does not appear to contain canonical MREs within the *bcd* sequence, had the highest representation amongst miRNAs (4510 counts). Let-7 and miR-311 were also found amongst the pulldown targets, with 661 and 631 counts, respectively (Fig. 5.12a).

We next examined all the miRNAs present in the pulldown (Fig. 5.13a) for their potential to target *bcd* (Fig. 5.13b). A previous run of the miRanda algorithm had already predicted several of these as capable of targeting *bcd*. To assess the potential the remaining miRNAs to regulate *bcd*, we loosened the miRanda algorithm parameters and subjected all miRNAs to another round of alignment against the full length *bcd* sequence. Briefly, the miRanda score threshold was lowered from 140 to 120 which gave rise to multiple new predictions. We filtered the predicted targeting events to select for those that had at least 6mer complementarity at the seed motif and were detected by Nanostring in the ovary fraction. This revised prediction revealed the presence of one target site for the most abundant miRNA found in the pulldown, miR-184. Furthermore, another MRE for miR-989 was identified, in addition to the two previously reported sites. Also, the miRNA let-7 was found to have one MRE in the *bcd* transcript. Finally, both miR-310 and miR-311, which belong to the same family of miRNAs, have two predicted sites. Importantly, all miRNAs identified through the pulldown have target sites within *bcd*, underscoring the reliability of this approach.

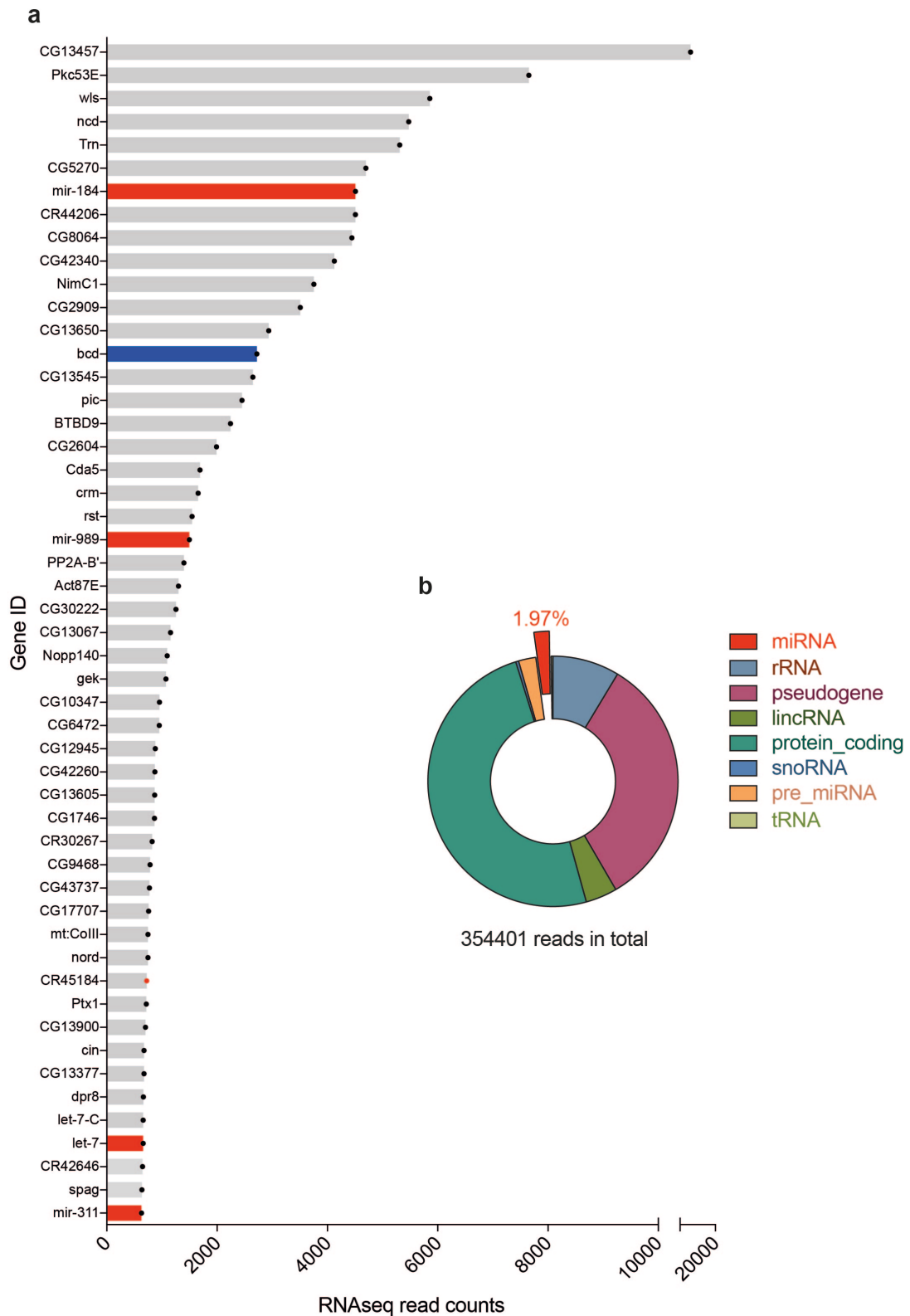


Figure 5.12 | Top candidates of the *bcd* pulldown. **a**, top 51 genes identified in the pulldown with their respective read counts. **b**, a pie chart represents the distribution of gene biotypes. Half of the reads mapped to protein coding genes, including *bcd*. The second most abundant group are pseudogenes. miRNAs (exploded, red) only represent 1.97% of the reads.

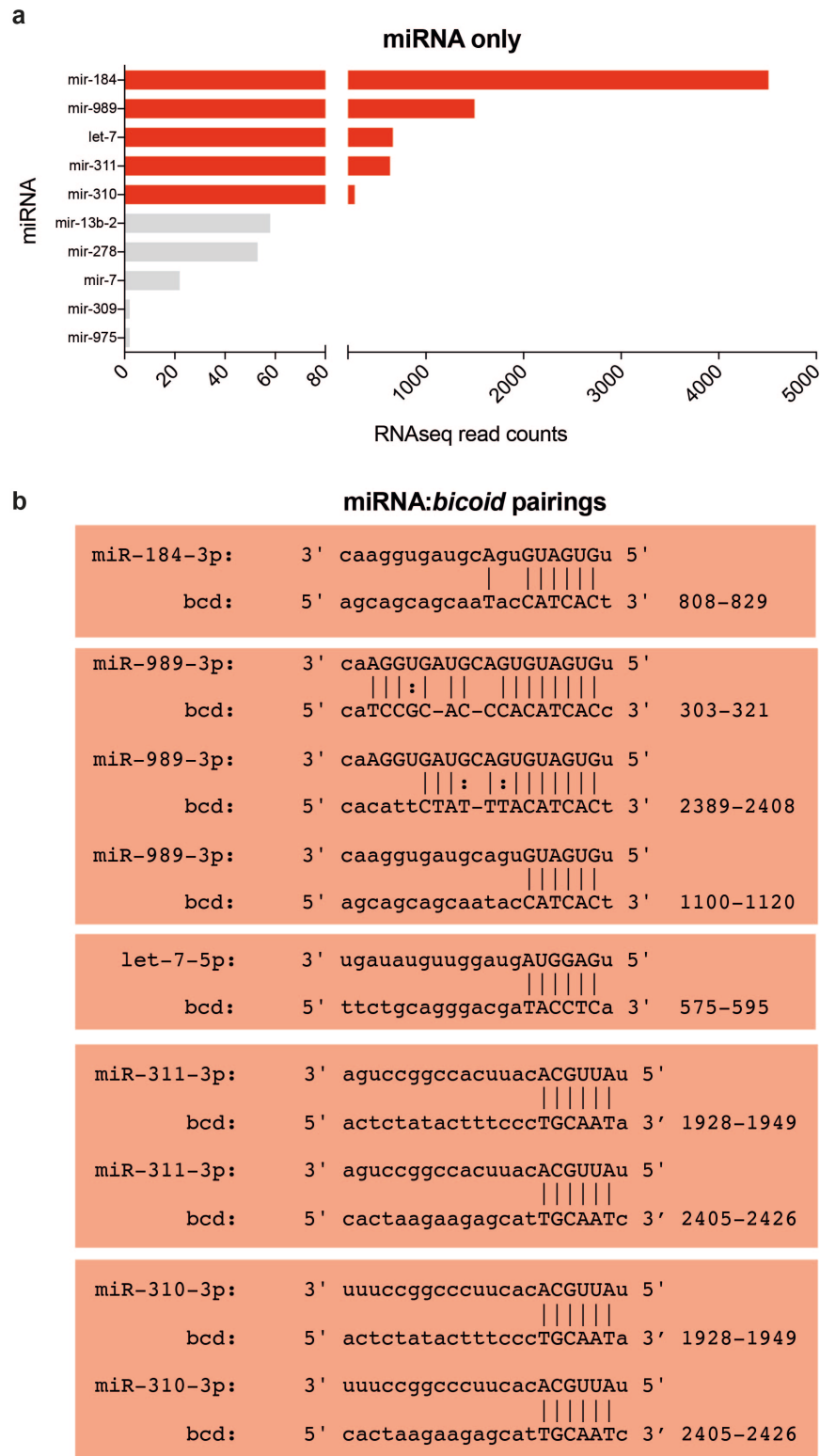


Figure 5.13 | *Bcd* contains MREs for all top miRNA pulldown candidates. **a**, 10 miRNA species were detected by sequencing, with miR-184 being the most abundant. **b**, re-executing the miRanda algorithm with more permissive (6mer seed allowed) criteria revealed that the top 5 candidate miRNAs have target sites in *bcd*. miR-184 has one predicted target site, miR-989 has three target sites, including an 8mer. let-7 has one target site. The miR-310/311 family has two overlapping target sites each.

5.3.11 miRNA-*bcd* interactions are decreased in early embryos

The pulldown and small RNAseq was performed only on a single ovary replicate due to technical difficulties with library preparation (see discussion). To validate the expression of candidate miRNAs isolated from the ovary, we performed Taqman qRT-PCR on the same eluate RNA fraction. Furthermore, as *bcd* is relieved from repression and actively translated in embryos, we reasoned that any associated repressor RNA would be absent from the embryo interactome. Hence, we sought to compare the levels of candidate miRNAs isolated from the ovary pulldown to their embryonic counterpart. To this end, we performed a second *bcd* RNA pulldown, employing identical conditions, from 30-90min old embryos (prior to *bcd* clearance as part of maternal-to-zygotic transition).

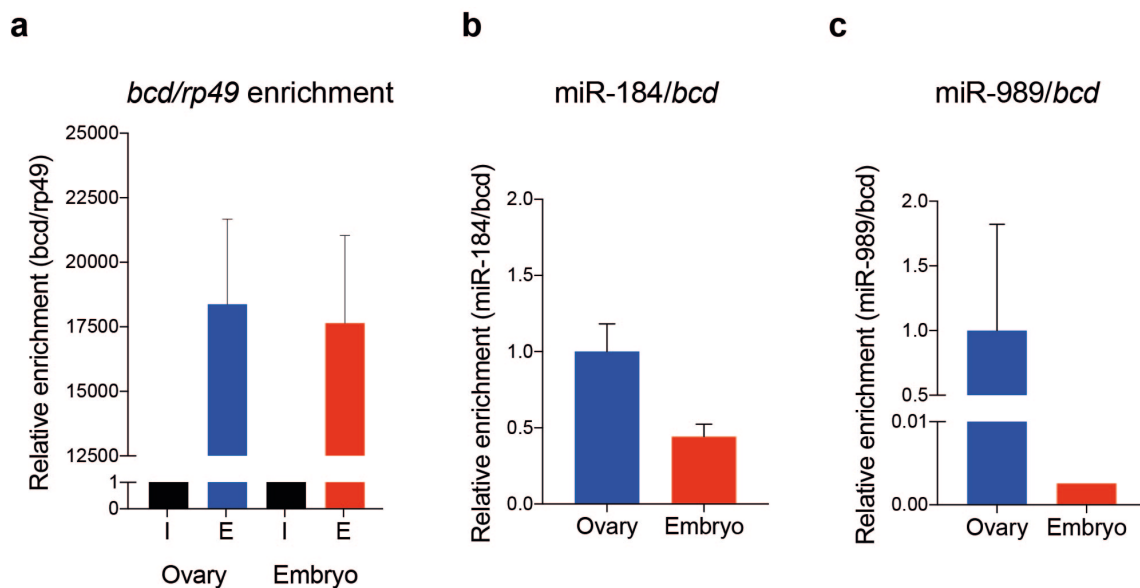


Figure 5.14 | qRT-PCR validation of pulldown. **a**, subjecting embryos to the same pulldown procedure as ovaries engendered comparably specific *bcd* capture. **b**, Taqman qPCR validation of miR-184 revealed that it is less associated with *bcd* in embryos than in ovaries. **c**, miR-989 is heavily depleted from embryonic *bcd*. Technical qPCR triplicates were performed for a). For b) and c) two independent pulldowns were performed. Error bars = standard deviation. N.b. embryo column in c) lacks error bar as miR-989 could only be detected in one of the two replicates.

Subjecting embryos to the exact same pulldown procedure as ovaries, resulted in a similar *bcd* enrichment in the eluate (Fig. 5.14a). We next measured the levels of the two most abundant candidate miRNAs (miR-184 and miR-989) in ovary and embryo eluates by Taqman qPCR. We were curious to see how much of each miRNA is associated to *bcd* and whether this changes during the transition from ovary to embryo. When miR-184 levels were normalised to *bcd* we found 56% less miRNA associated with embryonic *bcd* than with ovarian *bcd* (Fig. 5.14b). Analysis of miR-989 levels revealed a more drastic reduction of 99.99% as the egg transitions into an embryo (Fig. 5.14c). The temporal profile of miRNA-*bcd* binding, especially for miR-989, is entirely consistent with the expected binding profile of a putative repressor, which is thought to be associated to repress *bcd* in ovaries and relieved from its target in early embryos, when *bcd* is translated. To confirm the functional relevance of these interactions, future work will entail *in situ* mutagenesis of the MREs with subsequent monitoring of *bcd* translation.

5.3.12 *bcd* 3'UTR confers post-transcriptional repression in S2R+ cells

Heterologous sensor assays are frequently employed to validate miRNA::target interactions. In an attempt to pinpoint the location of a putative repressing *cis* element, we assessed the ability of the *bcd* 3'UTR to silence a reporter construct. To this end, we generated a sensor construct in which the sequence for the fluorophore mCherry is appended with the *bcd* 3'UTR (Fig. 5.15a). The same sensor without the 3' UTR served as the control. If indeed the 3'UTR harbours *cis* elements responsible for specific *bcd* silencing, it is likely that these also silence the expression of a reporter construct. To test this hypothesis we transfected *Drosophila* S2R+ cells with the mCherry sensor either with or without the *bcd* 3'UTR and measured mCherry fluorescence normalised to EYFP which served as a transfection control (Fig. 5.15a).

Remarkably, the *bcd* 3'UTR was capable of reducing mCherry fluorescence by ~78% (Fig. 5.15b).

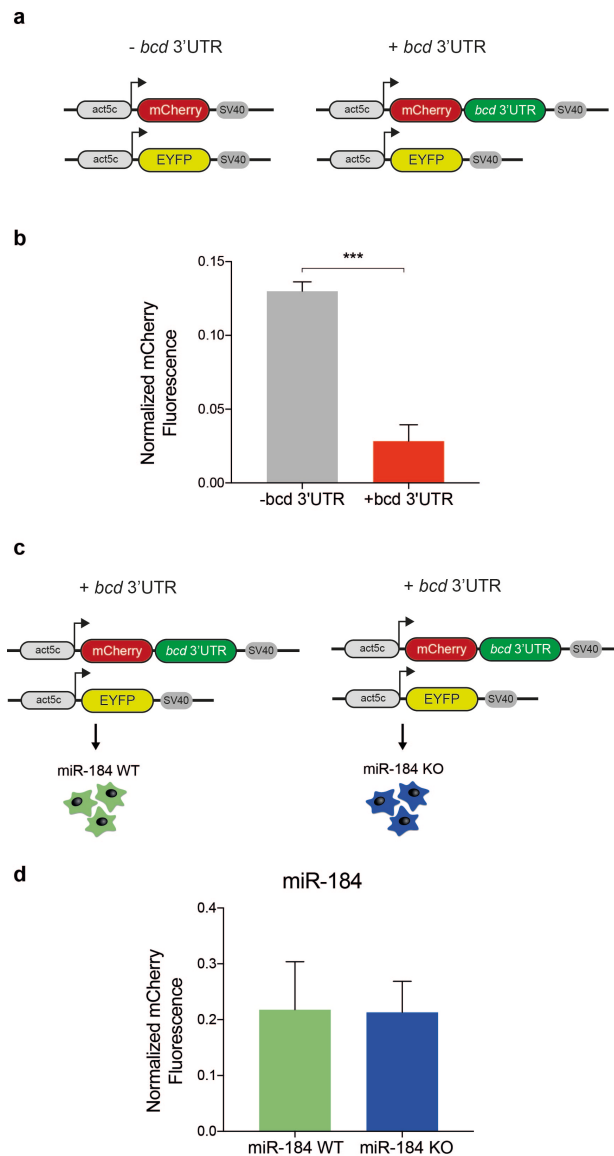


Figure 5.15 | *bcd* 3' UTR represses reporter fluorescence but not *via* miR-184. **a**, a two fluorophore sensor system was developed. mCherry expression is measured both in the absence and presence of the *bcd* 3'UTR. A separate plasmid expressing enhanced yellow fluorescent protein (EYFP) served as the transfection control. **b**, Flow cytometric fluorescence analysis revealed that fusing the *bcd* 3'UTR to the reporter leads to a marked downregulation of fluorescence. **c**, to test whether this effect is mediated *via* miR-184 we transfected mCherry**bcd* 3'UTR into both wildtype (WT) and miR-184 knock out (KO) S2 cells. **d**, no relief of repression is observed in the miR-184 KO cell line, suggesting miR-184 does not target *bcd* 3'UTR. Error bars = standard deviation, n = 3. *** $P < 0.0005$. Statistical significance was calculated using student's t-test.

Since a miR-184 mutant S2R+ cell line has been previously generated in our lab we decided to test whether this miRNA was responsible for the observed repression imparted by the *bcd* 3'UTR (Fig. 5.15c). Of note, the only predicted miR-184 canonical MRE resides in the ORF of *bcd*. However, considering the abundance of this miRNA in ovaries, it be possible that other (non-canonical) 3'UTR MREs may exist and mediate repression. Perhaps not surprising, in line with the absence of a strong MRE in the 3'UTR, transfecting the mCherry-*bcd*-3'UTR sensor into both WT and KO cells revealed no significant difference in reporter signal (Fig. 5.15d). This indicated that, if indeed miR-184 contributes to *bcd* repression, it is not *via* the 3'UTR.

5.4 Discussion

miRNAs provide the specificity code for many post-transcriptional silencing events by directing the miRISC complex to target RNAs. Based on several lines of evidence we hypothesised that miRNA-mediated repression could underlie the remarkable quiescence of *bcd* transcripts throughout oogenesis. We undertook a multipronged approach to provide some initial clues for this long-standing biological conundrum.

5.4.1 *Drosophila* ovaries contains miRNAs and piRNAs predicted to target *bcd*

In this chapter, we first employed the Nanostring platform to establish the miRNA complement at four distinct stages of oogenesis. We found 40 miRNAs to be significantly expressed in at least one of these developmental stages. Previously, Neumüller *et al.* reported 43 expressed miRNAs in the *Drosophila* female germline (Neumüller *et al.*, 2008). We noted 45.6% overlap (data not shown) between our dataset and the one reported in that study. At least two reasons could account for this discrepancy. First, Neumüller and colleagues used a more sensitive amplification system (qPCR) whilst we used digital counting of mature miRNAs in an amplification-free experiment. Second, perhaps more plausible, the number of known miRNAs has drastically increased between 2008 and 2012 (when our run was performed). Consequently, the miRBase miRNA registry used to inform the design of both qPCR and Nanostring assay probes has since then undergone several new releases.

We next screened these expressed miRNAs for putative MREs in the entire *bcd* transcript using the miRanda algorithm source code with standard parameters. This broad analysis identified several miRNAs predicted to target *bcd*. Seven out of the eight corresponding MREs were predicted in the *bcd* 3' UTR.

Previous studies reported miR-989 as the most abundant miRNA in the ovary, comprising 15.9% of mapped sequencing reads (Czech *et al.*, 2008; Kugler *et al.*,

2013). This profiling was performed on whole ovary tissue consisting of a very heterogenous cell population. We generated for the first time a miRNA profile with temporal resolution, enabling analysis of miRNA expression at distinct developmental transitions. Interestingly, miR-989, which has two predicted MREs in the *bcd* sequence, drops rapidly from stage 9 onwards. This is consistent with a function of miR-989 during border cell migration, which occurs immediately prior to and during stage 9 (Kugler et al., 2013). While the miR-989 decline following stage 9 makes it an unlikely candidate for *bcd* silencing in late oogenesis, it does not preclude a function for miR-989 in early *bcd* silencing. Interestingly, by stage 9 of oogenesis, *bcd* is already trapped in P-bodies (Weil et al., 2012). Considering the role of miRNAs in P-body-mediated translational silencing, it is intriguing to speculate that miR-989 may play a role in this process. Although *bcd* remains homogenous in regard to its translational state, it is certainly possible that temporarily distinct molecular processes underlie its complex translational quiescence control. As the precise post-transcriptional regulation of *bcd* constitutes a fundamental component in development, a multi-step mechanism similar to that reported for its localisation is conceivable (St Johnston et al., 1989).

A bioinformatics meta-analysis on ovary and embryo piRNA profiles, using the same pipeline employed by Rouget *et al.* to uncover the piRNA-mediated regulation of *nos* (Rouget et al., 2010), revealed several piRNA clusters that bear complementarity to *bcd*. Of particular interest were two piRNA families that mapped to the *bcd* ORF. These derived from the *roo* and *stalker* transposons and emerged as potential candidates not only due to their high expression in the ovary but also based on their temporal profile. Both piRNA clusters exhibited a considerable downregulation at the egg to embryo transition. Since piRNAs remain relatively novel players in post-transcriptional regulation, these observations will require further extensive validation.

Future experiments could include the mutation of their corresponding target sites or the abrogation of the *bcd*-piRNA interaction by genomically ablating the source transposons. However, considering that piRNAs represent the vanguard of genome defence against transposons in the germline and are required for proper gamete formation, discrete mutagenesis of the *bcd* target sequences may be the preferred choice.

5.4.2 TRACE uncovers the RNase H accessibility of the 3'UTR across the OET

To investigate which regions of *bcd* serve as potential regulatory *cis* elements mediating translational repression, we mapped the 3' UTR accessibility using TRACE. The technology had hitherto only been applied to candidate MREs, however, we reasoned that it could also serve in an exploratory manner, to 'walk' an RNA with tiled ASOs. The results in chapter 3 suggested that the assay sensitivity was high enough to compare RNA accessibilities between different biological conditions (e.g. miR-184 WT and KO). Hence, we took advantage of *bcd*'s differential fate in oocytes (repressed) and early embryos (translated) and compared the relative 3' UTR accessibility across these tissues. Overall, we noted a marked increase in accessibility in embryos, which matched the cellular fate of *bcd*. In the ovary, *bcd* is sequestered in the core of P-bodies and remains translationally inactive, most likely bound by a translational repressor (Weil et al., 2012). Conversely, in the early embryo *bcd* is freed from its repressor and released from the anterior cortex, consistent with a higher accessibility.

We next focused on the accessibilities of previously defined *cis* elements to assess whether the TRACE-generated maps correspond to previously reported occupancy at these sites. BLE1 had previously been defined as the *cis* element regulating early *bcd* localisation (Macdonald et al., 1993). Consistently, we noted a very low

accessibility at this site, compared to the mean ovarian *bcd* accessibility. Surprisingly, when examining the corresponding region in embryos, we did not notice an increase in accessibility, which is in contrast to the expected dissociation of *trans* factor mediating localisation at a much later stage (embryos). We analysed a second previously reported *cis* acting nonamer (Gottlieb, 1992) and noted a significant increase in accessibility in embryos. This suggested a function before egg activation, which may be why this site is no longer actively bound in the embryo. The multifactorial apparatus directing *bcd* localisation is highly dynamic and varies, depending on which stage of oogenesis *bcd* is undergoing. Thus, a limitation of TRACE maps is that in order to meet sufficient input requirements, whole ovaries had to be dissected. Although efforts were undertaken to stage egg chambers to a median developmental stage 10, this preparation inevitably contained a heterogenous population of egg chambers, which could obscure the study of distinct localisation elements active at different stages. Nevertheless, this should have a less pronounced effect on the study of *bcd* translational regulation, as it presumably contains a homogenous *bcd* population with regard to its translational state. Measuring differential occupancy at 6 predicted MREs, we noted a significant increase in *bcd* accessibility in embryos in 4/6 cases that exceeded the 'background' increase from oocytes to embryos. This observation is again consistent with the expected profile of repressor release in the embryo at one of these sites. An alternative to comparing ovaries to embryos could entail replacing the ovary fraction with *sarah* (*sra*) mutant embryos. Mutants of the calcipressin gene *sra* cannot re-enter meiosis and thus produce sterile eggs (Horner et al., 2006). Consequently, *bcd* is not activated and remains silenced. While the comparison of wild type ovaries to wild type embryos remains physiologically most relevant, *sra* mutant embryos are an alternative way in which translated and repressed *bcd* could be assessed in the

same tissue (embryo). This would reduce experimental artefacts introduced by different crosslinking efficiencies of ovaries vs. embryos or heterogeneous egg chamber populations. Nevertheless, to define a clearer picture of miRNA-*bcd* binding we sought to integrate these findings with an orthogonal pulldown approach, which captures *trans* interacting RNAs from native *in vivo* UV crosslinked *bcd*.

5.4.2 Development and implementation of a *bcd* small RNA interactome capture protocol

While profiling small RNAs provides a cost efficient and rapid overview of possible target interactions, the mere co-expression of computationally paired small RNA and target transcripts by no means guarantees functional relevance. The effector step of *bcd* activation is most likely manifested by the physical dissociation of the silencing agent from the *bcd* transcript. Profiling alone is incapable of detecting such events. Furthermore, profiling presupposes that in order to identify a silencing agent, this molecule should be absent from the whole tissue or at least strongly reduced in the translated state (i.e. early embryo for *bcd*). Since it remains unclear whether any miRNAs are cleared from embryos within the 1.5h collection window, profiling can only indirectly inform on the identity of putative repressors. Extending the time post-fertilisation is not a viable option either since *de novo* transcription of small RNAs as part of the MZT is expected to drastically shift the small RNA profile, thereby obscuring a comparative analysis with oocyte expressed miRNAs (Bushati et al., 2008). Therefore, we reasoned that the most appropriate way to identify *trans* interacting factors, and thus circumventing these caveats, is to carry out an *in vivo* crosslinked *bcd* pulldown followed by unbiased analysis of associated small RNAs by next generation sequencing.

To develop a suitable protocol, we surveyed multiple pulldown assays ranging from RNA chromatography to antisense purification of target RNAs from *in vivo* UV-

crosslinked lysate. We eventually compiled a protocol that incorporates intelligent probe design against accessible regions in the *bcd* transcript and the experimental framework developed by Castello and colleagues (Castello et al., 2012; 2013b; Simon, 2013). Following various rounds of optimisation we reached considerable levels of *bcd* RNA enrichment (eluate *versus* input, as quantified using a housekeeping gene), approaching 20,000-fold. This efficiency outperforms the previous miR-CATCH approach developed by Hassan *et al.* who used chemically crosslinked (formaldehyde) samples to achieve only 16-87 fold enrichment (Hassan et al., 2013). The authors of another protocol termed CHART claimed 50% recovery of target RNA from the input material but did not provide a fully quantitative analysis of the eluate fraction purity (Simon et al., 2011). Housekeeping genes *ep17* and *act5C* were measured by qRT-PCR but only quantified relative to the input (Simon et al., 2011). Engreitz and colleagues presented a technology termed RNA Antisense Purification (RAP), which used a mixture of formaldehyde and disuccinimidyl glutarate (DSG), and provided approximately 100-fold enrichment of *Xist* lincRNA relative to input (Engreitz et al., 2013). Notably, although conceptually similar, all these alternative approaches relied on chemical crosslinking, which lacks the specificity of UV crosslinking employed in our case.

Measuring fold enrichment by analog qPCR is inherently biased as the eluate is deliberately depleted of control RNAs, which leads to an overestimation of the actual amount of target molecule in the eluate. Absolute quantification of contaminant RNA species using ddPCR is a superior method to establish the purity of pulldown samples and provides a reliable quality control for subsequent small RNA interactome sequencing. We performed ddPCR on a preliminary pulldown sample (3rd iteration, prior to the protocol improvement of the 60 °C lysate pre-incubation) and noted absolute *bcd* enrichment accompanied by considerable depletion of *rp49*,

gapdh, and *act5c* housekeeping genes RNA. Although not a cost-effective option, a more accurate solution for assessing the actual RNA composition of eluate fractions could be regular long RNAseq.

The biochemistry of the pulldown protocol underwent several important improvements. Whilst a 20 min lysate pre-incubation at 60°C and an elevated DTT concentration improved the pulldown efficiency considerably (Sysoev et al., 2016), the most effective upgrade stemmed from the pre-absorption of capture beads with a synthetic IVT RNA. This observation suggests that fresh magnetic Dynabeads are 'sticky' and non-specifically bind to various RNA species. This pre-absorption step offers a massive improvement in pulldown efficiency and should represent an important consideration to any pulldown experiment beyond this study.

Although we successfully generated a pilot library from one ovary pulldown sample, we encountered several technical difficulties during the procedure, which hindered the generation of additional libraries. The standard small RNAseq library preparation recommends no more than 15 cycles of amplification. However, capturing a single RNA species naturally entails very low RNA yields. In fact, the RNA concentration in the eluate fraction was below the detection limit of the Qubit spectrophotometer. However, the presence of our target *bcd* RNA could be readily determined by qPCR. At 15 library amplification cycles no Bioanalyzer peak corresponding to the expected miRNA or piRNA sizes was detected. Upon increasing the cycle number to 25-35, we observed peaks within the region of interest. A 30 cycle amplification offered the best compromise between desired peak size and adaptor dimer contamination. Since these are within very close proximity of each other (20bp), we resorted to separation by Pippin targeted electrophoresis. To maximise our chances for successful sequencing, we chose to accept significant adaptor dimer contamination, as it was impossible to scale the library without compromising on adaptor dimers. However,

Pippin size selection led to a substantial loss of material. Although the Pippin manufacturer advertises only 50% sample loss, we noted a library loss amounting to ~95%. Nevertheless, the library was well detectable and subsequently sequenced. For indeterminate reasons, automatic demultiplexing of reads generated by this run failed. Although this phenomenon is not uncommon due to reverse complemented index sequences (Illumina, pers. comm.), we did not observe this phenomenon in our library. This left the option of un-effective base calling due to low sequencing across the indices. This explanation is supported by the presence of some poly(T) and poly(A) motifs in the index regions. Fortunately, we were able to resort to a different strategy to recover our data. Since only a single *Drosophila* ovary sample was sequenced we were able to circumvent the faulty indexing and isolate reads on the basis of the 5' or 3' RNA adaptors. These served as unique identifiers for our sample as other, unrelated libraries sequenced in the same run were generated using different adaptors.

Failure to multiplex our small RNA libraries meant no further replicates could be generated. The underlying cause for data demultiplexing difficulties was most likely contamination by adaptor dimers that emerged as a consequence of excessive PCR cycling. Unfortunately, with scarce input RNA the only way to generate libraries is through scaling the number of cycles. Interestingly, many of the recently developed protocols for single cell RNAseq also rely on extremely low RNA input, and higher cycle amplification is not uncommon (Picelli et al., 2014). Nonetheless, improvements on the quantity of RNA recovery are most likely the most promising avenue in making this protocol reproducible in the future. We identified several bottlenecks where input RNA quantity could possibly be maximised:

1. RNA degradation by exogenous and endogenous RNases is a common problem in RNAseq protocols. Although utmost care was taken during sample processing steps,

a 3h incubation at 40°C inevitably engenders RNA degradation even in the presence of broad-spectrum recombinant RNase inhibitors. While Bioanalyzer RNA Pico Chip analysis remained uninformative regarding RNA quality due to low input, we tried to control the integrity of *bcd* RNA by placing qPCR primers 500bp upstream of the pulldown probes target sites. Using this primer pair, very high *bcd* levels were detected at the end of the pulldown protocol, suggesting RNA degradation was not a major concern.

2. Elution of RNA from capture beads appears to be notoriously inefficient. Accordingly, we noticed that a considerable amount of *bcd* remained attached to the beads despite 5 min elution at 95°C. To maximise the *bcd* yield, we have explored strategies to improve elution efficiency. Analogous to the principle of elution by competition applied in protein chromatography, we delivered *bcd* sense oligonucleotides to the elution buffer, which would be expected to compete with *bcd* for the probe. Provided these decoys are delivered at a far higher concentration we reasoned they would outcompete and thus elute *bcd* in a specific manner. Unfortunately, we did not observe a meaningful improvement in elution using this implementation. Alternatively, the total probe-bound *bcd* fraction could be directly subjected to downstream analysis, thus circumventing an elution step, as previously reported for TINCR lncRNA interactome capture (Kretz et al., 2012). However, we were concerned that non-specific interactions on the beads (for example retention of the pre-absorption IVT RNA) could affect the quantity of the *bcd* interactome that is reliably sequenced, and hence hitherto have abstained from this option.

3. Another option that could increase the amount of captured RNAs is the use of a stronger crosslinking agent, for improved stability of mRNA-miRNA-protein ternary complexes. Although UV crosslinking has the advantage of superior specificity, it is inherently inefficient and only crosslinks *in vivo* RNA protein interactions at relatively

low frequency (Pashev et al., 1991) (Toth and Biggin, 2000). As an alternative strategy, we tested formaldehyde crosslinking. However, *bcd* enrichment was not satisfactory (<100-fold *versus* input) and housekeeping genes were massively abundant in the eluate fraction. This is not surprising considering that formaldehyde crosslinks not just RNAs to proteins but also proteins to proteins. Therefore, *de facto* formaldehyde crosslinking leads to massive amalgamations of ribonucleoparticles, which explains why this approach did not produce the purity required for a meaningful sequencing experiment. In cell culture systems, several photo-inducible and reversible RNA-RNA crosslinkers, such as psoralin, have been successfully implemented (Baigude et al., 2012; Harris and Christian, 2009). However, like nucleotide analogs, such as 4SU, they are difficult to incorporate into adult flies.

4. An obvious strategy is scaling up the input material. Presently, about 300 yeast fed female flies were used to dissect ovaries for each pulldown sample. This number becomes limited by how many flies can be manually dissected in the selected cut off time for proceeding to sample preparation (1 hour). The collection of embryos, however, can be scaled up more effectively with the use of population cages. These accommodate up to 100,000 flies per unit and once the flies are conditioned to synchronise egg lays, several grams of embryos can be harvested from a single one hour collection (Sandmann et al., 2007; Sysoev et al., 2016). Employing a smaller scale setup we collected ~200 mg of embryos from a short collection window of 1.5h which served as our input material for the translated *bcd* interactome. We envision that this setup, in conjunction with replacing ovary dissections with *sra* mutant embryo collections, could massively upscale input material. This would have the additional benefit of direct comparison of differential association of any repressor agent without the caveat of different tissue processing. Encouragingly, the poly(A)+ RNA protein interactome was recently isolated efficiently from whole embryos,

delivering proof-of-concept for this strategy (Sysoev et al., 2016). However, extrapolating this strategy from isolating whole tissues polyadenylated RNAs (~1% of all cellular RNAs) to a single RNA species remains a challenging task.

Taken together, including these alternative strategies in future studies should improve *bcd* small RNA interactome sequencing and improve the direct comparison between ovary (or *sra* mutant embryos) and embryo developmental stages.

5.4.3 *bcd* repression – small RNA or protein?

Since Wispy mediated polyadenylation of other maternal mRNAs is evidently occurring during late oogenesis, it is clear that *bcd* translational silencing in the ovary is selective (Benoit et al., 2008). Recently, the increase in *bcd* translational efficiency (TE) with regards to its poly(A) tail elongation was measured across the OET (Eichhorn et al., 2016). While the median fold-increase in TE of all RNAs across the OET was only 1.2-fold, *bcd* increased 36-fold. The corresponding median poly(A) tail increases were 1.2-fold for global RNAs and 1.7-fold for *bcd*. The ratio of TE and poly(A) tail length for *bcd* is considerably higher than for other RNAs, suggesting that *bcd* undergoes additional, poly(A) tail independent regulation.

Remarkably, transplantation of wild type stage 14 ooplasm taken from the anterior pole can rescue *bcd* null mutant oocytes, while cytoplasm of younger oocytes or of nurse cells cannot (Sander and Lehmann, 1988). This observation supports the view that *bcd* is actively repressed during earlier stages of oogenesis and suggests that by stage 14 *bcd* is already prepared for translation. Corroborating this hypothesis, a recent study challenged the dogma of Bicoid absence during oogenesis and presented evidence of the protein during stage 14c oocytes (Ali-Murthy and Kornberg, 2016). However, oocytes of this maturity are already passaged into the lateral oviduct where a calcium wave is known to trigger activation. Therefore, it is possible that the authors were imaging activated eggs as opposed to stage 14

oocytes (York-Andersen et al., 2015). Nevertheless, the observation that stage 14 ooplasm readily rescues *bcd* null mutant oocytes, suggests two plausible scenarios: *i*) the cytoplasm of egg chambers younger than stage 14 contains a factor that physically represses *bcd*; *ii*) in stage 14 oocytes the *bcd* RNA sequence is modified to prepare it for active translation. It is well established that the effector step of *bcd* translation is the elongation of its poly(A) tail (Eichhorn et al., 2016; Sallés et al., 1994). However, if stage 14 oocytes contained translation competent poly(A) tail elongated *bcd*, this would have been noticed in total ovary extract polyadenylation assays (where *bcd*'s poly(A) tail is clearly shortened) (Benoit et al., 2008; Sallés et al., 1994). Furthermore, the recent study by Eichhorn and colleagues also suggests that *bcd* isolated from a clean preparation of stage 14 oocytes does in fact contain a significantly shorter poly(A) tail. Based on these considerations, the explanation that oocytes contain a translational repressor is more plausible. A re-examination of our Nanostring data revealed no marked changes in miRNA profiles between stage 10-13 and 14 of oogenesis. However, intriguingly, miR-989, one of the most abundant miRNA in early oogenesis, rapidly declines after stage 9. miR-989 is conserved, has three predicted miRNA target sites in *bcd* and is specifically expressed in ovaries but not embryos (Ruby et al., 2007). Moreover, we abundantly isolated miR-989 in a *bcd* pulldown from ovaries and did not detect it in embryos. Based on these observations, one could speculate that miR-989 mediates *bcd* translocation to the P-body (which takes place prior to miR-989 downregulation), and thus constitutes one mechanism underlying *bcd* repression.

In addition to miR-989, another top candidate recovered in the *bcd* pulldown was miR-184. Although the reduction of *bcd*-associated miR-184 from ovary to embryo is nowhere close to the magnitude of miR-989, severe patterning defects have been reported in miR-184 mutant females (Iovino et al., 2009). 85% of flies lacking miR-

184 displayed severe anteroposterior axis defects. Interestingly, whilst pair-rule gene *fushi tarazu* (*ftz*) and *odd-skipped* (*odd*) expression was perturbed, the expression of gap genes (targets of Bicoid transcription factor) appeared unaffected. This phenotype was attributed to overexpression of *tramtrack69* due to the release of repression mediated by its 8mer miR-184 MRE. A potential link to *bcd* was not mentioned in this study despite the presence of a miR-184 6mer MRE in the *bcd* ORF. The *bcd* 3'UTR sensor experiment performed in our study remained inconclusive since the predicted MRE was not present in this construct. In general, it remains doubtful that the miR-184 6mer MRE (weakest class) can mediate the robust silencing effect underlying complete *bcd* translational quiescence.

Binding of miRNAs to their targets is necessary but not sufficient for effective silencing. Determining if miR-989 or miR-184 regulate *bcd* *in vivo* requires validation in form of deleting the putative MRE. The advent of CRISPR genome engineering now allows the precise dissection of MRE-miRNA axes with nucleotide resolution (Bassett et al., 2014). Future experiments aiming to establish the functional relevance of these interactions under physiological conditions will entail deletion of corresponding MREs and subsequent immunostaining for ectopic Bicoid expression and assessing *bcd* location with regards to P-bodies.

So far we have hypothesised that *bcd* activation entails the removal of a repressor. However, the opposite – association with an activator – also needs to be considered. Such a mechanism has been reported in the translational regulation of other maternal mRNAs. For instance, *grk* translation requires the association with Oo18 RNA binding protein (Orb) (Chang et al., 2001). Although miRNAs have been implicated in translational activation, these activities appear to remain a controversial exception (Truesdell et al., 2012). Translational activation is typically afforded by proteins. While *bcd* localisation is well known to depend on proteins, a super-

resolution and ultrastructural analysis of late stage oocytes and early embryos co-localised *bcd* with the P-body marker and translational repressor DEAD-box helicase Maternal expression at 31B (Me31B) (Weil et al., 2012). Interestingly, *bcd* overlaps with Me31B right up to egg activation, which coincides with P-body dispersal. The same spatio-temporal observation has been made for the repressor protein Trailer hitch (Tral) and Exuperantia. Furthermore, decapping enzyme Dcp2, a key effector immediately preceding 5'-to-3' RNA degradation, has also been highlighted in the core of P-bodies. While the co-localisation between *bcd* and these repressor proteins suggests regulation, it does not address the key question of how this specificity is achieved. With a few exceptions, most RBPs lack extensive sequence specificity but rather recognise short 6-8mer motifs (Jankowsky and Harris, 2015), which questions their involvement independently of miRNAs in coordinating *bcd* translation. By virtue of Watson-Crick base pairing, short non-coding RNAs, foremost miRNAs, are preferentially poised to serve as molecular specifiers of post-transcriptional silencing and hence likely to be involved in this remarkable silencing event. While the differential association of *bcd* with P-bodies provided a landmark discovery in the effector step, the underlying exact molecular mechanisms governing this process are still elusive. Since interactome capture was originally developed as a means to identifying RBPs, the *bcd* specific pulldown presented here is readily adaptable to investigate the protein complement. This would provide a full picture of the small RNA and RBP interactome of *bcd*. Deciphering the detailed *cis/trans* axis of the *bcd* regulation – be it small RNA or protein – will inevitably involve the mutagenesis of relevant *cis* elements in the *bcd* sequence. Our study suggests that the candidate MREs targeted by miR-989 and miR-184 serve as promising starting point for such investigations.

5.4.4 Final conclusions

In this chapter I have presented a multipronged approach directed at deciphering the translational regulation of *bcd* RNA, which uncovered the first link between *bcd* silencing and miRNAs. Employing profiling coupled to computational target prediction we identified several candidate miRNAs possibly facilitating *bcd* repression during oogenesis. We developed a robust pulldown protocol for isolation of *bcd* and its small RNA interactome which for the first time presented physical evidence that *bcd* is associated with certain miRNAs. Consistent with the temporal profile of *bcd* translation, the top candidate, miR-989, co-purifies with ovarian *bcd* but not with embryonic *bcd*. To confirm the role of miRNAs in *bcd* silencing and to uncover the functional impact of these interactions, future work will need to focus on the disruption of relevant *cis* regulatory elements (MREs).

Chapter 6: General Discussion and Future Perspectives

The diverse regulation of RNA – ranging from localisation to destabilisation – is typically governed by *cis* regulatory elements that are recognised by a plethora of *trans* acting RNA binding proteins (RBP) and short non-coding regulatory RNAs. Of particular consequence for the stability and translation activity of a mRNA are interactions with miRNAs (Bartel, 2009). Hence, deciphering the recognition code between miRNAs or RBPs and their target RNAs lies at the heart of understanding post-transcriptional regulation. Notably, progress in this field has largely been dictated by the technologies available to study such interactions. The relatively amenable implementation of next generation sequencing (NGS) as a read out system has led to unprecedented throughput in many of these applications. However, on its own, NGS is relatively limited; intelligent experimental setups continue to dictate to a large extent the advances made in the post-transcriptional regulation field. Over the past five years a slew of technologies has been developed to explore the regulatory potential of *cis/trans* RNA interactions. The most recent implementations allow unbiased, multiplexed screening for regulatory motifs on the basis of reporter abundance, and subsequent identification of the respective *cis* elements by sequencing (Yartseva et al., 2016; Zhao et al., 2014). While extremely valuable for exploratory studies, the physiological relevance of these interactions often remains unknown. Most of these assays, including the recently published RESA, do not assess *cis* elements under endogenous conditions (Yartseva et al., 2016). Therefore, these technologies rarely allow the study of *cis* elements under normal cellular homeostasis. Instead, RESA and other implementations rely on delivery of artificial constructs that are driven by exogenous promoters and thus exist at a non-physiological stoichiometry. This has particular impact for identifying

putative targets of miRNAs, as the overexpressed transcript levels do not necessarily reflect the regulation of their endogenous counterparts. Here, I have presented the development of *in vivo* detection of miRNA::target interaction events (TRACE), a novel and facile platform technology for detection of miRNA::target interactions in their endogenous context.

6.1 *In vivo* detection of miRNA-target interaction events (miR-TRACE) as a novel miRNA target identification platform

The conceptual framework of TRACE is based on *in vivo* crosslinking followed by RNase H cleavage of putative MREs. The binding of miRISC at individual target MREs is analysed *in situ*, without any external intervention. We have demonstrated the efficacy of this technology on a previously validated primary network of miR-184 targets. A direct comparison of target MRE accessibility was conducted between S2R+ wild type and miR-184 KO cells. As expected, MREs reported to be active by transcript fold change (miR-184 KO/WT) and CRISPR-Cas9 mediated MRE mutagenesis, revealed a significant increase in accessibility from WT to miR-184 KO cells. Conversely, inactive MREs revealed no significant differences in their accessibility between these conditions.

Originally, we developed this system in a 'one-ASO:one-MRE' modus; every MRE is hybridised to a perfectly complementary 25mer ASO to assess its accessibility. In an effort to scale the throughput of the technology, we considered two complementary options: *i)* grouping 25mers based on minimal inter-target off-targets and *ii)* employing a universal ASO targeted against the miRNA seed (present in every target of a primary miRNA network). Since the universal ASO approach could reproduce *solo* TRACE scores (one-ASO:one-MRE) and offered a massive scalability

advantage we used it to explore the global miR-184 MRE network based on differential accessibilities between WT and miR-184 KO cells.

To scale up the readout of the miR-TRACEseq assay we devised a seed-targeted reverse transcription protocol. The conceptual framework underlying this strategy relies on the assumption that the targets containing accessible MRE seeds will be destroyed by the complementary ASO upon RNase H digestion. Conversely, target RNAs bearing protected seeds will be reverse transcribed. To this end, we optimised RT conditions (primer concentration and hybridisation temperature) that allow seed specific priming and initiation of cDNA synthesis from this short unique sequence. A pilot MiSeq sequencing experiment revealed an unexpected technical issue: the majority of reads mapped to ribosomal RNA (rRNA) loci, thus absorbing critical bandwidth and preventing a meaningful target analysis. This outcome was surprising, since none of the top contaminating rRNA species exhibited obvious RT priming sites. Since poly(A) selection as a means to eliminate rRNA would prevent the analysis of multiple MREs per transcript (only the most proximal MRE to the poly(A) tail would be isolated) we resorted to a different method of ribo-depletion. The recently developed depletion of abundant sequences by hybridisation (DASH) protocol specifically eliminates undesired sequences in a library preparation by Cas9-targeted dsDNA cleavage *in vitro* (Gu et al., 2016). We multiplexed 18 CRISPR guide RNAs against the top contaminating rRNA species and observed almost complete destruction of rRNAs. In addition to the sequencing reads that are liberated by the rRNA elimination, we also scaled up the experiment to a NextSeq read out. This sequencing platform generated ~25 more reads than the previously used MiSeq. As expected, following these two modifications we noted a considerable increase in sequencing depth, which was reflected in sufficient reads over target sites, permitting downstream analysis.

The seed-targeted cDNA synthesis enabled genome-wide coverage of miR-184 seed binding. Accordingly, we built a general linearised logistics model to uncover differential miR-184 binding across the transcriptome. The model predicted 48 high confidence MREs responsive to miR-184. These included several previously validated MREs as well as novel ones. Importantly, unlike many computational prediction platforms, miR-TRACEseq is not limited to the 3'UTR, but rather explores all gene structures. Consistent with the previously reported relative location of MREs, we have noted 47% of our hits in the 3' UTR, 18% in the 5' UTR, and 35% in the CDS. We then conducted a motif analysis, which revealed that nucleotide positions immediately adjacent to the mRNA seed region were most frequently represented by nucleotides A (position 1) and U (position 8). This observation suggests the majority of TRACE identified MREs are 8mer target sites. This is consistent with the general rules of MRE strength, where 8mers are designated as the most potent motifs (Bartel, 2009). Future work will entail the validation of the novel MREs through orthogonal approaches (CRISPR mutagenesis). Interestingly, TRACE may also have uncovered several MREs that confer translational inhibition, positing the assay at a vantage point to measure both mRNA destabilisation and translational inhibition.

In its present format, miR-TRACEseq was applied to only search for perfect 6mer seed sites of miR-184. However, it is conceivable to assess the importance of each individual nucleotide in the seed through generating semi-random, partially degenerate 10mers with a mismatch at varying positions in the seed. Hence, with as few as six 6mers, the entire seed composition could be probed at nucleotide resolution.

In this study we have generated miR-TRACEseq MRE maps for a single miRNA only (miR-184). To establish the wider applicability of this technology, additional miRNAs have to be subjected to analysis. Although miR-TRACEseq has only been tested for

a small RNA, it is conceivable to use this setup for any RBP, which has a defined recognition motif (to guide the design of an ASO), and compare accessibility maps under WT and RBP-KO conditions.

Deconstructing the molecular architecture of RBP-RNA interactions also offers ample opportunity for therapeutic and biotechnological exploitation. For instance, understanding the specific requirements for *cis/trans* interactions may lead to the development of 'designer' RBPs. A breakthrough has recently been achieved by repurposing the CRISPR-Cas9 system, which had hitherto only been considered as a dsDNA binding agent, for RNA engineering (O'Connell et al., 2014). Jennifer Doudna and colleagues successfully applied the logic of Cas9 binding to single stranded RNA (ssRNA) through the provision of ssDNA PAMmer acting *in trans* (O'Connell et al., 2014). Following their demonstration of *in vitro* cleavage and an *in vitro* RNA pulldown, the most recent implementation reported native RNA labelling in living cells by virtue of a dCas9 fluorophore fusion protein (Nelles et al., 2016). CRISPR-based RNA targeting is not only useful for tag-free, native RNA labelling but also for engineering the fate of an RNA, for example through modifying splicing, or increasing/decreasing its translation rate (Nelles et al., 2015). Such developments could engender far-reaching benefits for pathologies originating from defective RNA metabolism or erroneous splicing. For example, the marked success achieved through 'exon skipping' in Duchenne muscular dystrophy (DMD) highlights the possibility to use targeted RBPs to mask defective bases and hence restore the correct reading frame (Mendell et al., 2013; Voit et al., 2014).

6.2 Elucidating the molecular basis for *bcd* translational quiescence during oogenesis

Besides the development of a new technology platform for generic miRNA target identification, I have also pursued in parallel a distinct biological question. *Bcd* RNA is translationally quiescent throughout oogenesis and is translated strictly after egg activation. Several lines of evidence, foremost *bcd*'s association with P-bodies in the oocyte, led us to hypothesise that small RNAs, particularly miRNAs, mediate this process. To this end I have developed a multi-faceted approach that incorporates a miR-TRACE generated *bcd* 3' UTR accessibility map from ovaries and embryos, a bioinformatics prediction analysis, and small RNA interactome capture analysis.

Based on the striking difference of the translational profile between ovarian *bcd* (quiescent) and embryonic *bcd* (translated) we hypothesised the RNA is bound by a silencer (e.g. miRNA) in the quiescent state, but not in the translated state. We hence reasoned that miR-TRACE should be capable of detecting differences in accessibility (repressor bound *versus* unbound) across *bcd* 3' UTR isolated from ovaries and embryos. Consistent with *bcd* release from its anchor at the anterior margin and dispersal of P-bodies following egg activation, we noted a significant increase in the mean accessibility in embryonic *bcd* (Weil et al., 2012). We measured the accessibility at six MREs and found that four of these even exceeded the 'background' increase from oocyte to embryo. This is consistent with the expected profile of repressor release in the embryo, and also provides further confidence in the sensitivity of the technology. Taken together, this suggested that TRACE can be readily applied to exploratory mapping of entire RNA regions (e.g. whole 830bp 3' UTR) across biological conditions and is hence amenable for unbiased *cis* element identification in candidate RNAs.

To substantiate this analysis, we performed a Nanostring-based miRNA profiling at four distinct developmental stages across oogenesis (germarium to stage 9, stage 9, stage 9-13, stage 14) and detected several miRNAs that have predicted MREs in the *bcd* transcript. Unfortunately, miRNA target prediction algorithms suffer from high false positive and false negative results (Rajewsky, 2006). Therefore, such algorithms are best used only in conjunction with complementary wet lab experiments. To this end, we have developed an RNA pulldown capable of isolating *bcd* with high specificity from UV crosslinked ovaries and embryos. We subjected the small RNA fraction from ovary isolated *bcd* to a small RNAseq experiment and retrieved several miRNAs predicted to target *bcd* as well as very few miRNAs that match non-canonical MREs.

The most abundant miRNA detected in the ovary pulldown was miR-184. Indicative of a putative repressor profile, miR-184 was reduced in a *bcd* pulldown from early embryos. Furthermore, AP and DV patterning defects had already been reported for the miR-184 null mutant (Iovino et al., 2009). However, the only predicted MRE is a relatively weak 6mer in the CDS region, raising concern regarding the efficacy of this site, especially in the context of the robust quiescence displayed by *bcd* during oogenesis. The second most abundant miRNA detected in the pulldown was miR-989. Intriguingly, miR-989 was completely depleted from the embryonic *bcd* interactome. This is consistent with its temporal profile as revealed by the Nanostring analysis. miR-989 is one of the most abundant miRNAs in oogenesis and sharply decreases following stage 9. This profile is compelling as by stage 9, *bcd* is already stored in P-bodies, suggesting a possible function in *bcd* shuttling into the P-body core (Weil et al., 2012). CRISPR-Cas9 mediated mutagenesis of the miR-989 MREs with subsequent immunostaining for translated Bicoid or a more refined

ultrastructural analysis of the *bcd* localisation relative to P-bodies could clarify the physiological significance of this interaction.

Following egg activation, P-bodies disperse, *bcd* is polyadenylated and becomes competent for translation. The external stimuli triggering egg activation is a single calcium wave (York-Andersen et al., 2015). It is presently unclear how the early embryo converts this signal to initiate *bcd* translation and complete meiosis. Future work could investigate whether miRNAs or perhaps *bcd* itself respond to calcium ions.

6.3 Final conclusions

The fascinating repertoire of cellular functions relies on various regulatory layers coordinating gene expression. Central to these processes, a complex code of *cis* regulatory elements and corresponding *trans* interacting factors, including RBPs and short regulatory RNAs, has evolved that allows the cell to direct RNAs to distinct fates. Here, I have developed in vivo detection of miRNA::target interaction events (TRACE), a novel method for the validation and discovery of MREs *in situ*. Future work in our lab will aim to demonstrate the general applicability of miR-TRACEseq principles to other miRNAs and cell types. Subsequently, studies using this technology in conjunction with other functional tools (such as GenERA) should engender a broader understanding of *cis/trans* interactions underlying the eukaryotic post-transcriptional regulatory landscape.

With regard to *bcd*, this study provides the first evidence for a potential role of miRNA-mediated regulation underlying *bcd* translational quiescence. This opens up an exciting new angle of investigation in the quest for defining the mechanism of what remains one of the most astounding accounts of gene expression regulation during the oocyte-to-embryo developmental transition.

Chapter 7: References

- Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, e05005.
- Ali-Murthy, Z., and Kornberg, T.B. (2016). Bicoid gradient formation and function in the *Drosophila* pre-syncytial blastoderm. *Elife* 5, e13222.
- Ameres, S.L., and Zamore, P.D. (2013). Diversifying microRNA sequence and function. *Nature Publishing Group* 14, 475–488.
- Arn, E.A., Cha, B.J., Theurkauf, W.E., and Macdonald, P.M. (2003). Recognition of a bicoid mRNA localization signal by a protein complex containing Swallow, Nod, and RNA binding proteins. *Developmental Cell* 4, 41–51.
- Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* 152, 844–858.
- Axtell, M.J. (2014). Butter: High-precision genomic alignment of small RNA-seq data. *bioRxiv*.
- Azzam, G., Smibert, P., Lai, E.C., and Liu, J.-L. (2012). *Drosophila* Argonaute 1 and its miRNA biogenesis partners are required for oocyte formation and germline cell division. *Dev Biol* 365, 384–394.
- Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes & Development* 22, 2773–2785.
- Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71.
- Baigude, H., Ahsanullah, Li, Z., Zhou, Y., and Rana, T.M. (2012). miR-TRAP: a benchtop chemical biology strategy to identify microRNA targets. *Angew. Chem. Int. Ed. Engl.* 51, 5880–5883.
- Baran, N., Haviv, Y., Paul, B., and Manor, H. (2002). Studies on the minimal lengths required for DNA primers to be extended by the *Tetrahymena* telomerase: implications for primer positioning by the enzyme. *Nucleic Acids Research* 30, 5570–5578.
- Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Bassett, A., and Liu, J.-L. (2014). CRISPR/Cas9 mediated genome engineering in *Drosophila*. *Methods* 69, 128–136.
- Bassett, A.R., Azzam, G., Wheatley, L., Tibbit, C., Rajakumar, T., McGowan, S., Stanger, N., Ewels, P.A., Taylor, S., Ponting, C.P., et al. (2014). Understanding

- functional miRNA-target interactions in vivo by site-specific genome engineering. *Nat Comms* 5.
- Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336, 233–237.
- Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E., and Meister, G. (2007). Identification of human microRNA targets from isolated argonaute protein complexes. *Rnabiology* 4, 76–84.
- Benes, V., Blake, J., and Doyle, K. (2011). Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nat Methods* 8, iii–iv.
- Benoit, P., Papin, C., Kwak, J.E., Wickens, M., and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. *Development (Cambridge, England)* 135, 1969–1979.
- Bergsten, S.E., and Gavis, E.R. (1999). Role for mRNA localization in translational activation but not spatial restriction of nanos RNA. *Development (Cambridge, England)* 126, 659–669.
- Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *The EMBO Journal* 7, 1749–1756.
- Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366.
- Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular Cell* 2, 437–445.
- Betancur, C., Sakurai, T., and Buxbaum, J.D. (2009). The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* 32, 402–412.
- Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11, R90.
- Béthune, J., Artus-Revel, C.G., and Filipowicz, W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* 13, 716–723.
- Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006a). Stress-induced reversal of microRNA repression and mRNA P-body localization in human cells. *Cold Spring Harb. Symp. Quant. Biol.* 71, 513–521.
- Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006b). Relief of microRNA-mediated translational repression in human cells

subjected to stress. *Cell* 125, 1111–1124.

Birikh, K.R., Berlin, Y.A., Soreq, H., and Eckstein, F. (1997). Probing accessible sites for ribozymes on human acetylcholinesterase RNA. *Rna* 3, 429–437.

Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L., and Ji, X. (2001). Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure* 9, 1225–1236.

Bohnsack, M.T., Czapinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* 10, 185–191.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 Proteins Directly Recruit Cytoplasmic Deadenylase Complexes to miRNA Targets. *Molecular Cell* 44, 120–133.

Braun, J., Misiak, D., Busch, B., Krohn, K., and Hüttelmaier, S. (2014). Rapid identification of regulatory microRNAs by miTRAP (miRNA trapping by RNA in vitro affinity purification). *Nucleic Acids Research* 42, e66–e66.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* 128, 1089–1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1387–1392.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of MicroRNA–Target Recognition. *Plos Biol* 3, e85.

Brizzard, B.L., and De Kloet, S.R. (1983). Reverse transcription of yeast double-stranded RNA and ribosomal RNA using synthetic oligonucleotide primers. *Biochim. Biophys. Acta* 739, 122–131.

Broccoli, S., Rallu, F., Sanscartier, P., Cerritelli, S.M., Crouch, R.J., and Drolet, M. (2004). Effects of RNA polymerase modifications on transcription-induced negative supercoiling and associated R-loop formation. *Mol. Microbiol.* 52, 1769–1779.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71, 515–526.

Broughton, J.P., and Pasquinelli, A.E. (2013). Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP. *Methods* 63, 119–125.

Brown, C.J., Lafreniere, R.G., Powers, V.E., Sebastio, G., Ballabio, A., Pettigrew, A.L., Ledbetter, D.H., Levy, E., Craig, I.W., and Willard, H.F. (1991). Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* 349, 82–84.

Brunel, C., and Ehresmann, C. (2004). Secondary structure of the 3' UTR of bicoid mRNA. *Biochimie* 86, 91–104.

- Bullock, S.L., and Ish-Horowicz, D. (2001). Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis. *Nature* 414, 611–616.
- Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dolken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *Rnabiology* 10, 1623–1630.
- Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in *Drosophila*. *Current Biology* 18, 501–506.
- Buszczak, M., and Cooley, L. (2000). Eggs to die for: cell death during *Drosophila* oogenesis. *Cell Death Differ.* 7, 1071–1074.
- Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* 10, 1957–1966.
- Cambronne, X.A., Shen, R., Auer, P.L., and Goodman, R.H. (2012). Capturing microRNA targets using an RNA-induced silencing complex (RISC)-trap approach. *Proc Natl Acad Sci USA* 109, 20473–20478.
- Carey, M.F., Peterson, C.L., and Smale, S.T. (2013). The primer extension assay. *Cold Spring Harbor Protocols* 2013, 164–173.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* 149, 1393–1406.
- Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell* 63, 696–710.
- Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013a). RNA-binding proteins in Mendelian disease. *Trends Genet.* 29, 318–327.
- Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsveld, J., and Hentze, M.W. (2013b). System-wide identification of RNA-binding proteins by interactome capture. *Nature Protocols* 8, 491–500.
- Cayirlioglu, P., Kadow, I.G., Zhan, X., Okamura, K., Suh, G.S.B., Gunning, D., Lai, E.C., and Zipursky, S.L. (2008). Hybrid neurons in a microRNA mutant are putative evolutionary intermediates in insect CO2 sensory systems. *Science* 319, 1256–1260.
- Cáceres, L., and Nilson, L.A. (2009). Translational repression of gurken mRNA in the *Drosophila* oocyte requires the hnRNP Squid in the nurse cells. *Dev Biol* 326, 327–334.
- Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. *Febs J* 276, 1494–1505.
- Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., and Crouch, R.J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in

Rnaseh1 null mice. *Molecular Cell* 11, 807–815.

Cha, B.J., Koppetsch, B.S., and Theurkauf, W.E. (2001). In vivo analysis of *Drosophila* bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway. *Cell* 106, 35–46.

Chang, J.S., Tan, L., Wolf, M.R., and Schedl, P. (2001). Functioning of the *Drosophila* orb gene in gurken mRNA localization and translation. *Development (Cambridge, England)* 128, 3169–3177.

Chekulaeva, M., Filipowicz, W., and Parker, R. (2009). Multiple independent domains of dGW182 function in miRNA-mediated repression in *Drosophila*. *Rna* 15, 794–803.

Chen, B., Harms, E., Chu, T., Henrion, G., and Strickland, S. (2000). Completion of meiosis in *Drosophila* oocytes requires transcriptional control by grauzone, a new zinc finger protein. *Development (Cambridge, England)* 127, 1243–1251.

Chen, C.-Y.A., and Shyu, A.-B. (2013). Deadenylation and P-bodies. *Adv. Exp. Med. Biol.* 768, 183–195.

Chen, C.-Y.A., Zheng, D., Xia, Z., and Shyu, A.-B. (2009). Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* 16, 1160–1166.

Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* 19, 321–327.

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

Chu, C.-Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *Plos Biol* 4, e210.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular Cell* 44, 667–678.

Chu, T., Henrion, G., Haegeli, V., and Strickland, S. (2001). Cortex, a *Drosophila* gene required to complete oocyte meiosis, is a member of the Cdc20/fizzy protein family. *Genesis* 29, 141–152.

Claycomb, J.M., and Orr-Weaver, T.L. (2005). Developmental gene amplification: insights into DNA replication and gene expression. *Trends Genet.* 21, 149–162.

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

Cooke, A., Prigge, A., and Wickens, M. (2010). Translational repression by deadenylases. *J. Biol. Chem.* 285, 28506–28513.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* 136, 777–793.

Crooke, S.T., Lemonidis, K.M., Neilson, L., and Griffey, R. (1995). Kinetic

characteristics of *Escherichia coli* RNase H1: cleavage of various antisense oligonucleotide-RNA duplexes. *Biochemical*

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., and Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in *Drosophila*. *Molecular Cell* 36, 445–456.

Davis, E., Caiment, F., Tordoir, X., Cavaillé, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted *Rtl1/Peg11* locus. *Curr Biol* 15, 743–749.

Dej, K.J., and Spradling, A.C. (1999). The endocycle controls nurse cell polytene chromosome structure during *Drosophila* oogenesis. *Development (Cambridge, England)* 126, 293–303.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* 432, 231–235.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA Hypothesis with Quantitative Measurements of miRNA and Target Abundance. *Molecular Cell* 54, 766–776.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336, 237–240.

Doane, W.W. (1960). Completion of meiosis in uninseminated eggs of *Drosophila melanogaster*. *Science* 132, 677–678.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. *Genes & Development* 18, 504–511.

Donis-Keller, H. (1979). Site specific enzymatic cleavage of RNA. *Nucleic Acids Research* 7, 179–192.

Driever, W., and Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in *Drosophila* embryos. *Cell* 54, 83–93.

Driever, W., Siegel, V., and Nüsslein-Volhard, C. (1990). Autonomous determination of anterior structures in the early *Drosophila* embryo by the bicoid morphogen. *Development (Cambridge, England)* 109, 811–820.

Dubnau, J., and Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. *Nature* 379, 694–699.

- Easow, G., Teleman, A.A., and Cohen, S.M. (2007). Isolation of microRNA targets by miRNP immunopurification. *Rna* 13, 1198–1204.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797.
- Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. *Elife* 5, e16955.
- Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. *Cell* 150, 100–110.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Endow, S.A., and Komma, D.J. (1997). Spindle dynamics during meiosis in *Drosophila* oocytes. *The Journal of Cell Biology* 137, 1321–1336.
- Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* 341, 1237973–1237973.
- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1.
- Etheridge, A., Lee, I., Hood, L., Galas, D., and Wang, K. (2011). Extracellular microRNA: a new source of biomarkers. *Mutat. Res.* 717, 85–90.
- Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2008). *Rna* 15, 21.
- Eulalio, A., Behm-Ansmant, I., Schweizer, D., and Izaurralde, E. (2007a). P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol. Cell. Biol.* 27, 3970–3981.
- Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. *Rna* 15, 21–32.
- Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.-F., Doerks, T., Dörner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007b). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes & Development* 21, 2558–2570.
- Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A., et al. (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Molecular Cell* 35, 868–880.
- Fasanaro, P., Greco, S., Lorenzi, M., Pescatori, M., Brioschi, M., Kulshreshtha, R., Banfi, C., Stubbs, A., Calin, G.A., Ivan, M., et al. (2009). An integrated approach for experimental target identification of hypoxia-induced miR-210. *J. Biol. Chem.* 284,

35134–35143.

Ferrandon, D., Elphick, L., Nüsslein-Volhard, C., and St Johnston, D. (1994). Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* 79, 1221–1232.

Forman, J.J., Legesse-Miller, A., and Collier, H.A. (2008). A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci USA* 105, 14879–14884.

Forrest, K.M., and Gavis, E.R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. *Curr Biol* 13, 1159–1168.

Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* 130, 287–297.

Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *Plos Biol* 3, e236.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Frohnhofer, H.G., and Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. *Nature* 324, 120–125.

Fukao, A., Mishima, Y., Takizawa, N., Oka, S., Imataka, H., Pelletier, J., Sonenberg, N., Thoma, C., and Fujiwara, T. (2014). MicroRNAs trigger dissociation of eIF4A1 and eIF4A11 from target mRNAs in humans. *Molecular Cell* 56, 79–89.

Fukaya, T., and Tomari, Y. (2011). PABP is not essential for microRNA-mediated translational repression and deadenylation in vitro. *The EMBO Journal* 30, 4998–5009.

Fukaya, T., and Tomari, Y. (2012). MicroRNAs Mediate Gene Silencing via Multiple Different Pathways in *Drosophila*. *Molecular Cell* 48, 825–836.

Fukaya, T., Iwakawa, H.-O., and Tomari, Y. (2014). MicroRNAs block assembly of eIF4F translation initiation complex in *Drosophila*. *Molecular Cell* 56, 67–78.

Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T., et al. (2007). DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat Cell Biol* 9, 604–611.

Fukunaga, R., Han, B.W., Hung, J.-H., Xu, J., Weng, Z., and Zamore, P.D. (2012). Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell* 151, 533–546.

Gamberi, C., Peterson, D.S., He, L., and Gottlieb, E. (2002). An anterior function for the *Drosophila* posterior determinant Pumilio. *Development (Cambridge, England)*

129, 2699–2710.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of *Isy-6* and other microRNAs. *Nat Struct Mol Biol* 18, 1139–1146.

Gavis, E.R., Lunsford, L., Bergsten, S.E., and Lehmann, R. (1996). A conserved 90 nucleotide element mediates translational repression of nanos RNA. *Development (Cambridge, England)* 122, 2791–2800.

Gavrilov, A., Razin, S.V., and Cavalli, G. (2015). In vivo formaldehyde cross-linking: it is time for black box analysis. *Briefings in Functional Genomics* 14, 163–165.

Gebauer, F., Preiss, T., and Hentze, M.W. (2012). From Cis-Regulatory Elements to Complex RNPs and Back. *Cold Spring Harbor Perspectives in Biology* 4, a012245–a012245.

Gebauer, F., Xu, W., Cooper, G.M., and Richter, J.D. (1994). Translational control by cytoplasmic polyadenylation of *c-mos* mRNA is necessary for oocyte maturation in the mouse. *The EMBO Journal* 13, 5712–5720.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 26, 317–325.

Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.M. (1978). Heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. *J Mol Biol* 126, 813–846.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312, 75–79.

Godt, D., and Laski, F.A. (1995). Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of *bric à brac*. *Development (Cambridge, England)* 121, 173–187.

González-Reyes, A., and St Johnston, D. (1998). Patterning of the follicle cell epithelium along the anterior-posterior axis during *Drosophila* oogenesis. *Development (Cambridge, England)* 125, 2837–2846.

González-Reyes, A., Elliott, H., and St Johnston, D. (1995). Polarization of both major body axes in *Drosophila* by *gurken*-*torpedo* signalling. *Nature* 375, 654–658.

Gottlieb, E. (1992). The 3' untranslated region of localized maternal messages contains a conserved motif involved in mRNA localization. *Proc Natl Acad Sci USA* 89, 7164–7168.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engle, P., Lim, L.P., and Bartel,

D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular Cell* 27, 91–105.

Grosswendt, S., Filipchuk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous Identification of miRNA:Target Site Interactions by Different Types of Ligation Reactions. *Molecular Cell* 1–13.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmann, P.N., and Kay, M.A. (2012). The Loop Position of shRNAs and Pre-miRNAs Is Critical for the Accuracy of Dicer Processing In Vivo. *Cell* 151, 900–911.

Gu, W., Crawford, E.D., O'Donovan, B.D., Wilson, M.R., Chow, E.D., Retallack, H., and DeRisi, J.L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* 17, 41.

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

Gutzeit, H.O. (1986). The role of microfilaments in cytoplasmic streaming in *Drosophila* follicles. *J Cell Sci* 80, 159–169.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. *Nature Publishing Group* 15, 509–524.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr, Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* 141, 129–141.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125, 887–901.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18, 3016–3027.

Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.-K., Yeom, K.-H., Yang, W.-Y., Haussler, D., Blelloch, R., and Kim, V.N. (2009). Posttranscriptional Crossregulation between Drosha and DGCR8. *Cell* 136, 75–84.

Hansen, K.D., Brenner, S.E., and Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* 38, e131–e131.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388.

Harris, M.E., and Christian, E.L. (2009). Chapter 7 -RNA Crosslinking Methods (Elsevier Inc.).

- Hassan, T., Smith, S.G.J., Gaughan, K., Oglesby, I.K., O'Neill, S., McElvaney, N.G., and Greene, C.M. (2013). Isolation and identification of cell-specific microRNAs targeting a messenger RNA using a biotinylated anti-sense oligonucleotide capture affinity technique. *Nucleic Acids Research* 41, e71–e71.
- Hatfield, S.D., Shcherbata, H.R., Fischer, K.A., Nakahara, K., Carthew, R.W., and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. *Nature* 435, 974–978.
- Hausser, J., and Zavolan, M. (2014). Identification and consequences of miRNA-target interactions--beyond repression of gene expression. *Nat. Rev. Genet.* 15, 599–612.
- Hausser, J., Syed, A.P., Bilen, B., and Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res.* 23, 604–615.
- He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., et al. (2005). A microRNA polycistron as a potential human oncogene. *Nature* 435, 828–833.
- Heifetz, Y., Yu, J., and Wolfner, M.F. (2001). Ovulation triggers activation of *Drosophila* oocytes. *Dev Biol* 234, 416–424.
- Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* 153, 654–665.
- Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *Plos Biol* 7, e1000238.
- Hentze, M.W., Caughman, S.W., Rouault, T.A., Barriocanal, J.G., Dancis, A., Harford, J.B., and Klausner, R.D. (1987). Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* 238, 1570–1573.
- Herbert, K.M., Pimienta, G., DeGregorio, S.J., Alexandrov, A., and Steitz, J.A. (2013). Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile. *Cell Rep* 5, 1070–1081.
- Herschlag, D. (1991). Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules in vivo: more isn't always better. *Proceedings of the National Academy of Sciences* 88, 6921–6925.
- Hindson, C.M., Chevillet, J.R., Briggs, H.A., Gallichotte, E.N., Ruf, I.K., Hindson, B.J., Vessella, R.L., and Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* 10, 1003–1005.
- Ho, J.J.D., and Marsden, P.A. (2014). Competition and collaboration between RNA-binding proteins and microRNAs. *WIREs RNA* 5, 69–86.
- Ho, S.P., Bao, Y., Leshner, T., Malhotra, R., Ma, L.Y., Fluharty, S.J., and Sakai, R.R. (1998). Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. *Nat Biotechnol* 16, 59–63.

- Hofacker, I.L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research* 31, 3429–3431.
- Horner, V.L., and Wolfner, M.F. (2008). Mechanical stimulation by osmotic and hydrostatic pressure activates *Drosophila* oocytes in vitro in a calcium-dependent manner. *Dev Biol* 316, 100–109.
- Horner, V.L., Czank, A., Jang, J.K., Singh, N., Williams, B.C., Puro, J., Kubli, E., Hanes, S.D., McKim, K.S., Wolfner, M.F., et al. (2006). The *Drosophila* calpripresin sarah is required for several aspects of egg activation. *Curr Biol* 16, 1441–1446.
- Hsu, R.-J., Yang, H.-J., and Tsai, H.-J. (2009). Labeled microRNA pull-down assay system: an experimental approach for high-throughput identification of microRNA-target mRNAs. *Nucleic Acids Research* 37, e77–e77.
- Hu, H.Y., Yan, Z., Xu, Y., Hu, H., Menzel, C., Zhou, Y.H., Chen, W., and Khaitovich, P. (2009). Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* 10, 413.
- Humphreys, D.T., Westman, B.J., Martin, D.I.K., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences* 102, 16961–16966.
- Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* 12, 99–110.
- Huntzinger, E., Braun, J.E., Heimstädt, S., Zekri, L., and Izaurralde, E. (2010). Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. *The EMBO Journal* 29, 4146–4160.
- Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* 65, 274–287.
- Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838.
- Huynh, J.-R., and St Johnston, D. (2004). The origin of asymmetry: early polarisation of the *Drosophila* germline cyst and oocyte. *Curr Biol* 14, R438–R449.
- Hwang, S.P., Eisenberg, M., Binder, R., Shelness, G.S., and Williams, D.L. (1989). Predicted structures of apolipoprotein II mRNA constrained by nuclease and dimethyl sulfate reactivity: stable secondary structures occur predominantly in local domains via intraexonic base pairing. *Journal of Biological Chemistry* 264, 8410–8418.
- Ikeda, S., Kong, S.W., Lu, J., Bisping, E., Zhang, H., Allen, P.D., Golub, T.R., Pieske, B., and Pu, W.T. (2007). Altered microRNA expression in human heart disease. *Physiol. Genomics* 31, 367–373.
- Imig, J., Brunschweiler, A., Brümmer, A., Guennewig, B., Mittal, N., Kishore, S., Tsikrika, P., Gerber, A.P., Zavolan, M., and Hall, J. (2015). miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction. *Nature Chemical*

Biology 11, 107–114.

Inoue, H., Hayase, Y., Iwai, S., and Ohtsuka, E. (1987). Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *Febs Letters* 215, 327–330.

Iovino, N., Pane, A., and Gaul, U. (2009). miR-184 Has Multiple Roles in *Drosophila* Female Germline Development. *Developmental Cell* 17, 123–133.

Iwakawa, H.-O., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol* 25, 651–665.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Publishing Group* 11, 113–127.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318–356.

Jankowsky, E., and Harris, M.E. (2015). Specificity and nonspecificity in RNA–protein interactions. *Nat Rev Mol Cell Biol* 16, 533–544.

Jansen, R.P. (2001). mRNA localization: message on the move. *Nat Rev Mol Cell Biol* 2, 247–256.

Janssen, H.L.A., Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patick, A.K., Chen, A., Zhou, Y., et al. (2013). Treatment of HCV Infection by Targeting MicroRNA. *N Engl J Med* 368, 1685–1694.

Januschke, J., Gervais, L., Dass, S., Kaltschmidt, J.A., Lopez-Schier, H., St Johnston, D., Brand, A.H., Roth, S., and Guichet, A. (2002). Polar transport in the *Drosophila* oocyte requires Dynein and Kinesin I cooperation. *Curr Biol* 12, 1971–1981.

Januschke, J., Gervais, L., Gillet, L., Keryer, G., Bornens, M., and Guichet, A. (2006). The centrosome-nucleus complex and microtubule organization in the *Drosophila* oocyte. *Development (Cambridge, England)* 133, 129–139.

Jens, M., and Rajewsky, N. (2015). Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* 16, 113–126.

Jin, Z., and Xie, T. (2007). Dcr-1 maintains *Drosophila* ovarian stem cells. *Curr Biol* 17, 539–544.

Jinek, M., Fabian, M.R., Coyle, S.M., Sonenberg, N., and Doudna, J.A. (2010). Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nat Struct Mol Biol* 17, 238–240.

Jovanovic, M., Reiter, L., Clark, A., Weiss, M., Picotti, P., Rehrauer, H., Frei, A., Neukomm, L.J., Kaufman, E., Wollscheid, B., et al. (2012). RIP-chip-SRM—a new combinatorial large-scale approach identifies a set of translationally regulated bantam/miR-58 targets in *C. elegans*. *Genome Res.* 22, 1360–1371.

Juge, F., Zaessinger, S., Temme, C., Wahle, E., and Simonelig, M. (2002). Control of

poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in *Drosophila*. *The EMBO Journal* **21**, 6603–6613.

Kadener, S., Rodriguez, J., Abruzzi, K.C., Khodor, Y.L., Sugino, K., Marr, M.T., Nelson, S., and Rosbash, M. (2009). Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. *Rna* **15**, 537–545.

Kalifa, Y., Huang, T., Rosen, L.N., Chatterjee, S., and Gavis, E.R. (2006). Glorund, a *Drosophila* hnRNP F/H homolog, is an ovarian repressor of nanos translation. *Developmental Cell* **10**, 291–301.

Karolchik, D., Hinrichs, A.S., and Kent, W.J. (2011). The UCSC Genome Browser. *Curr Protoc Hum Genet Chapter 18*, Unit18.6–18.6.33.

Kawamata, T., Seitz, H., and Tomari, Y. (2009). Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* **16**, 953–960.

Kedde, M., and Agami, R. (2008). Interplay between microRNAs and RNA-binding proteins determines developmental processes. *Cell Cycle* **7**, 899–903.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A.F., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Ørom, U.A., et al. (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273–1286.

Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J.A.F., Elkon, R., and Agami, R. (2010). A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat Cell Biol* **12**, 1014–1020.

Kelley, R.L. (1993). Initial organization of the *Drosophila* dorsoventral axis depends on an RNA-binding protein encoded by the squid gene. *Genes & Development* **7**, 948–960.

Kemény-Beke, A., Berényi, E., Facskó, A., Damjanovich, J., Horváth, A., Bodnár, A., Berta, A., and Aradi, J. (2006). Antiproliferative effect of 4-thiouridylate on OCM-1 uveal melanoma cells. *Eur J Ophthalmol* **16**, 680–685.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics* **39**, 1278–1284.

Khan, A.A., Betel, D., Miller, M.L., Sander, C., Leslie, C.S., and Marks, D.S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* **27**, 549–555.

Khan, I.M., and Coulson, J.M. (1993). A novel method to stabilise antisense oligonucleotides against exonuclease degradation. *Nucleic Acids Research* **21**, 4433.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216.

Kim, H.H., Kuwano, Y., Srikantan, S., Lee, E.-K., Martindale, J.L., and Gorospe, M. (2009a). HuR recruits let-7/RISC to repress c-Myc expression. *Genes & Development* **23**, 1743–1748.

- Kim, V.N., Han, J., and Siomi, M.C. (2009b). Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 10, 126–139.
- Kim-Ha, J., Kerr, K., and Macdonald, P.M. (1995). Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. *Cell* 81, 403–412.
- King, R.C., Aggarwal, S.K., and Aggarwal, U. (1968). The development of the female *Drosophila* reproductive system. *J. Morphol.* 124, 143–166.
- Kloosterman, W.P., Wienholds, E., Ketting, R.F., and Plasterk, R.H.A. (2004). Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Research* 32, 6284–6291.
- Klusza, S., and Deng, W.-M. (2011). At the crossroads of differentiation and proliferation: precise control of cell-cycle changes by multiple signaling pathways in *Drosophila* follicle cells. *Bioessays* 33, 124–134.
- Kole, R., Krainer, A.R., and Altman, S. (2012). RNA therapeutics: beyond RNA interference and antisense oligonucleotides. 1–16.
- Kozak, M. (1984). Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Research* 12, 3873–3893.
- Kozlov, G., Safaee, N., Rosenauer, A., and Gehring, K. (2010). Structural basis of binding of P-body-associated proteins GW182 and ataxin-2 by the Mlle domain of poly(A)-binding protein. *J. Biol. Chem.* 285, 13599–13606.
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* 17, 909–915.
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2011). iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* e2638–e2638.
- Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2012). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493, 231–235.
- Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*
- Kudla, G., Granneman, S., Hahn, D., Beggs, J.D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci USA* 108, 10010–10015.
- Kudlow, B.A., Zhang, L., and Han, M. (2012). Systematic analysis of tissue-restricted miRISCs reveals a broad role for microRNAs in suppressing basal activity of the *C. elegans* pathogen response. *Molecular Cell* 46, 530–541.
- Kugler, J.-M., and Lasko, P. (2012). Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during *Drosophila* oogenesis. 1–14.

- Kugler, J.-M., Verma, P., Chen, Y.-W., Weng, R., and Cohen, S.M. (2013). miR-989 is required for border cell migration in the *Drosophila* ovary. *PLoS ONE* 8, e67075.
- Kuhn, D.E., Martin, M.M., Feldman, D.S., Terry, A.V., Nuovo, G.J., and Elton, T.S. (2008). Experimental validation of miRNA targets. *Methods* 44, 47–54.
- Kundu, P., Fabian, M.R., Sonenberg, N., Bhattacharyya, S.N., and Filipowicz, W. (2012). HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Research* 40, 5088–5100.
- Kurreck, J., Wyszko, E., Gillen, C., and Erdmann, V.A. (2002). Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Research* 30, 1911–1918.
- Kuzuoglu-Öztürk, D., Huntzinger, E., Schmidt, S., and Izaurralde, E. (2012). The *Caenorhabditis elegans* GW182 protein AIN-1 interacts with PAB-1 and subunits of the PAN2-PAN3 and CCR4-NOT deadenylase complexes. *Nucleic Acids Research* 40, 5651–5665.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858.
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. *Rna* 9, 175–179.
- Lai, E.C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genetics* 30, 363–364.
- Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. (2003). Computational identification of *Drosophila* microRNA genes. *Genome Biol.* 4, R42.
- Lal, A., Thomas, M.P., Altschuler, G., Navarro, F., O'Day, E., Li, X.L., Concepcion, C., Han, Y.-C., Thiery, J., Rajani, D.K., et al. (2011). Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet.* 7, e1002363.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr Biol* 14, 2162–2167.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.
- Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862.
- Lazar, E., Jacob, M., Krol, A., and Branlant, C. (1982). Accessibility of U1 RNA to base pairing with a single-stranded DNA fragment mimicking the intron extremities at the splice junction. *Nucleic Acids Research* 10, 1193–1201.

- Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Molecular Cell* 43, 340–352.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854.
- Lee, T., Wang, N., Houel, S., Coutts, K., Old, W., and Ahn, N. (2015). Dosage and temporal thresholds in microRNA proteomics. *Mol. Cell Proteomics* 14, 289–302.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419.
- Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO Journal* 21, 4663–4670.
- Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal* 23, 4051–4060.
- Leung, A.K.L., Young, A.G., Bhutkar, A., Zheng, G.X., Bosson, A.D., Nielsen, C.B., and Sharp, P.A. (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* 18, 237–244.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* 120, 15–20.
- Lewis, B.P., Shih, I.-H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787–798.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, P., Peng, J., Hu, J., Xu, Z., Xie, W., and Yuan, L. (2011). Localized expression pattern of miR-184 in *Drosophila*. *Mol Biol Rep* 38, 355–358.
- Li, S., Zhu, J., Fu, H., Wan, J., Hu, Z., Liu, S., Li, J., Tie, Y., Xing, R., Zhu, J., et al. (2012). Hepato-specific microRNA-122 facilitates accumulation of newly synthesized miRNA through regulating PRKRA. *Nucleic Acids Research* 40, 884–891.
- Li, Y., and Zhang, Z. (2015). Computational Biology in microRNA. *WIREs RNA* 6, 435–452.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Lieberfarb, M.E., Chu, T., Wreden, C., Theurkauf, W., Gergen, J.P., and Strickland, S. (1996). Mutations that perturb poly(A)-dependent maternal mRNA activation block

the initiation of development. *Development (Cambridge, England)* 122, 579–588.

Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of *Caenorhabditis elegans*. *Genes & Development* 17, 991–1008.

Lima, W.F., Brown-Driver, V., Fox, M., Hanecak, R., and Bruce, T.W. (1997). Combinatorial screening and rational optimization for hybridization to folded hepatitis C virus RNA of oligonucleotides with biological antisense activity. *J. Biol. Chem.* 272, 626–638.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* 7, 723.

Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. (2009). C3PO, an Endoribonuclease That Promotes RNAi by Facilitating RISC Activation. *Science* 325, 750–753.

Loeb, G.B., Khan, A.A., Canner, D., Hiatt, J.B., Shendure, J., Darnell, R.B., Leslie, C.S., and Rudensky, A.Y. (2012). Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. *Molecular Cell* 48, 760–770.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* 303, 95–98.

Ma, H., Wu, Y., Choi, J.-G., and Wu, H. (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc Natl Acad Sci USA* 110, 20687–20692.

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

Ma, J., He, F., Xie, G., and Deng, W.-M. (2016). Maternal AP determinants in the *Drosophila* oocyte and embryo. *Wiley Interdiscip Rev Dev Biol*.

Macdonald, P.M., and Kerr, K. (1997). Redundant RNA recognition events in bicoid mRNA localization. *Rna* 3, 1413–1420.

Macdonald, P.M., and Kerr, K. (1998). Mutational analysis of an RNA recognition

element that mediates localization of bicoid mRNA. *Mol. Cell. Biol.* **18**, 3788–3795.

Macdonald, P.M., and Struhl, G. (1988). cis-acting sequences responsible for anterior localization of bicoid mRNA in *Drosophila* embryos. *Nature* **336**, 595–598.

Macdonald, P.M., Kerr, K., Smith, J.L., and Leask, A. (1993). RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development (Cambridge, England)* **118**, 1233–1243.

Macdonald, P.M., Leask, A., and Kerr, K. (1995). exl protein specifically binds BLE1, a bicoid mRNA localization element, and is required for one phase of its activity. *Proceedings of the National Academy of Sciences* **92**, 10787–10791.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198.

Mahowald, A.P., Goralski, T.J., and Caulton, J.H. (1983). In vitro activation of *Drosophila* eggs. *Dev Biol* **98**, 437–445.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**, 522–535.

Maragkakis, M., Reczko, M., Simossis, V.A., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., et al. (2009). DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Research* **37**, W273–W276.

Maroni, G. (1994). The organization of *Drosophila* genes. *DNA Sequence*.

Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S., et al. (2007). MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* **317**, 1764–1767.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Molecular Cell* **54**, 751–765.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620.

McLaughlin, J.M., and Bratu, D.P. (2015). *Drosophila melanogaster* Oogenesis: An Overview. In *Drosophila Oogenesis*, D.P. Bratu, and G.P. McNeil, eds. (New York, NY: Springer New York), pp. 1–20.

Meier, J., Hovestadt, V., Zapatka, M., Pscherer, A., Lichter, P., and Seiffert, M. (2013). Genome-wide identification of translationally inhibited and degraded miR-155 targets using RNA-interacting protein-IP. *Rnabiology* **10**, 1018–1029.

Meisner, N.-C., and Filipowicz, W. (2011). Properties of the Regulatory RNA-Binding

- Protein HuR and its Role in Controlling miRNA Repression. *Adv. Exp. Med. Biol.* **700**, 106–123.
- Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* **14**, 447–459.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338.
- Mendell, J.R., Rodino-Klapac, L.R., Sahenk, Z., Roush, K., Bird, L., Lowes, L.P., Alfano, L., Gomez, A.M., Lewis, S., Kota, J., et al. (2013). Eteplirsen for the treatment of Duchenne muscular dystrophy. *Ann. Neurol.* **74**, 637–647.
- Miles, W.O., Tschöp, K., Herr, A., Ji, J.-Y., and Dyson, N.J. (2012). Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes & Development* **26**, 356–368.
- Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *Rna* **10**, 1692–1694.
- Mishima, Y., Fukao, A., Kishimoto, T., Sakamoto, H., Fujiwara, T., and Inoue, K. (2012). Translational inhibition by deadenylation-independent mechanisms is central to microRNA-mediated silencing in zebrafish. *Proc Natl Acad Sci USA* **109**, 1104–1109.
- Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Curr Biol* **16**, 2135–2142.
- Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* **3**, e215.
- Monteys, A.M., Spengler, R.M., Wan, J., Tecedor, L., Lennox, K.A., Xing, Y., and Davidson, B.L. (2010). Structure and activity of putative intronic miRNA promoters. *Rna* **16**, 495–505.
- Moretti, F., Kaiser, C., Zdanowicz-Specht, A., and Hentze, M.W. (2012). PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. *Nat Struct Mol Biol* **19**, 603–608.
- Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* **15**, 902–909.
- Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Tuschl, T., Ohler, U., and Keene, J.D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Molecular Cell* **43**, 327–339.
- Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. *Nature* **486**, 368–374.

- Nam, J.-W., Rissland, O.S., Koppstein, D., Abreu-Goodger, C., Jan, C.H., Agarwal, V., Yildirim, M.A., Rodriguez, A., and Bartel, D.P. (2014). Global Analyses of the Effect of Different Cellular Contexts on MicroRNA Targeting. *Molecular Cell* 53, 1031–1043.
- Nelles, D.A., Fang, M.Y., Aigner, S., and Yeo, G.W. (2015). Applications of Cas9 as an RNA-programmed RNA-binding protein. *Bioessays* 37, 732–9.
- Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* 165, 488-96.
- Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *The EMBO Journal* 23, 150–159.
- Neuman-Silberberg, F.S., and Schupbach, T. (1993). The Drosophila dorsoventral patterning gene *gurken* produces a dorsally localized RNA and encodes a TGF alpha-like protein. *Cell* 75, 165–174.
- Neumüller, R.A., Betschinger, J., Fischer, A., Bushati, N., Poernbacher, I., Mechtler, K., Cohen, S.M., and Knoblich, J.A. (2008). Mei-P26 regulates microRNAs and cell growth in the Drosophila ovarian stem cell lineage. *Nature* 454, 241–245.
- Ngan K Voa, Daltona, R.P., Liub, N., Olsonb, E.N., and Goodman, R.H. (2010). Affinity purification of microRNA-133a with the cardiac transcription factor, Hand2. *Proc Natl Acad Sci USA* 107, 19231-6
- Nonne, N., Ameyar-Zazoua, M., Souidi, M., and Harel-Bellan, A. (2010). Tandem affinity purification of miRNA target mRNAs (TAP-Tar). *Nucleic Acids Research* 38, e20–e20.
- Nordman, J., and Orr-Weaver, T.L. (2012). Regulation of DNA replication during development. *Development (Cambridge, England)* 139, 455–464.
- Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* 121, 1005–1016.
- Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* 287, 795–801.
- O'Neill, L.A., Sheedy, F.J., and McCoy, C.E. (2011). MicroRNAs: the fine-tuners of Toll-like receptor signalling. *Nat Rev Immunol* 11, 163–175.
- Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* 216, 671–680.
- Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S., and Fisher, D.E. (2008). Chromatin structure analyses identify miRNA promoters. *Genes & Development* 22, 3172–3183.
- O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and

- Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 1–15.
- Page, A.W., and Orr-Weaver, T.L. (1996). The *Drosophila* genes *grauzone* and *cortex* are necessary for proper female meiosis. *J Cell Sci* 109 (Pt 7), 1707–1715.
- Page, A.W., and Orr-Weaver, T.L. (1997). Activation of the meiotic divisions in *Drosophila* oocytes. *Dev Biol* 183, 195–207.
- Pallan, P.S., and Egli, M. (2008). Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme-dsDNA complex. *Cell Cycle* 7, 2562–2569.
- Park, C.Y., Jeker, L.T., Carver-Moore, K., Oh, A., Liu, H.J., Cameron, R., Richards, H., Li, Z., Adler, D., Yoshinaga, Y., et al. (2012). A Resource for the Conditional Ablation of microRNAs in the Mouse. *Cell Rep* 1, 385–391.
- Park, J.-E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475, 201–205.
- Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* 434, 663–666.
- Pashev, I.G., Dimitrov, S.I., and Angelov, D. (1991). Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* 16, 323–326.
- Pawlicki, J.M., and Steitz, J.A. (2008). Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *The Journal of Cell Biology* 182, 61–76.
- Payre, F., Crozatier, M., and Vincent, A. (1994). Direct control of transcription of the *Drosophila* morphogen *bicoid* by the serendipity delta zinc finger protein, as revealed by in vivo analysis of a finger swap. *Genes & Development* 8, 2718–2728.
- Peng, J., Wang, C., Wan, C., Zhang, D., Li, W., Li, P., Kong, Y., and Yuan, L. (2015). miR-184 is Critical for the motility-related PNS development in *Drosophila*. *Int. J. Dev. Neurosci.* 46, 100–107.
- Petersen, C.P., Bordeleau, M.-E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell* 21, 533–542.
- Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grässer, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., et al. (2005). Identification of microRNAs of the herpesvirus family. *Nat Methods* 2, 269–276.
- Piao, X., Zhang, X., Wu, L., and Belasco, J.G. (2010). CCR4-NOT deadenylates mRNA associated with RNA-induced silencing complexes in human cells. *Mol. Cell Biol.* 30, 1486–1494.
- Picelli, S., Faridani, O.R., rklund, A.S.K.B.O., Winberg, G.O.S., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* 9, 171–181.

- Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* 309, 1573–1576.
- Pokrywka, N.J., Zhang, H., and Raley-Susman, K. (2014). Distinct roles for hu li tai shao and swallow in cytoskeletal organization during *Drosophila* oogenesis. *Developmental Dynamics* 243, 906–916.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Raal, F.J., Santos, R.D., Blom, D.J., Marais, A.D., Charng, M.-J., Cromwell, W.C., Lachmann, R.H., Gaudet, D., Tan, J.L., Chasan-Taber, S., et al. (2010). Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. *Lancet* 375, 998–1006.
- Rajewsky, N. (2006). microRNA target predictions in animals. *Nature Genetics* 38 *Suppl*, S8–S13.
- Ramos, A., Grünert, S., Adams, J., Micklem, D.R., Proctor, M.R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000). RNA recognition by a Staufen double-stranded RNA-binding domain. *The EMBO Journal* 19, 997–1009.
- Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *Rna* 11, 1640–1647.
- Reich, J., Snee, M.J., and Macdonald, P.M. (2009). miRNA-Dependent Translational Repression in the *Drosophila* Ovary. *PLoS ONE* 4, e4669.
- Reilly, S.N., Liu, X., Carnicer, R., Recalde, A., Muszkiewicz, A., Jayaram, R., Carena, M.C., Wijesurendra, R., Stefanini, M., Surdo, N.C., et al. (2016). Up-regulation of miR-31 in human atrial fibrillation begets the arrhythmia by depleting dystrophin and neuronal nitric oxide synthase. *Sci Transl Med* 8, 340ra74–340ra74.
- Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. *Genes & Development* 16, 1616–1626.
- Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. *Cell* 110, 513–520.
- Richter, J.D. (1991). Translational control during early development. *Bioessays* 13, 179–183.
- Riechmann, V., and Ephrussi, A. (2004). Par-1 regulates bicoid mRNA localisation by phosphorylating Exuperantia. *Development (Cambridge, England)* 131, 5897–5907.
- Riley, K.J., Yario, T.A., and Steitz, J.A. (2012). Association of Argonaute proteins and microRNAs can occur after cell lysis. *Rna* 18, 1581–1585.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat Biotechnol* 29, 24–26.

- Rogers, K., and Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* 25, 2383–2399.
- Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature* 467, 1128–1132.
- Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* 17, 1850–1864.
- Rukov, J.L., Hagedorn, P.H., Høy, I.B., Feng, Y., Lindow, M., and Vinther, J. (2015). Dissecting the target specificity of RNase H recruiting oligonucleotides using massively parallel reporter analysis of short RNA motifs. *Nucleic Acids Research* 43, 8476–8487.
- Ruohola, H., Bremer, K.A., Baker, D., Swedlow, J.R., Jan, L.Y., and Jan, Y.N. (1991). Role of neurogenic genes in establishment of follicle cell fate and oocyte polarity during oogenesis in *Drosophila*. *Cell* 66, 433–449.
- Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* 38, 317–331.
- Saito, Y., Liang, G., Egger, G., Friedman, J.M., Chuang, J.C., Coetzee, G.A., and Jones, P.A. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9, 435–443.
- Sallés, F.J., Lieberfarb, M.E., Wreden, C., Gergen, J.P., and Strickland, S. (1994). Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* 266, 1996–1999.
- Salomon, W.E., Jolly, S.M., Moore, M.J., Zamore, P.D., and Serebrov, V. (2015). Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* 162, 84–95.
- Sander, K., and Lehmann, R. (1988). *Drosophila* nurse cells produce a posterior signal required for embryonic segmentation and polarity. *Nature* 335, 68–70.
- Sandmann, T., Jakobsen, J.S., and Furlong, E.E.M. (2007). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nature Protocols* 1, 2839–2855.
- Scherr, M., and Rossi, J.J. (1998). Rapid determination and quantitation of the accessibility to native RNAs by antisense oligodeoxynucleotides in murine cell extracts. *Nucleic Acids Research* 26, 5079–5085.
- Schirle, N.T., and Macrae, I.J. (2012). The crystal structure of human Argonaute2. *Science* 336, 1037–1040.
- Schnorrer, F., Bohmann, K., and Nüsslein-Volhard, C. (2000). The molecular motor dynein is involved in targeting swallow and bicoid RNA to the anterior pole of *Drosophila* oocytes. *Nat Cell Biol* 2, 185–190.

- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199–208.
- Scott, G.K., Mattie, M.D., Berger, C.E., Benz, S.C., and Benz, C.C. (2006). Rapid alteration of microRNA levels by histone deacetylase inhibition. *Cancer Res.* 66, 1277–1281.
- Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev Biol* 243, 215–225.
- Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63.
- Sen, G.L., and Blau, H.M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature Publishing Group* 7, 633–636.
- Shcherbata, H.R., Ward, E.J., Fischer, K.A., Yu, J.-Y., Reynolds, S.H., Chen, C.-H., Xu, P., Hay, B.A., and Ruohola-Baker, H. (2007). Stage-specific differences in the requirements for germline stem cell maintenance in the *Drosophila* ovary. *Cell Stem Cell* 1, 698–709.
- Sheets, M.D., Wu, M., and Wickens, M. (1995). Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature* 374, 511–516.
- Shi, M., Han, W., and Spivack, S.D. (2013). Accepted Manuscript. *Analytical Biochemistry* 1.
- Simon, M.D. (2013). Capture hybridization analysis of RNA targets (CHART). *Curr Protoc Mol Biol Chapter 21*, Unit21.25.
- Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* 504, 465–469.
- Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. *Pnas* 108, 20497–20502.
- Slack, J. (2014). Establishment of spatial pattern. *Wiley Interdiscip Rev Dev Biol* 3, 379–388.
- Smibert, C.A., Wilson, J.E., Kerr, K., and Macdonald, P.M. (1996). *smaug* protein represses translation of unlocalized nanos mRNA in the *Drosophila* embryo. *Genes & Development* 10, 2600–2609.
- Snee, M., Benz, D., Jen, J., and Macdonald, P.M. (2014). Two distinct domains of Bruno bind specifically to the *oskarmRNA*. *Rnabiology* 5, 49–57.

- Song, J.-J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305, 1434–1437.
- Spradling, A.C. (1993). Developmental genetics of oogenesis. *The Development of Drosophila Melanogaster*. 1–70.
- St Johnston, D., Beuchle, D., and Nüsslein-Volhard, C. (1991). Staufén, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell* 66, 51–63.
- St Johnston, D., Driever, W., Berleth, T., Richstein, S., and Nüsslein-Volhard, C. (1989). Multiple steps in the localization of bicoid RNA to the anterior pole of the *Drosophila* oocyte. *Development (Cambridge, England)* 107 Suppl, 13–19.
- Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of *Drosophila* MicroRNA targets. *Plos Biol* 1, E60.
- Stein, H., and Hausen, P. (1969). Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. *Science* 166, 393–395.
- Steinkraus, B.R., Toegel, M., and Fulga, T.A. (2016). Tiny giants of gene regulation: experimental strategies for microRNA functional studies. *Wiley Interdiscip Rev Dev Biol* 5, 311–362.
- Stetina, Von, J.R., and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. *Cold Spring Harbor Perspectives in Biology* 3, a005553.
- Struhl, G., Struhl, K., and Macdonald, P.M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* 57, 1259–1273.
- Su, H., Trombly, M.I., Chen, J., and Wang, X. (2009). Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & Development* 23, 304–317.
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71.
- Sullivan, D.E., Brisson, N., and Verma, D.P. (1980). Reverse transcription of 25S soybean ribosomal RNA in the absence of exogenous primer. *Biochem. Biophys. Res. Commun.* 94, 144–150.
- Sysoev, V.O., Fischer, B., Frese, C.K., Gupta, I., Krijgsveld, J., Hentze, M.W., Castello, A., and Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*. *Nat Comms* 7, 12128.
- Taft, R.J., Pheasant, M., and Mattick, J.S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29, 288–299.
- Takeo, S., Tsuda, M., Akahori, S., Matsuo, T., and Aigaki, T. (2006). The calcineurin regulator sra plays an essential role in female meiosis in *Drosophila*. *Curr Biol* 16, 1435–1440.

- Tan, J.Y., Vance, K.W., Varela, M.A., Sirey, T., Watson, L.M., Curtis, H.J., Marinello, M., Alves, S., Steinkraus, B.R., Cooper, S., et al. (2014). Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7. *Nat Struct Mol Biol* 21, 955–961.
- Tang, X., Li, M., Tucker, L., and Ramratnam, B. (2011). Glycogen synthase kinase 3 beta (GSK3 β) phosphorylates the RNAase III enzyme Drosha at S300 and S302. *PLoS ONE* 6, e20391.
- Tang, X., Zhang, Y., Tucker, L., and Ramratnam, B. (2010). Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. *Nucleic Acids Research* 38, 6610–6619.
- Theurkauf, W.E., Smiley, S., Wong, M.L., and Alberts, B.M. (1992). Reorganization of the cytoskeleton during *Drosophila* oogenesis: implications for axis specification and intercellular transport. *Development (Cambridge, England)* 115, 923–936.
- Thum, T., Catalucci, D., and Bauersachs, J. (2008). MicroRNAs: novel regulators in cardiac development and disease. *Cardiovasc. Res.* 79, 562–570.
- Tian, Y., Simanshu, D.K., Ma, J.-B., Park, J.-E., Heo, I., Kim, V.N., and Patel, D.J. (2014). A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. *Molecular Cell* 53, 606–616.
- Tokumaru, S., Suzuki, M., Yamada, H., Nagino, M., and Takahashi, T. (2008). let-7 regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis* 29, 2073–2077.
- Toth, J., and Biggin, M.D. (2000). The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in *drosophila* embryos. *Nucleic Acids Research* 28, e4.
- Truesdell, S.S., Mortensen, R.D., Seo, M., Schroeder, J.C., Lee, J.H., LeTonqueze, O., and Vasudevan, S. (2012). MicroRNA-mediated mRNA translation activation in quiescent cells and oocytes involves recruitment of a nuclear microRNP. *Sci Rep* 2, 842.
- Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., and Tomari, Y. (2011). Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nat Struct Mol Biol* 18, 1153–1158.
- Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell* 147, 1537–1550.
- van Gurp, T.P., McIntyre, L.M., and Verhoeven, K.J.F. (2013). Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS ONE* 8, e85583.
- van Kouwenhove, M., Kedde, M., and Agami, R. (2011). MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat. Rev. Cancer* 11, 644–656.
- Vance, K.W., Sansom, S.N., Lee, S., Chalei, V., Kong, L., Cooper, S.E., Oliver, P.L., and Ponting, C.P. (2014). The long non-coding RNA Paupar regulates the expression of both local and distal genes. *The EMBO Journal* 33, 296–311.

- Vassalli, J.D., Huarte, J., Belin, D., Gubler, P., Vassalli, A., O'Connell, M.L., Parton, L.A., Rickles, R.J., and Strickland, S. (1989). Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes & Development* 3, 2163–2171.
- Vasudevan, S. (2012). Posttranscriptional upregulation by microRNAs. *WIREs RNA* 3, 311–330.
- Vasudevan, S., Tong, Y., and Steitz, J.A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318, 1931–1934.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351.
- Verrotti, A.C., Wreden, C., and Strickland, S. (1999). Dissociation of mRNA cytoplasmic polyadenylation from translational activation by structural modification of the 5'-UTR. *Nucleic Acids Research* 27, 3417–3423.
- Vo, N.K., Dalton, R.P., Liu, N., Olson, E.N., and Goodman, R.H. (2010). Affinity purification of microRNA-133a with the cardiac transcription factor, Hand2. *Proceedings of the National Academy of Sciences* 107, 19231–19236.
- Voit, T., Topaloglu, H., Straub, V., Muntoni, F., Deconinck, N., Campion, G., De Kimpe, S.J., Eagle, M., Guglieri, M., Hood, S., et al. (2014). Safety and efficacy of drisapersen for the treatment of Duchenne muscular dystrophy (DEMAND II): an exploratory, randomised, placebo-controlled phase 2 study. *Lancet Neurol* 13, 987–996.
- Wada, T., Kikuchi, J., and Furukawa, Y. (2012). Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8. *EMBO Rep* 13, 142–149.
- Wagner, C., Palacios, I., Jaeger, L., St Johnston, D., Ehresmann, B., Ehresmann, C., and Brunel, C. (2001). Dimerization of the 3'UTR of bicoid mRNA involves a two-step mechanism. *J Mol Biol* 313, 511–524.
- Wagner, R.W., Matteucci, M.D., Grant, D., Huang, T., and Froehler, B.C. (1996). Potent and selective inhibition of gene expression by an antisense heptanucleotide. *Nat Biotechnol* 14, 840–844.
- Wahlestedt, C., Salmi, P., Good, L., Kela, J., Johnsson, T., Hökfelt, T., Broberger, C., Porreca, F., Lai, J., Ren, K., et al. (2000). Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proceedings of the National Academy of Sciences* 97, 5633–5638.
- Wakiyama, M., Takimoto, K., Ohara, O., and Yokoyama, S. (2007). Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes & Development* 21, 1857–1862.
- Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., et al. (2015). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709.
- Wang, B., Love, T.M., Call, M.E., Doench, J.G., and Novina, C.D. (2006).

Recapitulation of short RNA-directed translational gene silencing in vitro. *Molecular Cell* 22, 553–560.

Webster, P.J., Liang, L., Berg, C.A., Lasko, P., and Macdonald, P.M. (1997). Translational repressor bruno plays multiple roles in development and is widely conserved. *Genes & Development* 11, 2510–2521.

Wee, L.M., Flores-Jasso, C.F., Salomon, W.E., and Zamore, P.D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151, 1055–1067.

Weil, T.T. (2014). mRNA localization in the *Drosophila* germline. *Rnabiology*.

Weil, T.T., Forrest, K.M., and Gavis, E.R. (2006). Localization of bicoid mRNA in Late Oocytes Is Maintained by Continual Active Transport. *Developmental Cell* 11, 251–262.

Weil, T.T., Parton, R.M., Herpers, B., Soetaert, J., Veenendaal, T., Xanthakis, D., Dobbie, I.M., Halstead, J.M., Hayashi, R., Rabouille, C., et al. (2012). *Drosophila* patterning is established by differential association of mRNAs with P bodies. *Nat Cell Biol* 14, 1305–1313.

Weil, T.T., Parton, R., Davis, I., and Gavis, E.R. (2008). Changes in bicoid mRNA Anchoring Highlight Conserved Mechanisms during the Oocyte-to-Embryo Transition. *Current Biology* 18, 1055–1061.

Weil, T.T., Xanthakis, D., Parton, R., Dobbie, I., Rabouille, C., Gavis, E.R., and Davis, I. (2010). Distinguishing direct from indirect roles for bicoid mRNA localization factors. *Development (Cambridge, England)* 137, 169–176.

Wharton, R.P. (1992). Regulated expression from maternal mRNAs in *Drosophila*. *Semin Dev Biol* 391–397.

Wharton, R.P., and Struhl, G. (1989). Structure of the *Drosophila* BicaudalD protein and its role in localizing the the posterior determinant nanos. *Cell* 59, 881–892.

Wharton, R.P., and Struhl, G. (1991). RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen nanos. *Cell* 67, 955–967.

Wickersheim, M.L., and Blumenstiel, J.P. (2013). Terminator oligo blocking efficiently eliminates rRNA from *Drosophila* small RNA sequencing libraries. *Biotech.* 55, 269–272.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862.

Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 11, 228–234.

Wissink, E.M., Fogarty, E.A., and Grimson, A. (2016). High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics* 17, 177.

- Wolter, J.M., Kotagama, K., Babb, C.S., and Mangone, M. (2015). Detection of miRNA Targets in High-throughput Using the 3'LIFE Assay. *J Vis Exp* e52647–e52647.
- Woolf, T.M., Melton, D.A., and Jennings, C.G. (1992). Specificity of antisense oligonucleotides in vivo. *Proceedings of the National Academy of Sciences* 89, 7305–7309.
- Worley-Morse, T.O., and Gunsch, C.K. (2015). A computational analysis of antisense off-targets in prokaryotic organisms. *Genomics* 105, 123–130.
- Wu, L., Fan, J., and Belasco, J.G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA* 103, 4034–4039.
- Wu, Y.-C., Chen, C.-H., Mercer, A., and Sokol, N.S. (2012). Let-7-complex microRNAs regulate the temporal identity of *Drosophila* mushroom body neurons via *chinmo*. *Developmental Cell* 23, 202–209.
- Xu, G., Fewell, C., Taylor, C., Deng, N., Hedges, D., Wang, X., Zhang, K., Lacey, M., Zhang, H., Yin, Q., et al. (2010). Transcriptome and targetome analysis in MIR155 expressing cells using RNA-seq. *Rna* 16, 1610–1622.
- Yang, L., Chen, D., Duan, R., Xia, L., Wang, J., Qurashi, A., Jin, P., and Chen, D. (2007). Argonaute 1 regulates the fate of germline stem cells in *Drosophila*. *Development (Cambridge, England)* 134, 4265–4272.
- Yartseva, V., Takacs, C.M., Vejnar, C.E., Lee, M.T., and Giraldez, A.J. (2016). RESA identifies mRNA-regulatory sequences at high resolution. *Nat Methods*.
- Yekta, S., Shih, I.-H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304, 594–596.
- Yeom, K.-H., Lee, Y., Han, J., Suh, M.R., and Kim, V.N. (2006). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Research* 34, 4622–4629.
- Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* 17, 3011–3016.
- Yoon, J.-H., Srikantan, S., and Gorospe, M. (2012). MS2-TRAP (MS2-tagged RNA affinity purification): Tagging RNA to identify associated miRNAs. *Methods* 1–7.
- York-Andersen, A.H., Parton, R.M., Bi, C.J., Bromley, C.L., Davis, I., and Weil, T.T. (2015). A single and rapid calcium wave at egg activation in *Drosophila*. *Biology Open* 1–8.
- Zekri, L., Huntzinger, E., Heimstädt, S., and Izaurralde, E. (2009). The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol. Cell. Biol.* 29, 6220–6231.
- Zeng, Y., Yi, R., and Cullen, B.R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *The EMBO Journal*

24, 138–148.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Publishing Group* 29, 607–614.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* 118, 57–68.

Zhang, L., Ding, L., Cheung, T.H., Dong, M.-Q., Chen, J., Sewell, A.K., Liu, X., Yates, J.R., and Han, M. (2007). Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Molecular Cell* 28, 598–613.

Zhang, L., Hammell, M., Kudlow, B.A., Ambros, V., and Han, M. (2009). Systematic analysis of dynamic miRNA-target interactions during *C. elegans* development. *Development (Cambridge, England)* 136, 3043–3055.

Zhao, T., Graham, O.S., Raposo, A., and St Johnston, D. (2012). Growing microtubules push the oocyte nucleus to polarize the *Drosophila* dorsal-ventral axis. *Science* 336, 999–1003.

Zhao, W., Pollack, J.L., Blagev, D.P., Zaitlen, N., McManus, M.T., and Erle, D.J. (2014). Massively parallel functional annotation of 3' untranslated regions. *Nat Biotechnol* 32, 387–391.

Zisoulis, D.G., Lovci, M.T., Wilbert, M.L., Hutt, K.R., Liang, T.Y., Pasquinelli, A.E., and Yeo, G.W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat Struct Mol Biol* 17, 173–179.

Ørom, U.A., Nielsen, F.C., and Lund, A.H. (2008). MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Molecular Cell* 30, 460–471.

Appendix: Primers and ASOs

TRACE ASOs

Name	Sequence 5'-3'
crok	ATTTTTTTGGACGGAAATAGTATTCC
Pck	ACTATGGACGGCGTCGATT
CG10217	TTCGTTTGGGACGGAATGGCT
CG12880	AGTGGAACCTGGACGGAAGAGG
Cont	GTCTATAAAATGGACGGAATTGCAT
lqfr (epsin-like)	CGATCAGGACGGATGTGGAAT
Gli	AAGTTATTGGACGGTTCACACTGAT
Nrx-IV	AACTAAAAATGGACGGTATTCTGTG
yrt	CTGCTGGACGGAACGGAAT
CG8010	AATCACTGTGGACGGTTGTCC
l(2)35Bg	CAGTGGACGGATGGCGTTTT
cdc5	AGGCCAGGACGGAGCGCTTT
Sinu (MRE1)	CTTTGAGGGGACGGAATGTTTTTG
Sinu (MRE2)	ACGTGCAAAGGACGGATGATT
Mga2	TTGGATCTGGACGGAGCCTAAA
CG1105	GCTTTCTATTTGGACGGAAAGC
CG1332	TCTGTGTGGACGGACTGTGATC
CG31195 (MRE1)	TTGTTTGGACGGAAGTTTGCG
CG31195 (MRE2)	GACTACTCATTGGACGGATGTTTTTG
CG8121	CATATGCGTTTGGACGGAATAGAC
Mcr	GCATTTTTTGGACGGAATGCAATTCG
Tsf2	CATTTTCGGACGGAAGGGATTTC
cals	TTATTGGGACGGAGAACTTCCTCC
emp	TGTATTTCGAGGACGGAATTTTGT
Ptp99A	CAGATACTAGGACGGACTGAATCG
Sema-1b	TTTAGTGGGACGGACAACAGCT
CG14059	CTATAACTATGGACGGATTCAGAAG
CG4313	ATATCCCGCTAGGACGGAGGACGA
CG6583	GGAAAAAGGTGGACGGAATGTGTT
CG7713	ATGCGCATGGACGGAGTTCTT
CG1298	GTGTAAATGGACGGATAATTCCTTCG
CG2813	AGCTTATGGACGGCTTCGTT
CG13088	CGTATACCGGACGGATTGAGGTA
CG6965	TATTAGTGTGGACGGTGGCTC
CG3446	TATTGCAGCTGGACGGCAAC
CG15154	TGGATGAGGACGGAGCCGAT
Atet	CAACTTAGGGACGGAGTGAAGAAC
BetaInt-nu	AAACGGACGGACGGACAGACAA
Bves	AAGGTGAAGGGACGGATGGGT
Cad96Cb	AACTTTATTGGACGGATCCTCTTAC
Cah2	TCCTCGGACGGAAGGGAATAT
CG14785	AGGACAATGGACGGAAGTACA
CG31495	TCCTGGAGGACGGAATGCTCT

CG4542 (MRE1)	TCACCCACTCGGACGGACGGC
CG4542 (MRE2)	TTCGTCTGGACGGAAGCTCGT
CG6038	GTGGAATGTTGGACGGAATTCGTAG
CG8776	AGCCGGACGGACCGTCTAAAT
Dok	TGAGAAGACGGACGGATAAACGG
drd	AGCACATGGACGGATTGGATG
fy	TTTGTAAGGACGGAATTGTAAGTC
l(2)gl	AATGGGGACGGACCAATTTATCGG
Oseg1	TGTACAGTTTTTGGACGGAAATTAT
Pka-R2	GTCTAGGACTGGACGGATTTTTGAG
Ppcs	GGGCGGGACGGAAGTGGTTA
RecQ5	GTTTCATATGGACGGACAAACGG
Sbr	CGGATCTGCGGACGGAACCTAAA
SPoCk	AGACATACGGACGGACAGAGT
ttk	CGTTCTGGACGGAGCAAGTGTTA
Lac	AGCAAAACGGACGGGATCAAG
CG12789	CTAAAGCACTAGGACGGCTTAAC
XLOC_035581	GGTTCTTGGACGGTCCACACA
XLOC_003055	ATCCTTTGGACGGACTTTCCG
XLOC_001425	CAACGGACGGACGGATGGCTA
XLOC_030523	ATAATGAGGACAGATAACACAC
XLOC_030181	AACGGAGAACGGAGACGAGT
XLOC_030521 (MRE1)	GACAAACGGACGGACAGAACG
XLOC_030521 (MRE2)	GACAGCCGGACGGACGGTCG
XLOC_000338	TGTTTCAGAAGAACGGAGAGCAGAA
XLOC_022242	ACTTAGACCGGACGGAAATTGC
XLOC_022396	GGATCCTGGACGGGGATCC
XLOC_023568	ATTGTGTGGACGGCAGTTTGG
XLOC_031931	GAACGCTTGGACGGGTCAGC
XLOC_019797	TTTCTAGACGGGACGGACATCG
XLOC_005141	AAGAGGCGAACGGAGAATGGT
XLOC_001669 (del1)	TTggacggTTggacggTTGGAC
XLOC_001669 (del2)	TTggacggTTggacggTTGGAC
XLOC_001669 (MRE5)	GCTCGACGGACGGATACTAGAA
CR44786-RA	GGCCGGTTTTTGACGGATTAATTC
Ctr1B	GAACCAAGAACTGATAAGGGCTTAG
CG9422	CGAATTTCAACTGATAAGGGAACCC
CG5850	AGCATATAACTGATAAGGGGCTTT
comm	ACTGAAAAACTGATAAGGAAATGAA
CG11594	GACGCGAACTGATAAGCAACCT
CR44033	ACATTTATCTGATAAGGGCACATGG
Kaz-m1	ATCAGGGGAACTGATAAGCTTGC
CG4705	CGAAGAACTGAACTGATAAGAACG
MRE23	CCAAAACTGATAAGGCCGCGA
Dp1	CTGCGACCAACTGATAAGGAACTAC
Ubc12	GCGTTAACAATAGAACTGATGTAAA

TRACE qPCR Primers

Name	Forward 5'-3'	Reverse 5'-3'
crok	GCCTTTTCTGTAACGGACTCG	CGAGATGGAATGTCATAGGCGT
crok non-cut	ACAACATCTTCATGGAGTTCTGC	AGTCCTAGCCTATGGATGCCC
pck	CGCCCGACTTCCATACATTC	CGTGCTATTTAATAACAATTACAAAC
CG10217	TTTACAACCAACATAAAGTCACT	GG ACTGCTTTAATTATTTGCACAATAC
CG12880	T	A
cont	GCCGGGCGTTCTATTTTCAA	CGCCAGGATTGGGTATTCCA
lqfr (epsin-like)	GCTCTCTTGCAAGCCCCTTA	AGTTCAGCAACATACTGGACCAT
gli	AAAGTTTGCGACACGTGCAA	CGAATCCGTTTAGTACAACCATT
Nrx-IV	TGTGTGTCGCATATAAGCAAAAT	AGTTATTGGACGGTTCACACTGA
yrp	GA	
CG8010	TGGTTCTGTCTGACGGCATT	CTCGGGCAGCTCCCTTAAAT
l(2)35Bg	TCCGAGCAATTATCCGGACA	TACGCGGAGTAGGGGTAC
cdc5	TGCTTCACAATCAACAAACAA	TAACGTATTATGCTTTCGTTATCAC
rp49	TGAACCTCCTTGGTACTGGCC	GAACGACAAATCCTACCGCG
	CATGATGTGGTTACGGCAGA	TTGAGGAGACGGAGGACTT
	TGCTAAGCTGTCTGCACAAATGGC	CGATCCGTAACCGATGTTGGGC

Multiplex ASO

Name	Sequence 5'-3'
crok 12mer	TTTGGACGGAAA
crok 11mer	TTTGGACGGAA
crok 10mer	TTGGACGGAA
cdc5 10mer	CAGGACGGAG
cdc5 11mer	CCAGGACGGAG
miR-184 universal 6mer	GGACGG
miR-184 universal randomized flank 10mer	NNGGACGGNN

Multiplex RT Primers

Name	Sequence 5'-3'
miR-184 seed	GGACGG
hsp83	TACCAGCCGTCTTGTTAGGG
rp49	CGATCCGTAACCGATGTTGGGC
egfp	CTTGTGCCCCAGGATGTTG

Multiplex qPCR Primers

Name	Forward 5'-3'	Reverse 5'-3'
gli non-cut	GCCACTCCCACACATCATCA	GTGGGTCCATCGAGGACAAT
egfp	CAACTACAAGACCCGCGC	CTTGTGCCCCAGGATGTTG
rp49	TGCTAAGCTGTCTGCACAAATGGC	CGATCCGTAACCGATGTTGGGC

DASH sgRNA primers

Name	Sequence
TRACR_RP	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTA TTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC
sgRNA1.1F P	GAAATTAATACGACTCACTATAGGCAGTACGTGAAACTGCTTAGGTTTTAG AGCTAGAAA
sgRNA1.2F P	GAAATTAATACGACTCACTATAGGTTCTTTTCAACTTTCCCTCAGTTTTAG AGCTAGAAA
sgRNA1.3F P	GAAATTAATACGACTCACTATAGGCTTGATAGTGCAGCACTAAGGTTTTAG AGCTAGAAA
sgRNA1.4F P	GAAATTAATACGACTCACTATAGGTCATTAACGTTATACGGGCCGTTTTAG AGCTAGAAA
sgRNA1.5F P	GAAATTAATACGACTCACTATAGGCAAATGTGAGATGCAGTGTAGTTTTAG AGCTAGAAA
sgRNA2.1F P	GAAATTAATACGACTCACTATAGGAGATGTGGTAGCCGTTTCTCGTTTTAG AGCTAGAAA
sgRNA2.2F P	GAAATTAATACGACTCACTATAGGCCCGTTACCCGTTGCAACCAGTTTTAG AGCTAGAAA
sgRNA2.3F P	GAAATTAATACGACTCACTATAGGTAGATAACATGCAGATCGTAGTTTTAG AGCTAGAAA
sgRNA3.1F P	GAAATTAATACGACTCACTATAGGAGTTAATGACTCACACTTACGTTTTAG AGCTAGAAA
sgRNA3.2F P	GAAATTAATACGACTCACTATAGGTTACCCGGACCTCTCGGTCTGTTTTAG AGCTAGAAA
sgRNA3.3F P	GAAATTAATACGACTCACTATAGGTAATGAGATTGAGCAATAACGTTTTAG AGCTAGAAA
sgRNA4.1F P	GAAATTAATACGACTCACTATAGGCTTACCAAAGTGGCCCACTGTTTTAG AGCTAGAAA
sgRNA4.2F P	GAAATTAATACGACTCACTATAGGTTAAGATCGTTTCGACCCTAGTTTTAG AGCTAGAAA
sgRNA4.3F P	GAAATTAATACGACTCACTATAGGATCCTGAGGGAACTTCGGAGTTTTAG AGCTAGAAA
sgRNA5.1F P	GAAATTAATACGACTCACTATAGGATGGTATGCGTAGAAGTGTTGTTTTAG AGCTAGAAA
sgRNA5.2F P	GAAATTAATACGACTCACTATAGGCTACTTTAGCGGTAATGTATGTTTTAG AGCTAGAAA
sgRNA6.1F P	GAAATTAATACGACTCACTATAGGCCATAAATTTTCAAGGTCCGGTTTTAG AGCTAGAAA
sgRNA1.4_ qPCR_FP	CAAGCAACACGACTCTTTGGA
sgRNA1.4_ qPCR_RP	GCAGTGTATGGAGCGTCAAT

***bicoid* TRACE ASO**

Name	Sequence 5'-3'
bcd-1	CTCTCTAACACGCCTCTCATCCAGG
bcd-2	AGTGGTTAACCTAAAGCTAATGAAA
bcd-3	CAATCAGGAACAACAGTGGTTAACC
bcd-4	TATCTACAATCACTTGGTATTTGTA
bcd-5	TCTACGCGTAGATATCTACAATCAC
bcd-6	TAGGACTAGACCTAACTTTCTACGC
bcd-7	CCCTGGGAACCATTTACACGGATCT
bcd-8	TGACTAGGCTAGTACATAAAACTTC
bcd-9	CGTGCGGCCTGCTGACTAGGCTAGT
bcd-10	GATATGCACTGGAATCCGTGCGGCC
bcd-11	GGAGTATCACTAAGATATGCACTGG
bcd-12	AGTTAACTGGAGTATCACTAAGATA
bcd-13	GCGTATTGCAGGGAAAGTATAGAGT
bcd-14	CACCCAGATACATCTAAGGCGAATA
bcd-15	TATTCCCGGGCTTTAGTGGAGCAGC
bcd-16	ATGGCCTCAAATGTAAGTGGTTGCA
bcd-17	TTCCATGGAATACGCTTAAGCCCAA
bcd-18	TAATTTCCGAAATGTGGGACGATAA
bcd-19	AAGATTTTCTTGCTGGCTCGGAATA
bcd-20	TTAGCTATGTCAAATTGTAACAGAG
bcd-21	TTTTTCATTTTGATTAGTACAGTTT
bcd-22	TATGAGATTACGCCCCAAGAGAAACA
bcd-23	GTTTCGATCTTTAAGGGTAATCATTG
bcd-24	TCATATCAAATATTATTGTTTAAAT
bcd-25	AAAATATCATATCAAATATTATTGT
bcd-26	CACTTTGGCATAGCATAGAAATTGA
bcd-27	TGCGCAAATGTTTGATTATGTCAGA
bcd-28	TTGCTGACTATTCTTGGTCAAAGAA
bcd-29	GGTCTGCATTGATTGAAAATACAAT
bcd-30	AAATCTCCGAATCTGAAACAAATGG
bcd-31	ATGATAGTTATTCCGTTTGGCAGCA
bcd-32	AGTGATGTAAATAGAATGTGAGCTA
bcd-33	GATTGCAATGCTCTTCTTAGTGATG
bcd-34	TTAAAATTAACTTGAGGCCTAACA
bcd-35	GAGACAACATCAAAGGTGCAGCATT
bcd-36	CGTAATTAAAAATACAAAGCTTAAA
bcd-37	GAGTAGAGTAGTTCTTATATATTTT
bcd-38	TTAGTCACAATTTACCCGAGTAGAG

***bicoid* TRACE primers**

Name	Forward 5'-3'	Reverse 5'-3'
ORF	TGAGCACCGGAATAAGAGCC	GGCGTTCCGATGGGGATTAT
bcd-A	TGAGAGGCGTGTTAGAGAGTTTC	GGAATCCGTGCGGCCTG
bcd-B	CACGGATTCCAGTGCATATCT	TACGCTTAAGCCCAAATGGC

bcd-C	TTGGGCTTAAGCGTATTCCAT	TGAGATTACGCCCCAAGAGAAACA
bcd-D	TCTCTTGGGCGTAATCTCATACA	CGAATCTGAAACAAATGGTCTGC
bcd-E	TTTTTGCTGCCAAACGGAAT	ACAATTTACCCGAGTAGAGTAGTTC

***bicoid* pulldown ASO**

Name	Sequence 5'-3'
bcd_oligo1	TATCTACAATCACTTGGTATTTGTA
bcd_oligo2	CCCTGGGAACCATTTACACGGAT
bcd_oligo3	GCAGCCACCCAGATACATCTAAGG
bcd_oligo4	CAAATGTAAGTGGTTGCATATTCCC
bcd_oligo5	TAATTTCCGAAATGTGGGACGATAA
bcd_oligo7	TGCGCAAATGTTTGATTATGTCAGA
bcd_oligo8	TTTGCTGACTATTCTTGGTCAAAGA
bcd_oligo9	ATGATAGTTATTCCGTTTGGCAGCA
bcd_oligo10	GAGACAACATCAAAGGTGCAGCATT

***bicoid* antisense pull down probes**

Name	Sequence 5'-3'
Oligo5_Antisense_C18_N	TAATTTCCGAAATGTGGGACGATAA/iSp18//3AmMO/
Oligo9_Antisense_C18_N	ATGATAGTTATTCCGTTTGGCAGCA/iSp18//3AmMO/

***bicoid* pulldown qPCR primers**

Name	Forward 5'-3'	Reverse 5'-3'
ORF	TGAGCACCGGAATAAGAGCC	GGCGTTCCGATGGGGATTAT
A	ATGAGAGGCGTGTTAGAGAGTT	CTAAGATATGCACTGGAATCCGT
B	CACGGATTCCAGTGCATATCT	TTTCTTGCTGGCTCGGAATA
C	TTTCTCTTGGGCGTAATCTCA	CTCCGAATCTGAAACAAATGG
D	CAGACCATTGTTCAGATTCG	CACAATTTACCCGAGTAGAGTAGTTC