Dicer-dependent transfer RNA-derived small RNAs and their role in gene regulation



Sheng Kai Pong

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2021

Impacts of COVID-19 pandemic

The COVID-19 pandemic has significantly impacted the final year of my DPhil:

- I self-isolated, according to the UK Government's guidelines, from 18th March 2020 for 14 days as I displayed mild symptoms associated with SARS-Cov-2.
- On 24th March 2020, I was informed that all experimental work at the Sir William Dunn School of Pathology must cease from 27th March onwards; this news came as the UK went into a national lockdown.
- Between 27th March and 7th June 2020, I worked from home; however, my progress had been hindered by space limitation and the lack of access to lab resources and materials. Most importantly, I could not perform the experiments I planned to complete during this time.
- The department later reopened on 8th June 2020, allowing a maximum of 2 members from each research group to be present in the lab at one time (this was later increased a week later to 2 people per bay, with the department operating at maximum 50% capacity).
- During this period, external orders were often postponed or cancelled, and extra responsibilities were placed on the researchers in the department as core facilities were understaffed.
- With social distancing being imposed, the capacity in the tissue culture room was halved, which has significantly affected my work as all users were only allowed to book a maximum of three hours per day.
- In early October 2020, the arrival of a new DPhil student and a medical student forced the entire lab to work in shifts. I was also assigned the task to supervise the medical student, who was previously scheduled to join in July.

- I was later granted a three-month extension for my DPhil but was only informed in September that my funding would be fully sponsored by the University.
- At the time of writing, the UK has entered a third national lockdown to curb the spreading of a new variant of coronavirus.
- The ever-changing circumstances during this time has brought about unprecedented anxiety and stress which has inevitably impacted on my performance.

I hope that the assessors would take into consideration the difficult circumstances I faced during the final year of my DPhil while examining my thesis.

Abstract

Endogenous small non-coding RNAs (sncRNAs) are employed in mammalian cells to negatively regulate gene expression in a sequence-specific manner. This phenomenon is known as RNA interference (RNAi). Depending on the molecular mechanism, RNAi can be categorised into post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS). PTGS involves direct cleavage of mature messenger RNA (mRNA) or destabilisation of mRNA and/or translational repression in the cytoplasm. TGS is mediated by the establishment of heterochromatin at target loci in the nucleus. Transfer RNAs (tRNAs), which are essential for protein synthesis, have recently been identified as a source for a novel class of regulatory sncRNAs, termed tRNA-derived small RNAs (tsRNAs). The biogenesis of tsRNAs and their role in gene regulation remain largely elusive. By using a combination of unbiased bioinformatic prediction and classic molecular biology methodologies, I have studied the role of tsRNAs in the regulation of gene expression. Here I propose that Dicer is involved in tsRNA biogenesis and Dicer-dependent tsRNAs can promote gene silencing through a mechanism that is distinct from well-known PTGS and TGS. I show that Dicer-dependent tsRNAs may lead to gene silencing by targeting nascent transcripts, a mechanism termed nascent RNA silencing (NRS). Finally, I explored the therapeutic potential of tsRNAs and demonstrated the silencing of multiple genes by employing tsRNA mimics.

Acknowledgements

My journey as a DPhil student has been challenging at times but overall rewarding.

I would not have reached this stage without the support and mentorship of my supervisor, Dr Monika Gullerova. She has showed me not only the world of academic research but also the translational aspect of science, rendering my DPhil slightly unusual but more exciting than my peers. And for that, I am grateful.

I am privileged to have collaborated with two brilliant scientists, Dr Margarita Schlackow and Dr Adele Alagia. I am thankful for their generous contributions to my DPhil project. Other members of the lab, especially Ruth Ketley and Dr Qilin Long, have brought so much joy and happiness to my time in the lab. I would also like to thank Dr Kaspar Burger who trained me to become an independent researcher at the beginning of my DPhil.

This DPhil was made possible by the funding I received from the Oxford Interdisciplinary Bioscience Doctoral Training Partnership and the E. P. Abraham Research Fund from the Sir William Dunn School of Pathology. I would like to thank the committee members for giving me the opportunity to pursue my passion for science without any financial burden. I would also like to thank Cancer Research UK for sponsoring the translational development of this research project.

I am lucky to have a group of close friends that I can lean on to whenever I face a difficulty in the past four years. The relentless support I received from them and my family has been critical for the completion of my DPhil. I would like to extend my heartfelt gratitude to every single friend and family member who lent their ears to my sometimes-unreasonable complaints.

I am blessed to have a loving partner who accompanied me throughout the second half of my DPhil journey. Without his oversight and the hearty meals he prepared, I would not have been able to complete my DPhil in the midst of a global pandemic.

Contributions

I have included relevant data that were produced by Dr Margarita Schlackow, Dr Adele Alagia, and Kaiyang Song in this thesis to provide a fuller picture of this research project. M. S. performed all the bioinformatic analyses for the project (Figures 3-1, 3-4 & 5-1). I was very much involved in driving the direction of the computational analyses following the prediction of tsRNA target genes. A. A. was responsible for the *in vitro* Dicer and Ago2 cleavage assays (Figures 3-3, 4-8a). Under my supervision and mentorship, K. S. completed an 8-week research project exploring the use of tsRNA mimics for the silencing of *BCL2* in MCF7 cells (Figure 5-2a, b, 5-4 & 5-5). Both M. S. and A. A. were listed as co-authors of the manuscript we submitted for peer review.

Table of Contents

Contents

Impacts of COVID-19 pandemic i

Abstract iii

Acknowledgements iv

Contributions vi

Table of Contents vii

Abbreviations xii

List of Figures xv

List of Tables xvii

1 Background 1

1.1 Introduction 1

1.2 Non-coding RNA 1

1.2.1 Pervasive transcription 1

1.2.2 Small non-coding RNA 2

1.2.3 Regulation of gene expression by small non-coding RNA 3

1.3 RNA interference 4

1.3.1 Biogenesis of microRNA and small interfering RNA 4

1.3.2 Assembly of the RNA-induced silencing complex 5

1.3.3 Post-transcriptional gene silencing 5

1.3.4 Transcriptional gene silencing 9

1.3.5 Intronic gene silencing 11
1.4 tRNA-derived small RNAs 14
1.4.1 tRNA 14
1.4.2 tRNA halves: biogenesis and functions 16
1.4.3 tRNA-derived fragments 17
1.5 RNAi therapeutics 22
2 Materials and Methods 24
2.1 Tissue culture 24
2.1.1 List of cell lines 24
2.1.2 Culture conditions 25
2.2 Experimental techniques 25
2.2.1 RNA purification 25
2.2.2 Northern blot 25
2.2.3 Western blot 25
2.2.4 sRNA-seq and RNA-seq sample preparation 26
2.2.5 *In vitro* Dicer cleavage assay 26
2.2.6 qRT-PCR 27
2.2.7 Subcellular fractionation 27
2.2.8 Immunoprecipitation 28
2.2.9 ChIP-qPCR 28
2.2.10 chrRNA-seq sample preparation 29

2.2.11 Labelling of small RNAs with fluorescent dye 30
2.2.12 CRISPR-Cas9 knockout of *SPINT1* target site 30
2.2.13 *In vitro* Argonaute 2 cleavage assay 32
2.3 Statistical analysis of real time PCR data 33
2.4 Bioinformatics methods 33
2.4.1 sRNA-seq and PAR-CLIP analyses 33
2.4.2 chrRNA-seq analysis 34
2.4.3 Differential gene expression analysis 35
2.4.4 tsRNA target gene prediction 35
2.4.5 Disease-gene association heatmap 35
2.4.6 Data and materials availability 36
2.5 List of antibodies 36
2.6 List of oligonucleotides 37
2.6.1 Northern blot 37
2.6.2 qRT-PCR primers 37
2.6.3 ChIP-qPCR primers 40
2.6.4 T7 transcription 41
2.6.5 tsRNA mimics 42
3 Biogenesis of tsRNAs and prediction of target genes 44
3.1 Introduction 44
3.2 Dicer is involved in tsRNA biogenesis 44

3.3 Dicer cleaves tRNA *in vitro* 47

3.4 Bioinformatic prediction of tsRNA target genes 49

3.5 Silencing of tsRNA target genes may be miRNA-independent 51

3.6 Regulation of selected tsRNA target genes occurs in the nucleus 53

3.7 Transcription of selected target genes is unaffected by Dicer knockdown 57

3.8 Hypothetical model of nascent RNA silencing 59

3.9 Discussion 60

4 Molecular mechanism of tsRNA-mediated gene silencing 64

4.1 Introduction 64

4.2 Genome-wide analysis of chromatin-associated RNA reveals upregulation of many genes in Dicer and Ago2 knockdown 64

4.3 tsRNA are present in both the nucleus and cytoplasm 69

4.4 Transfection of tsRNA mimics leads to silencing of selected target gene *SPINT1* 71

4.5 CRISPR-Cas9 knockout of a predicted tsRNA target site results in elevated levels of the target gene 73

4.6 Role of Ago2 in tsRNA-mediated gene silencing 76

4.7 Ago2 slicer activity in nascent RNA silencing 78

4.8 Discussion 81

5 Therapeutic potential of tsRNAs 84

5.1 Introduction 84

5.2 Association between tsRNA target genes and diseases 84

5.3 Silencing of proto-oncogenic genes using tsRNA mimics 85

5.4 Effect of RNA modifications on the silencing of tsRNAs 89

5.5 Intron- versus exon-targeting tsRNA 91

5.6 Silencing of *TTR* 92

5.7 Discussion 97

6 Conclusions and perspectives 99

6.1 Summary of thesis 99

6.2 Future work 100

7 Bibliography 103

8 Appendices 121

Abbreviations

AARS	Aminoacyl-tRNA synthase
Ago	Argonaute
ANG	Angiogenin
ASGPR	Asialoglycoprotein receptor
ChIP	Chromatin immunoprecipitation
ChrRNA-seq	Chromatin-associated RNA sequencing
CLASH	Cross-linking, ligation, and sequencing of hybrids
CLRC	Cryptic loci regulator complex
co-TGS	Co-transcriptional gene silencing
DCP2	Decapping protein 2
DGCR8	DiGeorge syndrome critical region 8
DNA	Deoxyribonucleic acid
dsRNA	Double-stranded RNA
EC	European Commission
FDA	Food and Drug Administration
H3K9	Histone H3 lysine 9
H3K9me2	Histone H3 lysine 9 dimethylation
hATTR	Hereditary transthyretin amyloidosis
itRF	Internal tRNA-derived fragments
kb	Kilobases
lncRNA	Long non-coding RNA
MEF	Mouse embryonic fibroblast
miRNA	MicroRNA

mNET-seq	Mammalian nascent elongating transcript sequencing
mRNA	Messenger RNA
NMD	Nonsense-mediated messenger RNA decay
nt	Nucleotide
PAGE	Polyacrylamide gel electrophoresis
PAR-CLIP	Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
piRNA	PIWI-associating RNA
pre-mRNA	Premature messenger RNA
pre-miRNA	MicroRNA precursor
pri-miRNA	Primary microRNA transcript
PTC	Premature termination codon
qPCR	Real-time polymerase chain reaction
RDRC	RNA-directed RNA polymerase complex
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing complex
RNA	Ribonucleic acid
RNA-seq	Poly(A) RNA sequencing
RNAi	RNA interference
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
RT	Reverse transcription
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA

sncRNA	Small non-coding RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
sRNA	Small RNA
sRNA-seq	Small RNA sequencing
SS	Splice site
TAP	Tandem affinity purification
TGS	Transcriptional gene silencing
tRNA	Transfer RNA
tiRNA	Transfer RNA-derived, stress-induced RNA
TRBP	Transactivation response element RNA-binding protein
tRF	Transfer RNA-derived fragment
tsRNA	Transfer RNA-derived small RNA
UTR	Untranslated region
XRN1	5’-to-3’ exoribonuclease 1

List of Figures

Figure 1-1 Post-transcriptional gene silencing 8

Figure 1-2 Co-transcriptional gene silencing in *S. pombe* 10

Figure 1-3 Secondary and tertiary structures of tRNA 15

Figure 1-4 Classification of tRNA-derived small RNAs 18

Figure 3-1 Global tsRNA levels are reduced upon Dicer knockdown 45

Figure 3-2 Levels of selected tsRNAs are reduced upon Dicer knockdown 46

Figure 3-3 Dicer processes tRNA into small RNA *in vitro* 49

Figure 3-4 Bioinformatic prediction of tsRNA target genes 50

Figure 3-5 Regulation of selected tsRNA target genes may not be miRNA-dependent 52

Figure 3-6 Selected tsRNA target genes are regulated in the nucleus 56

Figure 3-7 Transcription of tsRNA target genes is unaffected by Dicer knockdown 58

Figure 3-8 Hypothetical model of nascent RNA silencing 60

Figure 4-1 Summary of chrRNA-seq results 66

Figure 4-2 chrRNA-seq profiles of target versus non-target genes 69

Figure 4-3 Cellular localisation of tsRNA 70

Figure 4-4 Transfection of tsRNA mimics silences target gene *SPINT1* 72

Figure 4-5 CRISPR-Cas9 knockout of *SPINT1* target site 74

Figure 4-6 *SPINT1* expression is elevated in homozygous knockout mutant 76

Figure 4-7 Role of Ago2 in tsRNA-mediated gene silencing 78

Figure 4-8 Ago2 slicer activity in NRS 80

Figure 5-1 Disease-tsRNA target gene association analysis 85

Figure 5-2 *BCL2* is silenced by tsBCL2 transfection 88

Figure 5-3 *EGFR* is silenced by tsEGFR transfection 89

Figure 5-4 Silencing effects of 5' phosphorylated and duplex tsBCL2 90

Figure 5-5 Comparison between intron- and exon-targeting tsBCL2 92

Figure 5-6 Predicted target sites of tsTTR candidates.....95

Figure 5-7 *TTR* is silenced by transfection of tsTTR candidates.....96

List of Tables

Table 1 Classification of small non-coding RNAs 2

1 Background

1.1 Introduction

It is well known that the human genome is pervasively transcribed into RNA[1,2]. How much of this pervasive transcription is functional remains a subject of debate[3]. We know that a portion of RNAs, which are known as messenger RNAs (mRNAs), act as templates for protein synthesis. Structural RNAs, such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) are essential for mRNA translation. The remaining RNAs do not code for proteins but serve a wide range of functions that are crucial for the survival and maintenance of cells.

In this thesis, I focus on studying the derivatives of tRNAs, a group of non-coding RNAs termed transfer RNA-derived small RNAs (tsRNAs), and their role in the regulation of gene expression. Here, I will introduce concepts that are relevant and important for the understanding of this research undertaking.

1.2 Non-coding RNA

1.2.1 Pervasive transcription

Technological advances in whole-genome sequencing techniques have enabled researchers to gain unprecedented insights into cellular processes at the molecular level. One of the most surprising findings is that approximately 85% of the human genome is transcribed into RNA, despite the fact that only 3% of the human DNA encode proteins[4]. This observation immediately raises one question: why does the cell generate so many RNAs that have no role in the synthesis of proteins, arguably the most important biological molecules besides DNA in the cell? This so-called ‘dark matter’ of the cell has caught the attention of many biologists, leading to the inception of the study of non-coding RNAs and their functional relevance. So far, increasing evidence has shown that

many non-coding RNAs are involved in a wide variety of cellular processes, dismissing the notion that RNA merely acts as a messenger between DNA and protein.

1.2.2 Small non-coding RNA

Non-coding RNAs can be arbitrarily categorised into long and small (or short) non-coding RNAs according to their lengths. Long non-coding RNAs (lncRNAs) are non-coding RNAs that are longer than 200 nucleotides (nt) while small non-coding RNAs (sncRNAs) are shorter than 200 nt[5]. Since the main subject of this thesis is tsRNA, I will be focussing on discussing sncRNAs and their biological roles in this chapter.

sncRNAs can be further divided into tRNAs, rRNAs, small nuclear RNA (snRNA), small nucleolar RNA (snoRNAs), microRNAs (miRNAs), small interfering RNA (siRNA), Piwi-interacting RNAs (piRNAs), and tsRNAs, based on their biogenesis and functions (Table 1).

Table 1 Classification of small non-coding RNAs

sncRNA class	**Size (nt)**	**Polymerase specificity**	**Functions**
tRNA	~70-90	RNA polymerase (pol) III	Translation
rRNA (5S and 5.8S)	~120-150	RNA pol I and RNA pol III	Components of ribosomes
snRNA	~150	RNA pol II and RNA pol III[6]	Components of the spliceosome
snoRNA	60-140	RNA pol II	rRNA modifications[7]

miRNA	~22	RNA pol II	RNA interference (RNAi)[8]
siRNA	20-25	Exogenous (humans)	RNAi
piRNA	26-30	RNA pol II[9]	Transposon silencing[10]
tsRNA	15-50	RNA pol III	Diverse[11,12]

In summary, sncRNAs have diverse biological functions, ranging from splicing to gene regulation. Among these functions, one of the most well-studied is the regulation of gene expression mediated by sncRNAs.

1.2.3 Regulation of gene expression by small non-coding RNA

sncRNAs are employed in plants and animals to regulate gene expression. The term RNAi was coined by Fire *et al.* in their landmark paper showing effective silencing of genes by the introduction of double stranded RNAs (dsRNA) matching the sequence of the targeted genes in *C. elegans*[13]. It was later revealed that exogenous siRNA of 21-25 nucleotides can trigger sequence-specific gene silencing in mammalian cell lines[14,15]. RNAi has since been used generally to describe a phenomenon where gene silencing results from the action of regulatory sncRNAs in a sequence-specific manner. In 2006, Andrew Fire and Craig Mello were awarded the Nobel Prize in Physiology or Medicine for their discovery of siRNA-mediated RNAi[16].

miRNA was first described in a study by Lee *et al.* in which a 22-nt sncRNA coded by the *lin-4* gene negatively regulates *lin-14* activity in *trans*[17]. Temporal negative regulation of *lin-14* is crucial for the larval development of *C. elegans*. Interestingly, they discovered that the sequences of *lin-4* miRNA are complementary to the repeated sequence element

embedded in the 3' untranslated region (UTR) of *lin-14* mRNAs. In 2000, a report showed that the *let-7* gene encodes a 21-nt sncRNA which is complementary to the 3' UTR of several heterochronic genes in *C. elegans* and negatively regulates them[18]. It was later found that *let-7* miRNA is evolutionarily conserved across the three main clades of bilaterian animals, but not in cnidarian and poriferan species and non-animals, such as *Saccharomyces cerevisiae*, *Escherichia coli* or *Arabidopsis*. The evolution of the gene was therefore suggested to occur after the divergence of diploblastic and bilaterian animals[19].

Chemically synthesized, exogenous siRNA duplexes are now widely used in the laboratory to downregulate gene expression specifically. Even though miRNA and siRNA are very similar functionally, they are in fact different in terms of biogenesis and mode of action. In the next section, I will elaborate on the biogenesis of these silencing sncRNAs and the mechanisms by which they negatively regulate gene expression.

1.3 RNA interference

1.3.1 Biogenesis of microRNA and small interfering RNA

siRNAs are derived from endogenous or exogenous long dsRNA sources. Long dsRNAs are processed by a Dicer family enzyme through its RNase III domains into 20-25 nt siRNA duplexes with 2 nt 3' OH overhangs and 5' monophosphates[20–23].

The biogenesis of miRNA begins with the transcription of primary miRNA transcripts (pri-miRNA) by RNA pol II[24]. In the nucleus, pri-miRNAs are folded into stem-loop structures that are recognised by the Microprocessor, which comprises one molecule of Drosha and two molecules of DiGeorge syndrome critical region 8 (DGCR8). Drosha is the catalytic subunit of the complex[25], while DGCR acts as a cofactor which stabilises Drosha through its C-terminal tail and dsRNA-binding domains[26]. The resulting

intermediate products (~70 nt), precursors of miRNA (pre-miRNA), are later exported into the cytoplasm by Exportin 5[27] and processed by Dicer into mature miRNA duplexes[28].

1.3.2 Assembly of the RNA-induced silencing complex

Even though siRNAs and miRNAs are not generated in the same pathway, both associate with Argonaute (Ago) proteins to form the RNA-induced silencing complex (RISC) which coordinates downstream silencing activities through interacting with other protein factors[29].

In humans, the assembly of the RISC requires Dicer and its binding partner transactivation response element RNA-binding protein (TRBP). Once processing of sRNA precursors is completed, the nascent siRNA duplex is released from the binding of Dicer and repositioned onto another dsRNA-binding site of the protein with the help of TRBP[30]. TRBP, acting as an asymmetry sensor, positions the RNA in an orientation that allows proper strand selection and loading[31,32]. The sRNA duplex is loaded onto Ago through an ATP-dependent Hsc-70/Hsp90 chaperone machinery[33]. The duplex is later unwound within Ago through the action of the N domain of the protein, resulting in the retention of the biologically active strand[34]. The strand choice is determined by the thermodynamics properties of the sRNA duplexes[35,36].

The RISC is then guided by the sRNA to its complementary target RNA to exert its regulatory function. Depending on the mode of gene silencing, RNAi can be sub-divided into post-transcriptional gene silencing (PTGS) and transcription gene silencing (TGS).

1.3.3 Post-transcriptional gene silencing

PTGS is characterised by RISC-mediated gene silencing at the post-transcriptional level, through either direct target mRNA cleavage or translational repression. The mode of PTGS is determined by two main factors:

1. The complementarity between the regulatory sRNA and its target mRNA
2. The Ago protein that constitutes the RISC

For direct target mRNA cleavage, the sRNA (siRNA or miRNA) must be fully complementary to its target sequence and they must be associated with Argonaute 2 (Ago2), which has an endonuclease activity embedded in its ribonuclease H domain[37–39]. Ago2 has long been considered to be the only Ago protein within the family to have the 'slicer' activity; however, a recent report has claimed that Ago3 also possesses an endonuclease ability but such activity is substrate-dependent and less common[40]. In animals, miRNAs are only partially complementary to their mRNA targets hence a majority of miRNA-mediated PTGS events are not mediated through slicing[41].

Most miRNAs silence gene expression through mRNA destabilisation and/or translational repression. miRNA-guided RISC finds its targets via base-pairing interactions with partially complementary sequences predominantly in the 3' UTR[42]. Pairing between the target mRNA sequences and the 5' portion of the miRNA (from nucleotide 2 to 8), called the seed sequence, is particularly important for target recognition[43]. The interactions between miRNA and their targets have been extensively studied and numerous bioinformatic tools for predicting target transcripts have been developed[43–50]. Following target recognition, gene silencing is achieved through the interaction with GW182 proteins, which connects the RISC::mRNA complex to the 5' to 3' mRNA decay pathway[51–53]. In this pathway, mRNAs are first deadenylated by the PAN2-PAN3 and CCR4-NOT deadenylase complexes[54] and decapped by the decapping protein 2 (DCP2) together with its cofactors[55]. The silencing is finalised by mRNA degradation, which is executed by the major cytoplasmic nuclease 5'-to-3' exoribonuclease 1 (XRN1)[8]. Accelerated mRNA degradation by miRNA-guided RISC accounts for most of the miRNA-mediated PTGS events in cultured mammalian cells.

Translational repression by miRNAs, however, is not yet fully understood. Several mechanisms have been proposed, mainly surrounding repressing the initial step of translation by either recruitment of translational repressor through GW182, or dissociating the mRNA from the initiation complex[56]. Figure 1-1 illustrates PTGS mediated by siRNA and miRNA, and highlights the differences in biogenesis and mode of action[57].

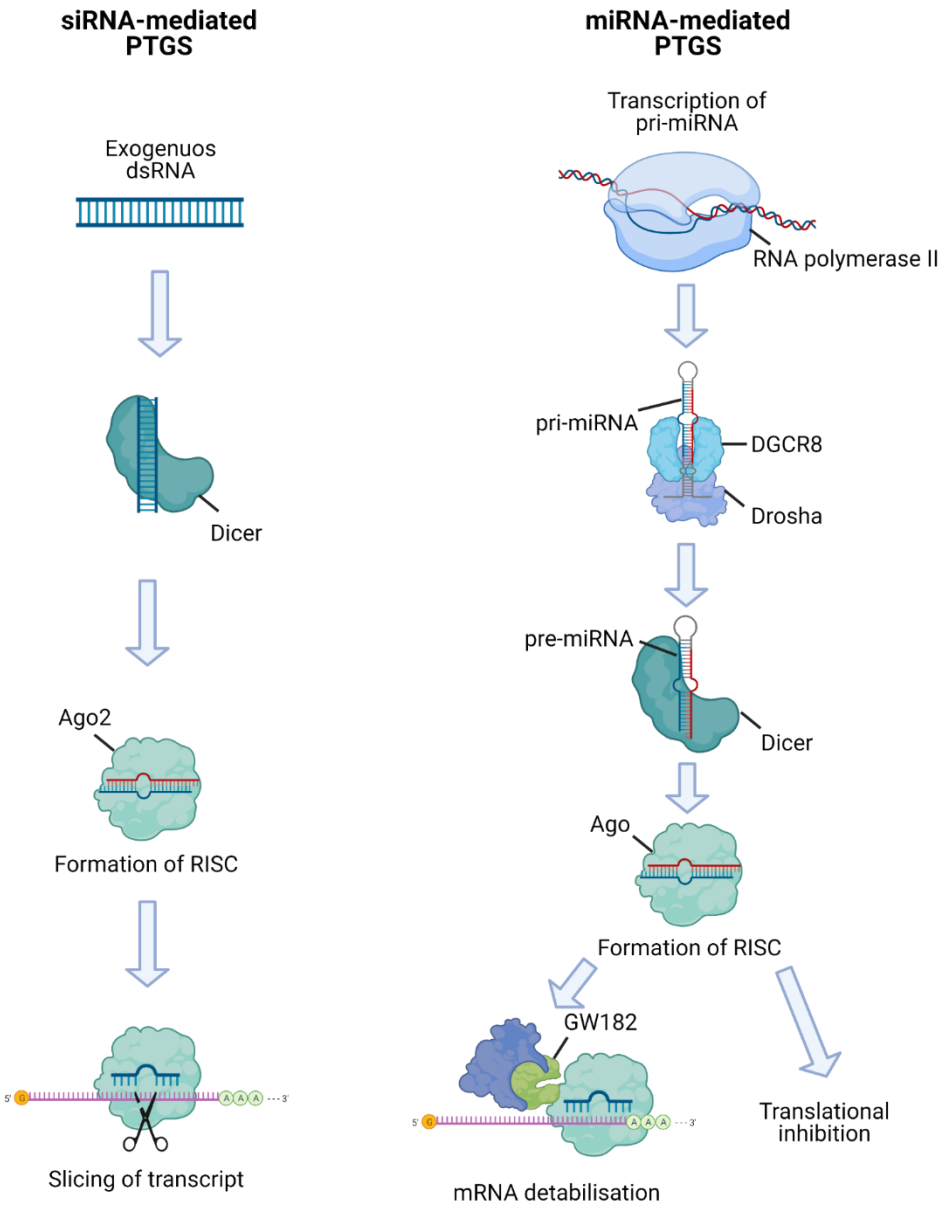


Figure 1-1 Post-transcriptional gene silencing Diagram summarising the biogenesis of siRNA and miRNA, and the mechanism by which they carry out post-transcriptional gene silencing (PTGS). siRNAs are derived from exogenous double-stranded RNA (dsRNA) sources while miRNAs are cleavage products of endogenous miRNA-encoding transcripts by the Microprocessor (Drosha and DGCR8) and Dicer. Both siRNAs and miRNAs associate with Argonaute (Ago) proteins to form the RNA-induced silencing complex (RISC). siRNA mediates PTGS through full complementary binding with target messenger RNA (mRNA), resulting in direct cleavage by Ago2. miRNA carries out PTGS by taking advantage of the 5'-3' mRNA degradation machinery through the interaction with GW182 and other accessory proteins. The matching between miRNA and its target mRNA is not fully complementary and largely depends on the 'seed sequence'. The outcome of miRNA-mediated PTGS can be either mRNA destabilisation and/or translational inhibition, the latter of which is not well understood. Adapted from Wilson & Doudna (2013)[57].

1.3.4 Transcriptional gene silencing

Another RNAi mechanism that has received increasing attention is TGS in the nucleus. In this pathway, gene silencing is achieved by the formation of heterochromatin through the interaction between the sRNA-guided silencing complex and the target loci[58,59].

Nuclear RNAi is best studied in the fission yeast *Schizosaccharomyces Pombe* (*S. Pombe*) in which RNA-induced transcriptional silencing complex (RITS) mediates the formation of constitutive epigenetically repressive regions (heterochromatin) at pericentromeric regions. Deletion of key RNAi factors, namely Argonaute, Dicer and RNA-dependent RNA polymerase (RdRP), results in accumulation of transcripts from the centromeric repeats and loss of histone H3 lysine-9 (H3K9) methylation in *S. pombe*[60]. Another independent study reported 12 sequences of ~22-nt siRNAs that are complementary to *S. pombe* centromeric repeats[61,62]. Both reports suggested a relationship between RNAi and histone modifications in eukaryotes.

The molecular mechanism of nuclear RNAi is rather complex (Figure 1-2). The initial siRNAs in this pathway are derived from dsRNAs formed by bidirectional transcription of the centromeric repeats[60]. Guided by siRNA, the RITS targets the nascent transcript at the repetitive region[63] and recruits RNA-directed RNA polymerase complex (RDRC), which interacts with both the RITS and Dicer, to generate dsRNA substrates for Dicer processing and further amplify the siRNA response[64,65]. A positive feedback loop is established through the targeting of additional nascent transcripts by siRNA-loaded RITS. The localisation of the RITS and the chromatin is stabilised by the interaction with Chp1, a component of the RITS, and H3K9me[62], allowing the recruitment of cryptic loci regulator complex (CLRC). Clr4, a constituent of CLRC, catalyses the methylation of H3K9 and the spreading of heterochromatin[66]. TGS is also known as co-TGS or CGTS[63]

since transcription is required for the initial recruitment of RITS to the target loci in *cis*. The model for TGS/CTGS is summarised in Figure 1-2.

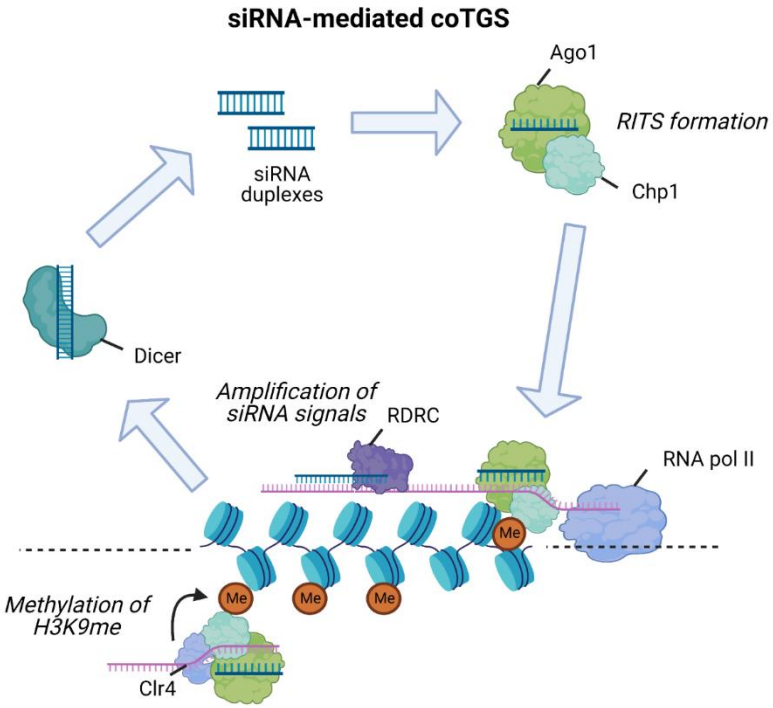


Figure 1-2 Co-transcriptional gene silencing in *S. pombe* Diagram illustrating the current model of co-transcriptional transcriptional gene silencing (co-TGS) in fission yeast. Bidirectional transcripts generated from pericentromeric regions form dsRNA substrates which are subject to Dicer processing. The resulting siRNAs enable the formation of RNA-induced transcriptional silencing complex (RITS) and subsequent recognition of non-coding transcript of the same site, allowing the recruitment of RNA dependent RNA polymerase complex (RDRC) and amplification of siRNA signals. Chp1 stabilises the interaction between the RITS and histone H3 lysine 9 (H3K9me) and recruit the cryptic localisation regulator complex (CLRC) to the target loci. A component within CLRC, Clr4, facilitates the methylation of H3K9me, leading to heterochromatin establishment. Adapted from Martienssen and Moazed (2015)[59].

Apart from constitutive heterochromatin sites, nuclear RNAi has been shown to be active around convergently transcribed genes. A report by Gullerova and Proudfoot showed that bidirectional readthrough transcripts generated in early G1 can form dsRNA substrates which are subject to Dicer processing. The resulting siRNA duplexes guide RITS to the target loci in *cis*, leading to the methylation of histone H3K9. Cohesin is later recruited through the interaction with Swi6, a heterochromatin protein before and during early S

phase. When transcription is reinitiated after replication, RNA pol II is blocked by the presence of cohesin in the intergenic region, promoting transcriptional termination[67].

In summary, RNAi comes in two main flavours, each mediated by Ago-containing RISC and RITS, respectively. Guided by sRNAs, the RISC can downregulate gene expression by directly cleaving target mRNA or destabilising the mRNA and exposing it to the mRNA degradation machinery. In co-TGS, RITS helps recruit proteins that methylate H3K9 to form heterochromatin at target loci.

1.3.5 Intronic gene silencing

The PTGS events I have described so far have largely been studied as a cytoplasmic mechanism and thought to be strictly cytoplasmic[68]. Very few have considered whether canonical PTGS can occur in the nucleus, even though increasing evidence has shown that RNAi factors can be detected in the nucleus[69]. A 2018 report showed that miRNA mediated PTGS can occur in the nucleus of adult muscle stem cells through the interaction with GW182 proteins and the CCR4-NOT deadenylase complex, similar to that of the cytoplasmic RNAi machinery. Using photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP), the authors revealed that 50% of the Ago PAR-CLIP binding sites in the nucleus were sequences derived from introns of protein-coding genes, suggesting an intron-targeting gene silencing mechanism[70]. However, how the RISC silences gene expression through interacting with intronic sequences was not discussed. Computational analysis of crosslinking, ligation and sequencing of hybrids (CLASH) data from Ago1 immunoprecipitation in HEK293 cells identified putative interactions between Ago1-loaded tsRNAs and intronic sequences, suggesting that introns can serve as platforms for gene silencing[71].

For introns to be involved in the regulation of gene expression, intronic sites must be available and accessible to regulatory proteins or ribonucleoprotein complexes. In this section, I will discuss recent findings which indicate that there is a possibility for introns to become regulatory elements.

1.3.5.1 Splicing and transcription kinetics

Splicing has long been established as a co-transcriptional mechanism[72,73]. However, due to its transient nature, it is technically difficult to study the kinetics of the splicing machinery and the dynamics between splicing and transcription.

In lower eukaryotes, such as budding yeast, splicing is halfway complete when RNA pol II is only 45 nt downstream of introns[74], meaning intron synthesis and splicing happen almost simultaneously. Long-read RNA sequencing data in *S. pombe* also supports the close coupling between transcription and splicing, documenting the loss of an intron when RNA pol II is positioned just downstream of the intron[75]. The picture, however, is slightly different in higher eukaryotes in which gene structures are more complex and alternative splicing more frequent.

Using mammalian native elongating transcript sequencing (mNET-seq), Nojima *et al.* was unable to detect spliced transcripts at RNA pol II active sites, an observation suggesting that splicing transpires after RNA pol II has transcribed a "significant" distance beyond the 3' splice site (SS)[76]. A recent paper by Drexler *et al.* has provided a clearer picture on splicing kinetics and the coordination between splicing and transcription in *Drosophila* and human cells by directly sequencing nascent RNA through nanopores. Even though the method can study splicing dynamics without biases from reverse transcription and amplification, the sequencing depth was relatively low, and specific splicing profiles of lowly expressed genes were unobtainable. However, by

looking at splicing dynamics globally, the authors concluded that most human introns are not spliced until RNA pol II has transcribed around 4 kilobases (kb) downstream of the intron (measured from the 3' SS)[77]. The same report also showed that Drosha processing of miRNA-containing introns occurs prior to splicing. These findings are intriguing because the gap between intron synthesis and the completion of splicing provides an opportunity for introns to be accessed by other regulatory molecules and proteins that are present in the nucleus co-transcriptionally.

1.3.5.2 Intron retention

Another explanation for the identification of intronic sequences in Ago PAR-CLIP and CLASH data is the retention of introns in some mature mRNAs.

Mature mRNAs are generally fully spliced and exported to the cytoplasm for translation. Alternative splicing, however, has led to production of a group of mature mRNAs retaining introns. Only until recently, many believe that intron retention is a result of an anomaly in splicing and most mRNAs with retained introns would be subject to nonsense-mediated decay (NMD) in the cytoplasm, as introns normally contain a premature termination codon (PTC)[78]. Recent studies have shown that mRNAs with retained introns can have other fates.

For example, retention of introns within the 3' UTR of mRNAs provides a larger platform for miRNA-mediated PTGS, resulting in an alternative pathway for retained introns to be destabilised. Alternative protein isoforms can also be synthesised from mRNA with retained introns. Interestingly, some transcripts with retained introns can be 'detained' within the nucleus, precluding export into the cytoplasm and translation[78]. Many 'detained' introns are evolutionarily conserved and have half-lives of more than an hour[79].

The question of whether retained or 'detained' introns can be targeted for regulation remains to be explored.

1.4 tRNA-derived small RNAs

In this section, I intend to provide a broad overview on tsRNAs with an emphasis on their potential roles in gene silencing, as this will the main theme of the thesis.

1.4.1 tRNA

tRNA is a type of sncRNA well-known for its evolutionarily conserved role in translating the information coded in mature mRNA into a sequence of amino acids through the ribosome, completing the transfer of genetic information from DNA through to protein[80]. This process involves the pairing between the anticodon embedded in the tRNA and the triplet codon on the mRNA alongside the formation of a peptide bond between the amino acid linked to the tRNAs. In humans, a total of more than 400 tRNA genes encode 64 anti-codons for 20 standard biological amino acids[81].

Biogenesis of tRNA involves synthesis of precursors by RNA pol III and a multi-step maturation process. During maturation, the 5' leader and 3' trailer sequences of a tRNA precursor are processed by RNases P and Z, respectively. In humans, only around 7% tRNAs have introns (28 out of 417). Intronic sequences in the anticodon stem of tRNAs are excised by an endonuclease complex and subsequently joined by a ligase[82,83]. Correctly processed tRNAs are then exported into the cytoplasm by Exportin-T[84] for the addition of an invariant 3' CCA tail and later an amino acid to form an aminoacyl-tRNA, the latter of which is catalysed by aminoacyl tRNA synthetases (AARSs)[85]. AARSs recognises the right tRNAs by interacting with the discriminator base (N73), the acceptor stem and the anticodon.

tRNA folds into a canonical clover-leaf secondary structure[86] and an L-shaped 3D structure[87] (Figure 1-3). This conserved structure is vital for the function of tRNA and its interaction with other proteins.

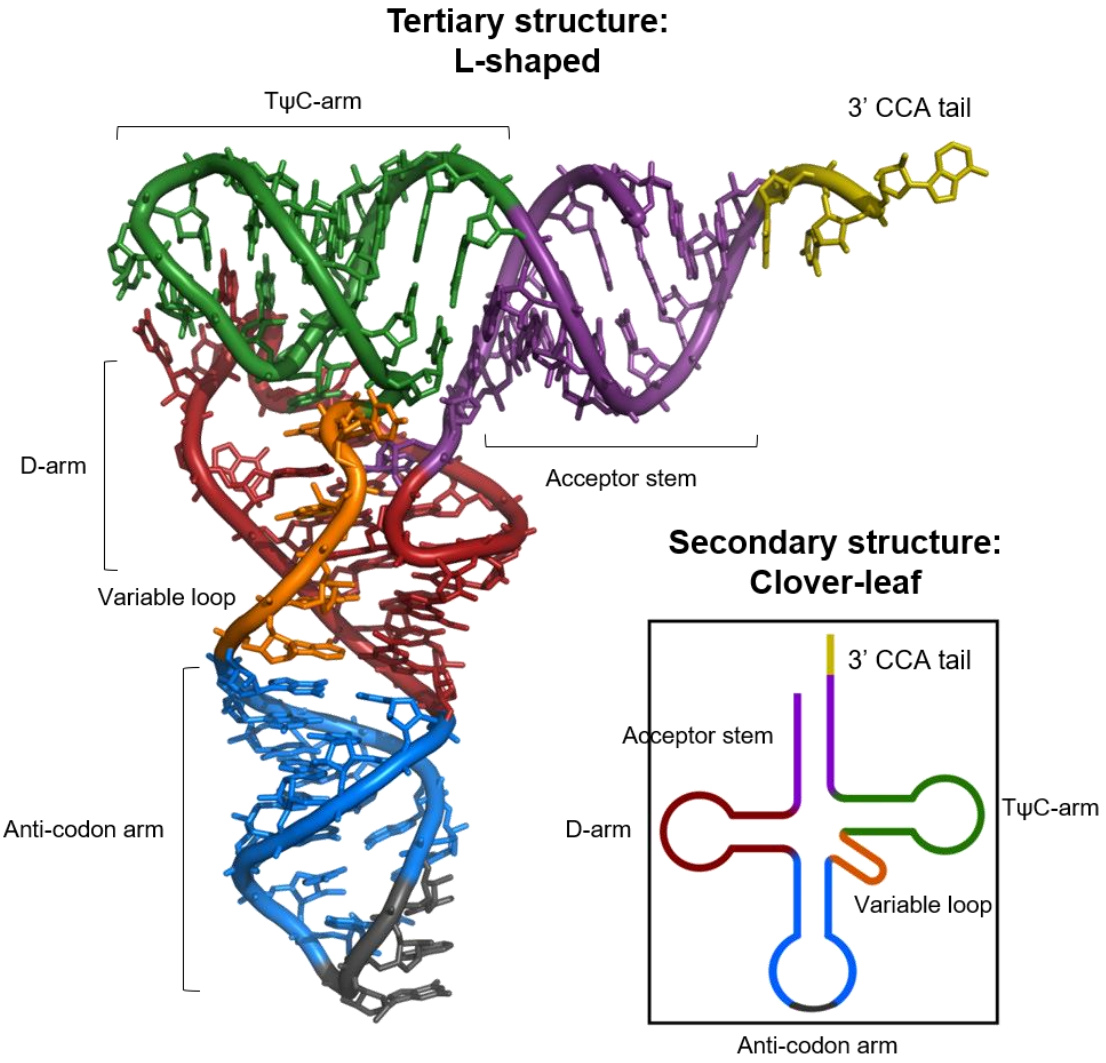


Figure 1-3 Secondary and tertiary structures of tRNA (Left) Diagram illustrating the canonical L-shaped tertiary structure of the cytosolic $tRNA^{Phe}$ in budding yeast (*Saccharomyces cerevisiae*). (Right) Diagram depicting the secondary clover-leaf structure of the tRNA. Important structural components of the tRNAs are labelled and colour coded. Adapted from Lorenz *et al.* (2017)[88].

Among all RNAs, tRNA are the most heavily modified RNA molecules; 93 different modifications (out of 112 known RNA modifications) are found in tRNAs[89]. Pseudouridine (ψ), the most abundant modified nucleotide in tRNA, at position 55 in the loop of TψC arm is important in maintaining the tertiary structure of tRNA[90]. Apart from

ψ, 1-methyladenosine (m^1A) at positions 9 and 55 also contribute to the stability of the L-shaped structure[91,92]. Due to the importance of post-transcriptional modifications in maintaining the structure of tRNA, hypomodified tRNAs are deemed to be less stable in cells. In fact, tRNAs lacking non-essential modifications are subjected to a rapid tRNA degradation pathway[93]. tRNA modifications can also affect tRNA cleavage and the biological activity of the cleavage products[94].

Besides being an adaptor for amino acids, emerging studies have shown that tRNAs have other non-canonical functions[95], one of which is that they can serve as precursors for another novel group of sncRNAs, broadly known as tsRNAs.

In the section below, I will discuss the biogenesis and functions of two major categories of tsRNAs: tRNA halves and tRNA-derived fragments (tRFs).

1.4.2 tRNA halves: biogenesis and functions

Originally detected in the bacterium *Escherichia coli*[96], tRNA fragments which result from cleavage of tRNA in the anticodon loop were also found in eukaryotes. In 2005, Lee and Collins reported that mature tRNAs, which have gone through 5' and 3' processing and base modification, are cleaved in their anticodon loops in response to starvation in the ciliate *Tetrahymena thermophila*[97]. Further investigations showed that RNaseA family protein angiogenin (ANG) cuts tRNA into tRNA halves, also known as tRNA-derived, stress-induced small RNAs (tiRNAs), in mammalian cells under stress conditions, promoting translational repression[98,99]. tRNA halves are around half the length of a mature tRNAs, ranging from 31 to 40 bases in length and can be subdivided into 5'- or 3'- tiRNAs, depending on their origins (Figure 1-4).

It was first proposed that cleavage of mature tRNAs into tRNA halves might lead to depletion of tRNAs but most reports have shown that the change of total tRNA levels is

marginal[100]. This finding suggests that tiRNAs are inherently biological active. Transfection of purified, endogenous 5'-tiRNAs into human bone cancer cell line U2OS can inhibit protein synthesis by approximately 20%. In contrast, transfection of 3' tiRNAs has little effect on translational efficiency, indicating the specificity of 5'-tiRNA in translational repression[98]. This observation is later elaborated by Ivanov *et al.* who showed that synthetic 5'-tiRNA^{Ala} and 5'-tiRNA^{Cys} can displace the translational initiation factor eIF4G from capped mRNA and form stress granules with YBX-1 protein, an RNA-binding protein[101].

1.4.3 tRNA-derived fragments

The increasing depth of whole-genome sequencing has enabled the discovery of tRFs. Previous reports have detected the presence of small RNAs originated from tRNA sequences by sequencing in mouse and humans[102–104]. However, these observations were not explored in detail until 2009 when three consecutive reports focused on tRFs were published[105–108]. Since then, tRFs have been receiving a great deal of attention and many studies have gradually shed light on the biogenesis and functions of these molecules[11,12,109–112].

1.4.3.1 Classification

Depending on their precursors and original positions within the tRNA, tRFs can be subdivided into four categories: tRF-5, tRF-3, tRF-1 and internal tRF (itRF).

tRF-5s are derived from the 5' end of tRNA, starting from the anticodon loop or the D loop till the 5' terminal. tRF-5s can be further classified into three main classes based on their lengths: tRF-5a, tRF-5b and tRF-5c, with tRF5-c being the longest (27 – 30 nt). tRF-3s originate from the 3' end of its mature precursor and are divided into two groups: tRF-3a (18 nt) and tRF-3b (20 nt). tRF-1 (or 3' U tRF) is a class of tRFs that come from the

3’ trailer of tRNA precursors. itRFs represent a group of tRFs corresponding to the region around the anticodon loop of mature tRNAs. Differences between the classes are illustrated in Figure 1-4.

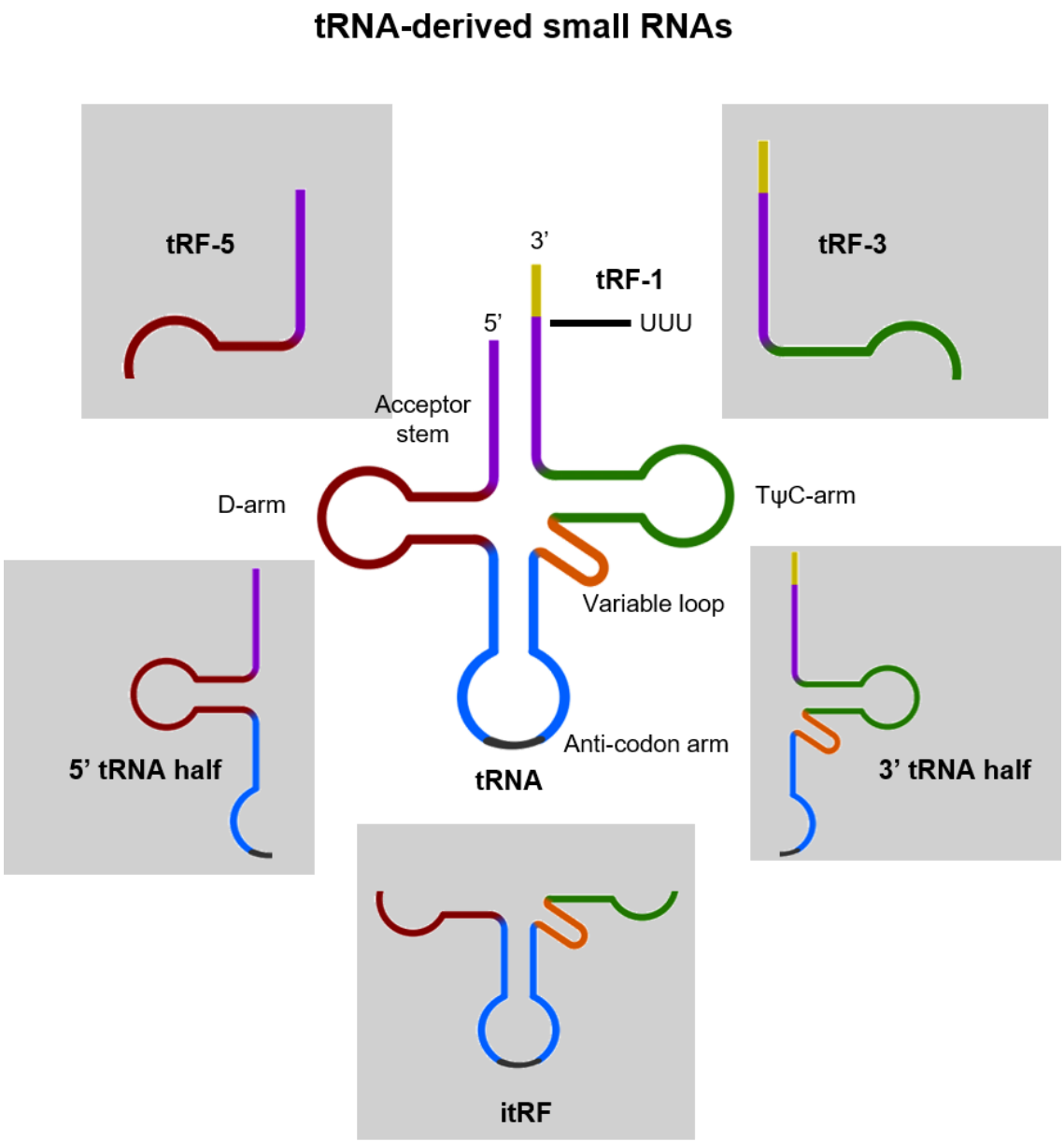


Figure 1-4 Classification of tRNA-derived small RNAs Diagram illustrating different classes of tRNA-derived small RNAs (tsRNAs) that have been identified. tsRNAs can be divided into two main categories: tRNA halves and tRNA-derived fragments (tRFs). tRNA halves are subdivided into 5’ and 3’ tRNA half depending on their origins. tRFs can be sub-divided into 4 classes. tRF-1 are derived from the 3’ trailer sequence of tRNA precursor while the other tRF classes are cleavage products of mature tRNAs. Components of the tRNAs are colour coded for clarity. Adapted from Su *et al.* (2020)[112].

1.4.3.2 Biogenesis

tRF-1 sequences correspond to the 3' trailer sequence removed by RNase Z, its 3' end matches the termination signal for RNA pol III, strongly suggesting that tRF-1s are generated during the maturation process of tRNA[105,113].

The biogenesis of the other types of tRFs is less clear. Cole *et al.* showed that a cytoplasmic tRF-5 derived from $tRNA^{Gln}$ is Dicer-dependent based on the evidence that the tRF is lost upon Dicer depletion in HeLa and HEK293 cells and that Dicer alone is sufficient to generate the tRF from $tRNA^{Gln}$ precursor[106]. CU1276, a 22-nt miRNA-like tRF-3 derived from the 3' end of $tRNA^{Gly}$ was described to be Dicer-dependent in another independent study[114]. Haussecker *et al.* described a group of Drosha-independent, Dicer-dependent tRF-3s (Type I) which associate with Ago proteins[107]. Evidence for Dicer-dependent tRFs are also found in other organisms. In zebrafish, tRF-5s derived from $tRNA^{Glu\text{-}CTC}$ and $tRNA^{Pro\text{-}CGG}$ are shown to be cleaved specifically by Dicer *in vitro*. Using RNA secondary structure prediction software RNAShapes, Soares *et al.* verified that 5% of $tRNA^{Glu}$ and 50% $tRNA^{Pro}$ are prone to form long hairpin structures, which can be recognised by Dicer[115]. In *Arabidopsis thaliana*, tRF-5 accumulation can be induced by the knockout of the chromatin remodelling protein Decrease in DNA Methylation 1 (DDM1); in *dcr1*/*ddm1* double mutants, the accumulation is reduced, implying that Dicer-like1 (DCL1) has a role in the biogenesis of tRF-5s[116].

On the other hand, several studies have claimed that Dicer is not necessary for the generation of most tRFs[117–119]. Meta-analysis of small RNA sequencing datasets from mouse, *S. pombe* and *Drosphilia* showed that mutation of Dicer does not influence the levels of tRFs in these organisms[117]. A subsequent report by the same group claimed that tRF-3 generation is intact in Dicer knockout HEK293T cells. However, the authors did not show or discuss whether the other tRF classes were affected upon full Dicer depletion.

As discussed above, Dicer is usually required for the loading of sRNAs onto Ago proteins; it is therefore unknown how mature tRFs are passed onto Ago, as many tRFs have been shown to be associated with Ago proteins[107,114,117,120].

tRNA modifications can also have an impact on the biogenesis of tRFs. PUS7, a ψ synthase, promotes the generation of a group of 18-nt tRF-5s which contain 5' terminal oligoguanine (originated from $tRNA^{Ala}$, $tRNA^{Cys}$ and $tRNA^{Val}$) in H9 human embryonic stem cells through the pseudouridylation at position 8[121]. The loss of function of a guanosine 9 tRNA methyltransferase TRMT10A precludes the biogenesis of tRF-5 from $tRNA^{Gln}$[122]. BCDIN3D, an RNA phosphomethyltransferase, negatively regulates the cleavage of $tRNA^{His}$ by Dicer, which produces a tRF-3 corresponding to a misannotated miRNA miR-4454[123].

1.4.3.3 Functions

In contrast to tRNA halves, tRFs have been reported to have diverse biological functions.

A tRF-5 from $tRNA^{Gln}$ was shown to inhibit protein translation without affecting mRNA abundance or polyadenylation in a sequence-independent manner. The exact mechanism, however, is unclear[124]. tRF-5s from $tRNA^{Ala}$, $tRNA^{Cys}$ and $tRNA^{Val}$, which is pseudouridylated at position 8, can potentially change the conformation of PABPC1 and precludes the recruitment of eIF4G to the cap complex, resulting in translational repression. This observation resonates with that of a report on displacement of eIF4G facilitated by 5'-tiRNAs[101]. Furthermore, tRFs play a role in regulating ribosome biogenesis. tRF-3 from $tRNA^{Leu\text{-}CAG}$ binds to the mRNAs of two ribosomal proteins, RPS28 and RPS15, promoting their translation. This in turn enhances processing of 18S ribosomal RNA and increases production of 40S ribosomal subunits. Interestingly,

transfection of antisense oligonucleotides against this tRF-3 can reduce tumour growth in a xenograft model, suggesting the importance of tRF regulation in cell proliferation[125].

The association of tRFs with Ago proteins implies that they may be involved in RNAi. Dicer-dependent (or Type I) tRFs co-immunoprecipitate with all four FLAG-tagged Ago proteins while RNaseZ-dependent, Dicer-independent tRFs show a preference for FLAG-tagged Ago3 and 4[107]. Meta-analysis of Ago PAR-CLIP data by Kumar *et al.* revealed that tRF-3s and tRF5-s strongly associate with Ago1, 3 and 4 but not Ago2[117]. In B cells, Dicer-dependent CU1276, a tRF-3, can downregulate endogenous RPA1 by interacting with its 3' UTR in a miRNA-like manner[114]. Additionally, a recent study by Kuscu *et al.* has shown that overexpression of tRNA can lead to increased levels of tRF-3s resulting in global silencing of mRNA targets whose 3' UTR sequences are complementary to at least the 6-mer sequence of the corresponding tRF-3. They also showed that tRF-3 associates with GW182, indicating the gene silencing is achieved by mRNA destabilisation[120].

tRFs can sequester RNA-binding proteins and prevent them from binding other RNAs. For instance, itRFs from $tRNA^{Glu}$, $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Tyr}$ share a common motif which can be recognised by YBX1. Under hypoxia, these itRFs sequester YBX1, leading to destabilisation of pro-oncogenic transcripts and suppression of growth[126].

tRFs are also involved in promoting retrotransposon silencing. tRF-3s found in preimplantation stem cells are involved in the silencing of endogenous retroviruses (ERV) in two distinct mechanisms: the 18-nt tRF-3s can block reverse transcription of ERV by acting as an inactive primer while 22-nt tRF-3s silence coding-competent ERVs in a miRNA-like manner[127].

Previously considered to be random degradation products, tsRNAs have gradually been recognised as *bona fide* regulatory sncRNAs. The discovery of tsRNAs has further expanded the sncRNA repertoire and painted a fuller picture of the regulatory networks mediated by RNA.

1.5 RNAi therapeutics

The story of regulatory sncRNAs does not end with the elucidation of their biogenesis and cellular functions. Given that sncRNA can function as potent gene silencing molecules, scientists have started exploring and harnessing their potential in the therapeutic context. Here in this section, I will provide a succinct update on the current developments in RNAi therapeutics.

The discovery of RNAi has instigated a great deal of interest in the pharmaceutical industry. The idea that a duplex RNA formulation can downregulate the expression of virtually every gene in the cell is deemed exciting and attractive, because such drugs potentially allow flexibility and programmability that very few can match.

However, drug developers face major challenges as initial clinical trials showed that siRNA drugs can have toxic effects in patients. dsRNAs can potentially induce immunogenic reaction through activating toll-like receptors (TLRs) which in turn causes rapid production of pro-inflammatory cytokines and interferon response. Toxicity of excipients, the inactive substances which act as the vehicle for a drug, also poses an issue to siRNA drug delivery. Furthermore, the flexibility of RNAi also comes with promiscuity, resulting in off-target effects. Another major issue faced by drug developers is the systemic effect of siRNA drugs; an siRNA drug must be delivered to the right cells for maximum effect. A wide variety of methods are being developed to overcome each of these challenges[128].

In August 2018, the first RNAi drug Onpattro® (patisiran) by Alnylam Pharmaceuticals was approved by the United States Federal Food and Drug Administration (FDA) for the treatment of hereditary transthyretin amyloidosis (hATTR). hATTR is caused by hereditary mutations in transthyretin (*TTR*) leading to deposition of TTR amyloids and has poor prognosis with a median survival of 5-15 years from diagnosis[129]. Patisiran works by targeting the less variable 3' UTR region of *TTR* mRNA to silence both wild type and mutant *TTR* to reduce serum level of TTR and halt the progression of hATTR[130]. The approval of patisiran has provided hope for safe and efficient siRNA therapeutics.

Another notable development within the field is the use of GalNAc conjugation as an efficient liver-targeted delivery platform. GalNAc is a sugar derivative of galactose and a ligand for asialoglycoprotein receptors (ASGPRs), which are highly expressed on hepatocytes[131]. At the time of writing, about one third of siRNA drugs in clinical trials are GalNAc based.

In the future, RNAi therapeutics is expected to gradually move beyond the liver by harnessing other metabolic stabilisation methods to target other tissues. Antibody – siRNA conjugates is another potent strategy in development for targeted delivery. For systemic delivery, exosomes have shown to be effective vehicles for the delivery of anti-KRAS siRNA against pancreatic cancers[132] Besides siRNA, miRNA mimics, inhibitors and agonists are also being considered as potent drug candidates[133].

2 Materials and Methods

2.1 Tissue culture

2.1.1 List of cell lines

Cell line	Organism	Tissue	Source
HEK293T	*Homo sapiens*, human	Embryonic kidney	ATCC® (CRL-3216™)
Dicer-kd/2b2[134] **(shDicer)**	*Homo sapiens*, human	Embryonic kidney	Filipowicz Lab
Ago2-kd[134] **(shAgo2)**	*Homo sapiens*, human	Embryonic kidney	Filipowicz Lab
BT549	*Homo sapiens*, human	Mammary gland; breast	ATCC® (HTB-122™)
Ago$^{-/-}$ MEF[135]	*Mus musculus*, mouse	Embryo	O' Carroll Lab
MCF7	*Homo sapiens*, human	Mammary gland; breast; derived from metastatic site: pleural effusion	ATCC® (HTB-22™)
HepG2	*Homo sapiens*, human	Liver	ATCC® (HB-8065™)
BWGT3	*Mus musculus*, mouse	Liver	Oxford Cell Bank

2.1.2 Culture conditions

All cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific) at 5% CO_2 and 37 °C.

2.2 Experimental techniques

2.2.1 RNA purification

Total RNA was isolated from whole-cell, cytoplasmic, nuclear or chromatin extracts using TRIzol LS Reagent (Invitrogen) according to manufacturer's instructions followed by treatment with DNase I (1 U, Roche) for 30 min at 37 °C. RNA sample concentration was measured by the NanoDrop™ Lite Spectrophotometer (Thermo Fisher Scientific).

2.2.2 Northern blot

For detection of tsRNAs, 20 µg of total RNA in 2x Novex TBE-Urea Sample Buffer (2x) (Thermo Fisher Scientific) were separated on 12% denaturing (50% g/v urea) polyacrylamide gel in 1x TBE. RNA was transferred onto hybond-N+ nylon membrane at 5 V for 30 min followed by UV-crosslinking and pre-hybridisation in ULTRAhyb-Oligo buffer (Thermo Fisher Scientific) at 42 °C for at least 30 min. Target-specific DNA probes were radiolabelled with ^{32}P-ATP (Perkin Elmer) by polynucleotide kinase (PNK) (NEB) at 37 °C for 30 min. Radio-labelled probes were purified in G-25 Sephadex columns (GE Healthcare) and hybridised onto the membrane at 42 °C followed by washes with 0.1x SCC washing buffer (0.05% SDS) and subjected to autoradiography.

2.2.3 Western blot

Whole-cell, cytoplasmic, nuclear or chromatin extracts were lysed in RIPA buffer (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris (pH 8.0)) and protein concentrations were normalised using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Samples were later treated with Laemmli buffer (0.2 M Tris-

HCl, 8% (w/v) SDS, 40% glycerol, 20% (v/v) β-Mercaptoethanol, 0.005% bromophenol blue), incubated at 95 °C for 8 min and sonicated. Protein samples were separated on mini-PROTEAN® TGX™ gels (Bio-Rad Laboratories) followed by transfer onto a nitrocellulose membrane (Protran, GE Healthcare) and probed with primary and secondary antibodies at varying concentrations (Section 2.5). Signals were visualised on Amersham™ Hyperfilm™ ECL(Cytiva) after the application of Pierce™ ECL Western Blotting Substrate (Thermo Scientific).

2.2.4 sRNA-seq and RNA-seq sample preparation

For small RNA sequencing (sRNA-seq), total RNA was isolated from cells treated with scrambled shRNA (as control) and shDicer cells for 7 d using the miRVana miRNA Isolation Kit (Thermo Fisher Scientific). Quality of purified RNA was confirmed with RNA 6000 Pico Kit (Agilent) on the Agilent 2100 Bioanalyzer. Sequencing libraries were prepared using the NEBNext® Multiplex Small RNA Library Prep Set (New England BioLabs) and sequenced on a HiSeq2000 (Illumina).

RNA was purified using the miRNEasy Kit (Qiagen) and treated with DNase (Thermo Fisher Scientific) for 37 ºC for 30 min followed by acidic phenol-chloroform extraction for poly(A) RNA sequencing (RNA-seq). Sample integrity was verified with a 1.25% formaldehyde gel. RNA samples were ribo-depleted and sequencing libraries preparation was performed with the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) followed by paired-end sequencing on HiSeq2000 (Illumina).

2.2.5 *In vitro* Dicer cleavage assay

TAP-tagged Dicer protein expression was induced with 3 µg/mL doxycycline for 1 week in HEK293T cells. 10^8 cells were harvested, washed twice in PBS, and lysed for 30 min at 4 °C in lysis buffer (20 mM TRIS-HCl pH 7.6, 137 mM NaCl, 3 mM MgCl2, 0.5%

NP-40, 5% glycerol and proteases inhibitor). Insoluble cellular debris were pelleted by centrifugation and the lysate was incubated with 50 µL IgG Sepharose 6 Fast Flow affinity resin (GE Healthcare) for 1.5 h at 4 °C. After binding reaction, Sepharose beads were extensively washed before proceeding with the elution.

TAP-tagged Dicer was eluted by using AcTEV protease (Invitrogen) following manufacturer's instructions, then purified Dicer protein was resuspended in Dicer reaction buffer (50 mM TRIS-HCl pH8, 20 mM HEPES, 5 mM MgCl2, 5% glycerol with proteases inhibitor). 500 ng of purified RNA was incubated with Dicer at 37 °C. At appropriated times, aliquots of the reaction mixtures were separated, added to a Novex TBE-Urea loading buffer, and immediately frozen at −80 °C. All samples were heated for 3 min at 95 °C immediately before gel electrophoresis. Cleavage products were separated on a 10% denaturing polyacrylamide gel and visualised by SYBR Gold (Invitrogen) staining.

2.2.6 qRT-PCR

Purified RNA (250 ng and 1 µg for exon- and intron-based qRT-PCR, respectively) was used for preparing cDNA template using SuperScript™ Reverse Transcriptase (Thermo Fisher Scientific) with specific reverse primers (Section 2.6.2). Real-time PCR was performed on Rotor-Gene RG3000 machine (Corbett Research) with SensiMix™ SYBR No-Rox Mastermix (Bioline Reagents) with specific primer pairs (Section 2.6.2). Relative fold change was computed using the comparative C_t method[136].

2.2.7 Subcellular fractionation

3 x 10^6 cells were trypsinised, washed twice in ice-cold PBS and lysed in five volumes of hypotonic lysis buffer (10 mM HEPES pH 7.0, 60 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 0.075% NP-40, 1x protease inhibitor cocktail (Roche)) for 10 min

on ice. Lysates were centrifuged at 500x g for 5 min and the resulting supernatant was collected as the cytoplasmic fraction. The nuclear pellet was washed gently for 3 times in 800 µl hypotonic lysis buffer without NP-40. Nuclei were lysed in one volume of nuclear lysis buffer (20 mM HEPES pH 7.9, 400 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT, 5% glycerol, 1x protease inhibitor cocktail (Roche)) and diluted with two volumes of dilution buffer (20 mM HEPES pH 7.9, 1.5% Triton-X-100, 0.2% sodium deoxycholate, 1x protease inhibitor cocktail (Roche)).

2.2.8 Immunoprecipitation

Cells were scrapped from 10-cm dishes, washed with ice-cold PBS, and lysed in hypotonic lysis buffer (10 mM HEPES pH 7.0, 60 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 0.075% NP-40, 1x protease inhibitor cocktail (Roche)) for 10 min on ice. Lysates were centrifuged at 500x g for 5 min and the supernatant was collected as the cytoplasmic fraction. The nuclear pellet was washed gently for 3 times in 800 µl hypotonic lysis buffer without NP-40. Nuclei were lysed in one volume of nuclear lysis buffer (20 mM HEPES pH 7.9, 400 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT, 5% glycerol, 1x protease inhibitor cocktail (Roche)). Immunoprecipitation with antibodies was performed overnight and samples were incubated with protein G magnetic beads (40 µl per sample, Invitrogen) for 1 h. For Ago2 immunoprecipitation, protein-RNA complexes were directly subjected to TRIzol treatment followed by northern blotting.

2.2.9 ChIP-qPCR

Approximately 7 x 10^6 cells were incubated with 1% formaldehyde in DMEM for 8 min followed by quenching with 0.125 M glycine in DMEM for 10 min at 37 °C. Cells were washed with ice-cold PBS and lysed in 500 µl cell lysis buffer (0.5% NP-40, 85 mM KCl, 5 mM PIPES, 1x protease inhibitor cocktail (Roche)). Chromatin was pelleted at 800x g

for 10 min and lysed in nuclear lysis buffer (50 mM Tris-HCl, 1% SDS, 10 mM EDTA, 1x protease inhibitor cocktail (Roche)) and sonicated at high power settings for 25 min at 4 °C. The fragmented chromatin lysate was pre-cleared with protein G magnetic beads (40 µl per sample, Invitrogen) for 1 h, divided equally into input, IP and beads only samples and diluted in dilution buffer (16.7 mM Tris-HCl, 0.01% SDS, 1.1% Triton X-100, 500 mM EDTA, 167 mM NaCl, 1x protease inhibitor cocktail (Roche)). Immunoprecipitation with antibodies was performed overnight and samples were incubated with protein G magnetic beads (40 µl per sample, Invitrogen) for 1 h. Beads were washed with washing buffers A (20 mM Tris-HCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 150 mM NaCl), B (20 mM Tris-HCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 500 mM NaCl), C (10 mM Tris-HCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate, 0.25 M LiCl) and D (10 mM Tris-HCl, 1 mM EDTA). Protein-DNA complexes were eluted with elution buffer (1% SDS, 0.1 M NaHCO3) for 30 min at room temperature and treated with with RNase A (1 µl) and proteinase K (2 µl) at 65 ºC overnight. DNA was extracted with phenol-chloroform-isoamyl alcohol (25:24:1) mix and used followed by qPCR.

2.2.10 chrRNA-seq sample preparation

Chromatin fraction was extracted using approximately 6.72 x 10^6 cells according to a published protocol[137] and treated with 40 µg of proteinase K in 1% SDS and 1 µl of Turbo DNase (2 U/µl) (Thermo Fisher Scientific) followed by TRIzol (Invitrogen) extraction. Incompletely dissolved chromatin pellet was dissolved by heating the samples at 55 ºC for 10 min on a heat block in safe lock tubes (Eppendorf). RNA samples were ribo-depleted and sequencing libraries preparation was performed with the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) followed by paired-end sequencing on HiSeq2000 (Illumina).

2.2.11 Labelling of small RNAs with fluorescent dye

T7-transcribed tsRNA (5'-UCCCUGGUGGUCUAGUGGUUAG-3') and siRNA against *EGFR* (5'-GAGGAAAUAUGUACUACGAUU-3') were labelled with Alexa Fluor® 488 fluorescent dye by FISH Tag RNA Multicolor Kit (Invitrogen) according to manufacturer's instructions. Cells were fixed by paraformaldehyde (3%) 24h post-transfection followed by confocal imaging. All images were captured using the same settings.

2.2.12 CRISPR-Cas9 knockout of *SPINT1* target site

2.2.12.1 RNP formation

For ribonucleoprotein (RNP) formation, 1 µg of each IVT sgRNA was added to 5 µg Cas9 HiFi protein (IDT) in a total volume of 2 µl IDT Duplex Buffer (IDT, cat. #11-05-01-12), resuspended 5 times and incubated at 37 ºC for 5 min. Resulting RNP mixes were added directly to cells prior to nucleofection.

2.2.12.2 RNP nucleofection

Cells (2 x 10^5) were washed twice with PBS and nucleofected with 5 µg Cas9 protein (IDT) and 1 µg sgRNA in a total volume of 10 µl "buffer R" using the Neon Transfection System (Invitrogen, cat. #MPK5000).

2.2.12.3 Materials

Neon Transfection System (Invitrogen, cat. #MPK5000) and consumables Neon Transfection System 10 µL Kit (Invitrogen, cat. #MPK1096).

Buffers etc. for nucleofection buffer R from Neon Transfection System 10 µL Kit (Invitrogen, cat. #MPK1096).

2.2.12.4 In vitro transcription of sgRNAs

Single guide RNAs (sgRNAs) were *in vitro* transcribed from double-stranded DNA templates as previously described by Gagnon *et al.*[138]. In brief, 60-mer DNA oligos harbouring an 18-20-mer protospacer sequence flanked 5' (upstream) by a T7 promoter sequence (TAATACGACTCACTATAGG) and 3' (downstream) by part of the conserved tracrRNA domain sequence (GTTTTAGAGCTAGAAATAGCAA) was annealed to a universal 80-mer oligo (AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTA TTTTAACTTGCTATTTCTAGCTCTAAAAC) harbouring the remainder of the tracrRNA and gap-filled using T4 DNA polymerase (NEB). The resulting double-stranded DNA template was column purified using a DNA Clean & Concentrator kit (Zymo Research, cat. #D4013) and 500 ng used in a 30 µl IVT reaction as per manufacturer's protocol (HiScribe T7 High Yield RNA Synthesis Kit, NEB, cat. #E2040S). After 4 h incubation at 37ºC, remaining DNA template was degraded by adding 2 ul DNAseI and 18ul water to the IVT reaction. Finally, sgRNAs were purified using a MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific, cat. # AM1908) according to the manufacturer's instructions and eluted in 25 µl nuclease-free water. Guide RNA concentration and purity was measured using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

2.2.12.5 Generation of knockout cell lines

Cells (2 x 10^5) were nucleofected with 5 µg Cas9 protein (IDT) and 1 µg sgRNA in a total volume of 10 µl "buffer R" using the Neon Transfection System (Invitrogen, cat. #MPK5000) recommended protocols for RKO and SH-SY5Y cell lines.

For each gene target, multiple guide sequences were used to generate the sgRNAs as follows:

gRNA_SPINT1_in1 TAATACGACTCACTATAGG ACCTTCTGTGAGTCCCCCAT GTTTTAGAGCTAGAAATAGCAA

gRNA_SPINT1_in2 TAATACGACTCACTATAGG AACTCTCTAGAGAGCCTCTG GTTTTAGAGCTAGAAATAGCAA

2.2.12.6 Screening of mutant cell line

Cells subjected to nucleofection were sorted into single cells onto 96-well plates for recovery. Genomic DNA was extracted from single cell pellets by phenol:chloroform:isoamyl alcohol (25:24:1) extraction followed by PCR using Phusion High-Fidelity PCR Master Mix (NEB) (forward primer: CAGCCACTTTTCGTTCCTGC; reverse primer: AACCATGCTACCTTAGGGTTTAT). PCR products were subjected to agarose gel electrophoresis, visualised by UV transilluminator and Sanger sequenced.

2.2.13 *In vitro* Argonaute 2 cleavage assay

SPINT1 RNA (hg38, chr15:40,847,819-40,848,367) was prepared using HiScribe™ T7 High Yield RNA Synthesis Kit (NEB) followed by DNAse I RNAse free (NEB) treatment, purified by phenol chloroform extraction and ethanol precipitation. 50 ng of recombinant Ago2 or Ago3 protein (Active Motif) were pre-incubated with 500 nM single-stranded tsRNA in cleavage buffer (HEPES-KOH 25 mM pH 7.5; $MgCl_2$ 5 mM; KCl 50 mM; DTT 0.5 mM; EDTA 0.2 mM; ATP 0.1 mM; GTP 0.02 mM; BSA 0.05 µg/µl; Ribolock 20 U; Protease inhibitors EDTA free 1x) for 2 h at 37 °C and followed by adding 400 ng of *SPINT1* RNA. Cleavage reactions were quenched by TBE-Urea loading buffer after 0, 3 h, 6 h, 24 h incubation at 37°C, respectively. Cleavage products were resolved by 4% polyacrylamide 7 M urea denaturing gel at 400 V in 1x TBE buffer. For blotting cleavage products were transferred onto Hybond N+ positive charged nylon

membranes (GE Healthcare) by semidry electroblotting in 0.5x TBE buffer at 300 mA for 60 min. After UV crosslinking at 120 mJ/cm^2 in a Stratalinker UV crosslinker, the membrane was blocked with ULTRAhyb Hybridization Buffer (Invitrogen) for 30 min at 45°C and then incubated with 5'-32P labelled probe (5'-TGATAGATATTGCCCAACTACC-3') overnight at 45 °C. The blot was washed twice with 0.1x SSC at 45 °C for 15 min. The blot was exposed to Hyperfilm (GE Healthcare) for 2 days at −70 °C.

2.3 Statistical analysis of real time PCR data

Real time PCR (qPCR) data were analysed using raw take-off (C_t) values (median of two technical replicates). When comparing two conditions, data were subjected to Shapiro-Wilk test and F test to assess for normality and equal variance, respectively. If the two conditions followed a normal distribution and had the same variance, unpaired, one-tailed t-test was performed to test for significant difference (*P*-value < 0.05 is considered as significant). If data did not follow a normal distribution, unpaired Mann Whitney test (one-tailed) was used instead. For comparison of two or more conditions, one-way ANOVA was performed followed by Tukey multiple comparisons test if the mean of every group was compared to other groups. In cases where the mean of every experimental group is compared to the control group, a Dunnett multiple comparisons test was used instead.

2.4 Bioinformatics methods

2.4.1 sRNA-seq and PAR-CLIP analyses

A bowtie index was built using bowtie-build[139] for tRNA gene sequences extended by 7 nt on each side. miRNA and snoRNA sequences were downloaded from UCSC and an index was built the same way.

sRNA-seq data for Dicer knockdown and scrambled shRNA control (3 biological replicates) and photoactivable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) data for AGO 1, 2, 3 and 4[140] and Dicer[141] were adapter trimmed using cutadapt 1.8.3 with --minimum-length 10. Further sequences consisting of partial adapters were removed using a custom written perl script. The remaining sequences were mapped to tRNAs ±7 nt/mi- & snoRNAs using bowtie *–S –v3 --all --best –strata*[139].

Only sequences between 19 nt and 22 nt in length were considered for further analysis. Reads were equalized to their genomic sequence.

For sRNA analyses, reads are supported by at least one AGO and one Dicer PAR-CLIP hit in either replicate. Hits in tRNA/miRNA/snoRNA regions were normalized to total number of mappable reads to the genome. These were determined by mapping reads with bowtie[139] *-m 1 -k 1* to hg38 and adding up "reads with at least one reported alignment" and "reads with alignments suppressed due to -m" from the output report.

For PAR-CLIP reads occurring 25 times in all AGO sets and 323, 41 or 19 times in either Dicer rep1, rep2 or rep3 sets respectively were considered further (cut-offs were due to different library sizes and distribution of read occurrence). These were considered to be tsRNAs.

2.4.2 chrRNA-seq analysis

cDNA and ncRNA sequence data were downloaded from Ensembl version 89[142] and a kallisto (v0.43.1) index was built. Read counts for RNA-seq data were generated using kallisto with the following options: *--rf-stranded -b 100 -t 5*.

2.4.3 Differential gene expression analysis

Differentially expressed genes were determined with DESeq2[143]. For mRNA-seq genes with the FDR adjusted $P < 0.001$ were considered as significantly differentially expressed. For chrRNA-seq, genes with the adjusted $P < 0.005$ for shDicer and shAgo2 were considered as significantly differentially expressed. Due to the low number of changing genes, a less stringent criterion of adjusted $P < 0.05$ was used in shDrosha samples.

2.4.4 tsRNA target gene prediction

tsRNA targets were predicted by running miRanda 3.3a[47,144] with the parameters *-sc 150 -en -30 -quiet* against significantly upregulated genes determined from mRNA-Seq. The same analysis was repeated for genes significantly upregulated in shDicer and shAgo2 in chrRNA-seq.

2.4.5 Disease-gene association heatmap

The table for all gene-disease associations was downloaded as well as a curated annotation of CUI, disease and disease classes from DisGeNET[145,146]. DisGeNET uses NCBI annotation as reference. The gene symbols and all synonyms were extracted for NCBI genes. Gene symbols for Ensembl 89 were extracted using BioMart. The gene universe was taken as the overlap of NCBI gene symbols and Ensembl 89 gene symbols and only disease and target genes with symbols in this universe were considered further. A contingency table was thus constructed, and a one-sided Fisher exact test was employed to assess the significance of the observation.

Heatmaps were plotted in R using the pheatmap package by constructing a binary matrix for target-gene – disease/disease class associations. This matrix was ordered first by column sums (genes), then by row sums (diseases).

2.4.6 Data and materials availability

All data are available in the supplement; raw data are deposited in GEO accession number GSE126751. Combined RNA pol II ChIP-seq and chrRNA-seq data can be viewed here:

https://genome-euro.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr12%3A6533927-6538371&hgsid=245170476_lUdLtIIcI6k0WJUqnqBaj06rtntf

Custom written codes are available in Github: https://github.com/rita1510/tRNA_codes.git

2.5 List of antibodies

Anti-	**Usage**	**Concentration**	**Manufacturer (catalogue number)**
Dicer (A-2)	WB	1:250	Santa Cruz Biotechnology (sc-136981)
p63 [EPR5701]	WB	1:1000	Abcam (ab124762)
Drosha	WB	1:500	Abcam (ab12286)
Ago2	WB	1:500	Abcam (ab32318)
Ago2	IP	5 µg	Abcam (ab57113)
β-tubulin	WB	1:1000	Abcam (ab6046)
H3	WB/ChIP	1:1000/5 µg	Abcam (ab1791)
RNA pol II (8WG16)	WB/ChIP	1:1000/5 µg	Abcam (ab817)
RNA pol II (S2P)	WB/ChIP	1:1000/5 µg	Abcam (ab5095)

RNA pol II (S5P)	WB/ChIP	1:1000/5 µg	Abcam (ab5131)
H3K9me2	WB/ChIP	1:1000/5 µg	Abcam (ab1220)
SPINT1	WB	1:500	Abcam (ab189511)
EGFR/Egfr	WB	1:500	Abcam (ab52894)
Cleaved Caspase-9	WB	1:1000	Cell Signaling Technology (9508)
BCL2	WB	1:1000	Abcam (ab32124)

2.6 List of oligonucleotides

2.6.1 Northern blot

Target sRNA	**DNA probe (5'-3')**
tRF-5$^{Ala-ACG}$ / tRNA$^{Ala-ACG}$	TACCACTGAGCTACACCCCC
tRF-5$^{Glu-CTC}$ / tRNA$^{Glu-CTC}$	CTAACCACTAGACCACCAGGGA
itRF$^{Arg-CCT}$ / tRNA$^{Arg-CCT}$	CGAACCCACAATCCCTGGCTTA
tRF-3^{Arg} / tRNA^{Arg}	TGCCTTATCCATTAGGCCAC
let7a-5p / pre-let7a	AACTATACAACCTACTACCTCA
snR38A	GATAACTAACCATCATAGGC
Cleaved RNA (*in vitro* Ago2 assay)	TGATAGATATTGCCCAACTACC

2.6.2 qRT-PCR primers

Target gene		**Sequence (5'-3')**	**Target region**
GAPDH	Forward	AACCTGCCAAATATGATGAC	Exon 8
	Reverse	AGGAAATGAGCTTGACAAAG	

HPRT1	Forward	AGATGTGATGAAGGAGATGG	Exon 3
	Reverse	AATAGCTCTTCAGTCTGATAAAATC	
CDKN1B	Forward	TGTTAAGAATTGACCATCTGC	Last exon
	Reverse	TTTCCTGTGGATAAATGGAC	
ETNK1	Forward	GGTGGATCAATCCCTTAGTC	Intron 1
	Reverse	AATCTGCAGTTGCTTGAATC	
NOV	Forward	CTCATGCTGTCTGGTGTG	3’ UTR
	Reverse	CAGATGCCAGTCTGGTTG	
	Forward	ATGCAATACCCAATTTGAAC	Intron 4
	Reverse	ATTGCCCAAGATTACATAGC	
GK	Forward	GAAGAATTCGGATCCTTACC	3’ UTR
	Reverse	TCAGTGAAGATGTTGCTGTC	
	Forward	TTTATAACCCCTGCTGTTTG	Intron 4
	Reverse	GCCACATTTGTAGAAGAAGC	
GUCY1A2	Forward	TCTTTTTGCCACCTTCTAAC	Last exon
	Reverse	GAATAGCAGTCGGCTCTTAG	
	Forward	GTCAATGCCTCTGATAATGC	Intron 1
	Reverse	TTCTGAAAACCATTCTCTGG	
RP4-639F20.1	Forward	CCTGGCTGAATTTTATTTTG	Last exon
	Reverse	TCTGTGTTCTTTTCCTGGAG	
	Forward	CCTGGCTGAATTTTATTTTG	Intron 2
	Reverse	TCTGTGTTCTTTTCCTGGAG	
RPB7	Forward	AAACAGACATTCCAGAGAGC	3’ UTR
	Reverse	TTGGACAGGTTTATTGAAGC	

	Forward	CTTGCCCTTGAATTCTTTC	Intron 1
	Reverse	GAAATCTCGCTACTGCACTC	
SPINT1	Forward	TTGTTTCCTCTGTTCAAAGC	3’ UTR
	Reverse	TGCCACAGCATTAAATAAAC	
	Forward	CACTACCATAAATGGGAAGC	Intron 6
	Reverse	TTTGTAGAGATGGGGTTTTG	
	Forward	TGTAGGCCTCTGAAACATTC	Target site
	Reverse	TGCAAAGAAAGAAGAGAAGG	
	Forward	TTGGGGTAATAAACCTGATG	Intron 2
	Reverse	CAAAGTTCTGGGATTACAGG	
hAgo2-FLAG	Forward	GCCAGCATACTACGCTCACC	
	Reverse	TCGTCCTTATAGTCCTTATCGTCG	
BCL2	Forward	CACAAGTGAAGTCAACATGC	Last exon
	Reverse	GGCAGTAAATAGCTGATTCG	
	Forward	ACTCAATCGAGGAGTGTGAC	Intron 1
	Reverse	CATGGTGTAGGAGAGAGTGC	
EGFR	Forward	CTGACTACCAGCAGGACTTC	3’ UTR
	Reverse	CTGGGTATCGAAAGAGTCTG	
TTR	Forward	CTGCTGAGCCCCTACTCCTA	Last exon
Gapdh	Forward	GCAGTGGCAAAGTGGAGATTG	Exon 3
	Reverse	GCCGTTGAATTTGCCGTGAG	
Hprt	Forward	AGGAGTCCTGTTGATGTTGCC	3’ UTR
	Reverse	CTACCAGAGGGTAGGCTGGC	
Egfr	Forward	TTTCATGAGGTAACTCCAGGGC	Intron 21

	Reverse	GCCTATGCTCCCTCTCGGTA	
	Reverse	TGAAATCCCATCCCTCGTCC	
Ttr	Forward	CCGTGTTAGCAGCTCAGGAA	Last exon
	Reverse	TGAACGGTTGGTCCACTCTG	

2.6.3 ChIP-qPCR primers

Target gene		**Sequence (5'-3')**	**Target region**
GK	Forward	ATCCAAAGCAGTTAATGGAG	TSS -5' UTR
	Reverse	TTTCGGTTTAGCATTTAAGG	
	Forward	TTTGCTAACTGAACTTCACAAC	Intron 4 – exon 5
	Reverse	AAGCCATGAATCAATAGTCC	
	Forward	TGGGATGACAGAATTACCTC	3' UTR
	Reverse	AGGCAATCTTTTCTTCACTG	
	Forward	GCACAGTGGCTTAAACGTC	Target site
	Reverse	CCAAGTAGCTGGGACTACAG	
GUCY1A2	Forward	ACTCCCAATGAGTATGATG	TSS – 5' UTR
	Reverse	GCTCCGAGAGTGTGTGAC	
	Forward	GTTACAGGGATGCAGAAAAG	Exon 3
	Reverse	ACTGAAGAATTCCTGAGACAT	
	Forward	AAAGGTTTCCTACAACATCG	Last exon – 3' UTR
	Reverse	AGAAATTTTGCTGCTTGTTC	
	Forward	CCAAGTTCCCGCCATGAG	Target site
	Reverse	GCAAGCGACAACGTTAAG	
SPINT1	Forward	TCAGTTTCACCAGAAACCAG	5' UTR – intron 1

	Reverse	CACGAGCGCAGAACAAAG	
	Forward	GAGCTACCTTCCTGGAGTC	Exon 2
	Reverse	CGTGAGGTAGTTGATGAAGC	
	Forward	AAAAACTTTGGAACCAGACTC	3’ UTR
	Reverse	AACTCCAACTTCCTCCAAAC	
	Forward	TCCTACCTCCATGTTAGGG	Target site
	Reverse	GGAGGTAGCTTGAACAAATG	
NOV	Forward	GAGTGGTTTCTCCTTGTCTC	Target site
	Reverse	ACACACCAGACAGCATGAG	
RPB7	Forward	GGTCTCAAATTCCTGACCTC	Target site
	Reverse	AGGCTGGTATTAAACTCCTG	

2.6.4 T7 transcription

sRNA	**Strand**	**Sequence (5’-3’)**
T7-tsEGFR / itRF$^{Glu-CCT}$	Coding	TAATACGACTCACTATAGTCCCTGGTGGTCTAGTGGTTAG
	Template	CTAACCACTAGACCACCAGGGACTATAGTGAGTCGTATTA
T7-siEGFR	Coding	TAATACGACTCACTATAGGAGGAAATATGTACTACGATT
	Template	AATCGTAGTACATATTTCCTCCTATAGTGAGTCGTATTA

T7-tRNA^{Arg}	Coding	TAATACGACTCACTATAGGCCCCAGTGGCCTA ATGGATAAGGCACTGGCCTCCTAAGCCAGGG ATTGTGGGTTCGAGTCCCACCTGGGGTG
	Template	CACCCCAGGTGGGACTCGAACCCACAATCCCT GGCTTAGGAGGCCAGTGCCTTATCCATTAGGC CACTGGGGCCTATAGTGAGTCGTATTA

2.6.5 tsRNA mimics

Name	Class	Sequence	Origin
tsSPINT1	tRF-3	AGUCUCGGUGGAACCUCCA	tRNA$^{Gln-CTG}$
tsGK	itRF	CACCCUGGACUCUGAAUCCAG	tRNA$^{Gln-CTG}$
tsBCL2	itRF (intronic)	UAAGCCAGGGAUUGUGGGUUCG	tRNA$^{Arg-CCT}$
	tRF-5 (exonic)	AGGGGGUGUAGAUCAGUGGUAG	tRNA$^{Arg-AGC}$
tsEGFR	tRF-5	UCCCUGGUGGUCUAGUGGUUAG	tRNA$^{Glu-CTC}$
tsTTR candidates	tRF-5	UGGGGGUGUAGCUCAGUGGUAG	tRNA$^{Val-AAC}$
	tRF-5	GGGGGCAUAGCUCAGUGGUAGA	tRNA$^{Cys-GCA}$
	itRF	UGGCCUCCUAAGCCAGGGAUUG	tRNA$^{Arg-CCT}$
	itRF	GGGGAUUGUGGGUUCGAGUCC	tRNA$^{Arg-CCG}$
	itRF	CACUCUGGACUUUGAAUCCAG	tRNA$^{Gln-TTG}$
	tRF-5	GGGGGGUGUAGCUCAGUGGUAG	tRNA$^{Ala-AGC}$
	itRF	UGCAGUGGUAGAAUUCUCG	tRNA$^{Gly-CCC}$
	itRF	UAGUGGCUAGGAUUCGGCG	tRNA$^{Glu-TTC}$

	tRF-3	CAAAUCUCGGUGGAACCUGCA	$tRNA^{Gln\text{-}CTG}$
	tRF-5	UCCCUGGUGGUCUAGUGGUUAG	$tRNA^{Glu\text{-}CTC}$

3 Biogenesis of tsRNAs and prediction of target genes

3.1 Introduction

The involvement of Dicer in the biogenesis of tsRNA has been implicated in multiple reports[106,107,115]. However, none of them showed definitively that Dicer processes tRNAs into tsRNAs. Subsequent meta-analyses of published bioinformatic data have suggested that Dicer is dispensable for the biogenesis of most tsRNAs in mouse, fission yeast and fruitfly[117,118]. Intrigued by these discrepancies, I decided to study the role of Dicer in the generation of tsRNA.

In addition to tsRNA biogenesis, I was keen on investigating the role of tsRNA in the regulation of gene expression. Previous reports have shown that tsRNA can regulate gene expression in a miRNA-like manner[120]. However, the decision to approach this investigation in an unbiased manner, which is by employing bioinformatic tools to generate a list of predicted tsRNA target genes, was taken. With the list of tsRNA target genes, I ventured into understanding how they are regulated.

3.2 Dicer is involved in tsRNA biogenesis

If Dicer is responsible for tsRNA biogenesis, depletion of Dicer should result in the loss of tsRNA molecules. Small RNA-sequencing (sRNA-seq) was performed to compare global small RNA (sRNA) levels in wild type HEK293T and Dicer knockdown cells. The amount of sRNA groups, namely tsRNAs, miRNAs and snoRNAs, were assessed. Global levels of tsRNAs were significantly decreased in Dicer knockdown cells. As expected, levels of miRNAs, which are generated by Dicer, were also significantly reduced. In contrast, the amount of snoRNAs, which acted as a negative control for this experiment, remained unchanged upon Dicer depletion (Figure 3-1).

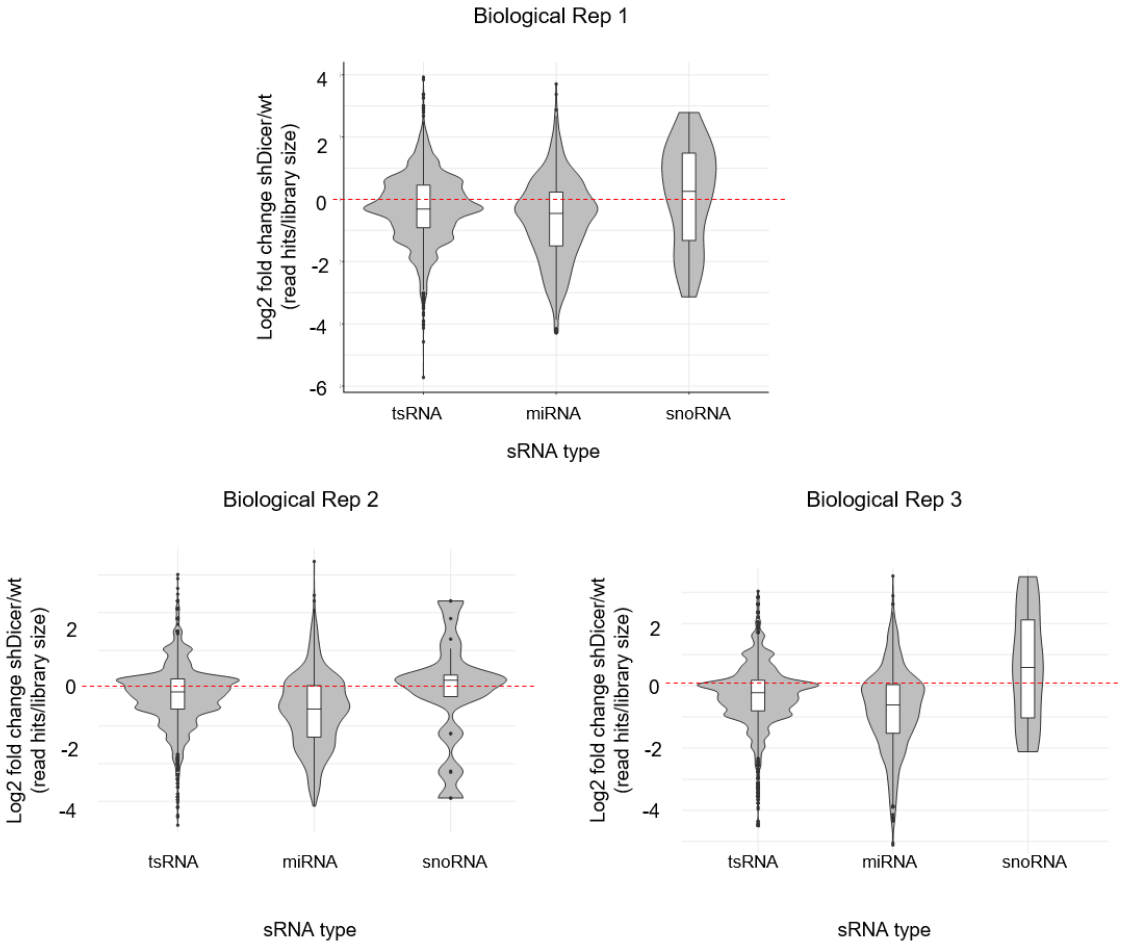


Figure 3-1 Global tsRNA levels are reduced upon Dicer knockdown Violin plots showing the log2 fold change of tRNA-derived small RNAs (tsRNAs), microRNAs (miRNAs) and small nucleolar RNAs (snoRNAs) upon Dicer knockdown (shDicer) (*n = 3*). Red dotted line is drawn across zero. t-tests results for tsRNA: *P* < 2.2e-16 with 95% confidence interval (CI) [-0.26, -0.21]; miRNA: *P* < 2.2e-16 with 95% CI [-0.52, -0.38]; snoRNA: *P* = 0.91 with 95% CI [-0.51,0.45]. Reads were normalised to library sizes. The analysis was done by Margarita Schlackow.

To support these observations, I performed northern blotting to compare the levels of selected tsRNAs identified in sRNA-seq in wild type HEK293T cells and an inducible Dicer knockdown cell line (Dicer-kd-2b2; shDicer)[134]. I managed to achieve a significant reduction of Dicer levels by inducing the expression of a stably integrated short hairpin RNA (shRNA) against *DICER* mRNA (shDicer) using doxycycline for 72 hours. Signals of tRF-5$^{\text{Ala-ACG}}$, tRF-5$^{\text{Glu-CTC}}$ and itRF$^{\text{Arg-CCT}}$ were sharply reduced upon induction of

Dicer knockdown (Figure 3-2). Altogether, these data suggest that Dicer is involved in the biogenesis or the stability of tsRNAs, including tRF-5s, tRF-3s and itRFs.

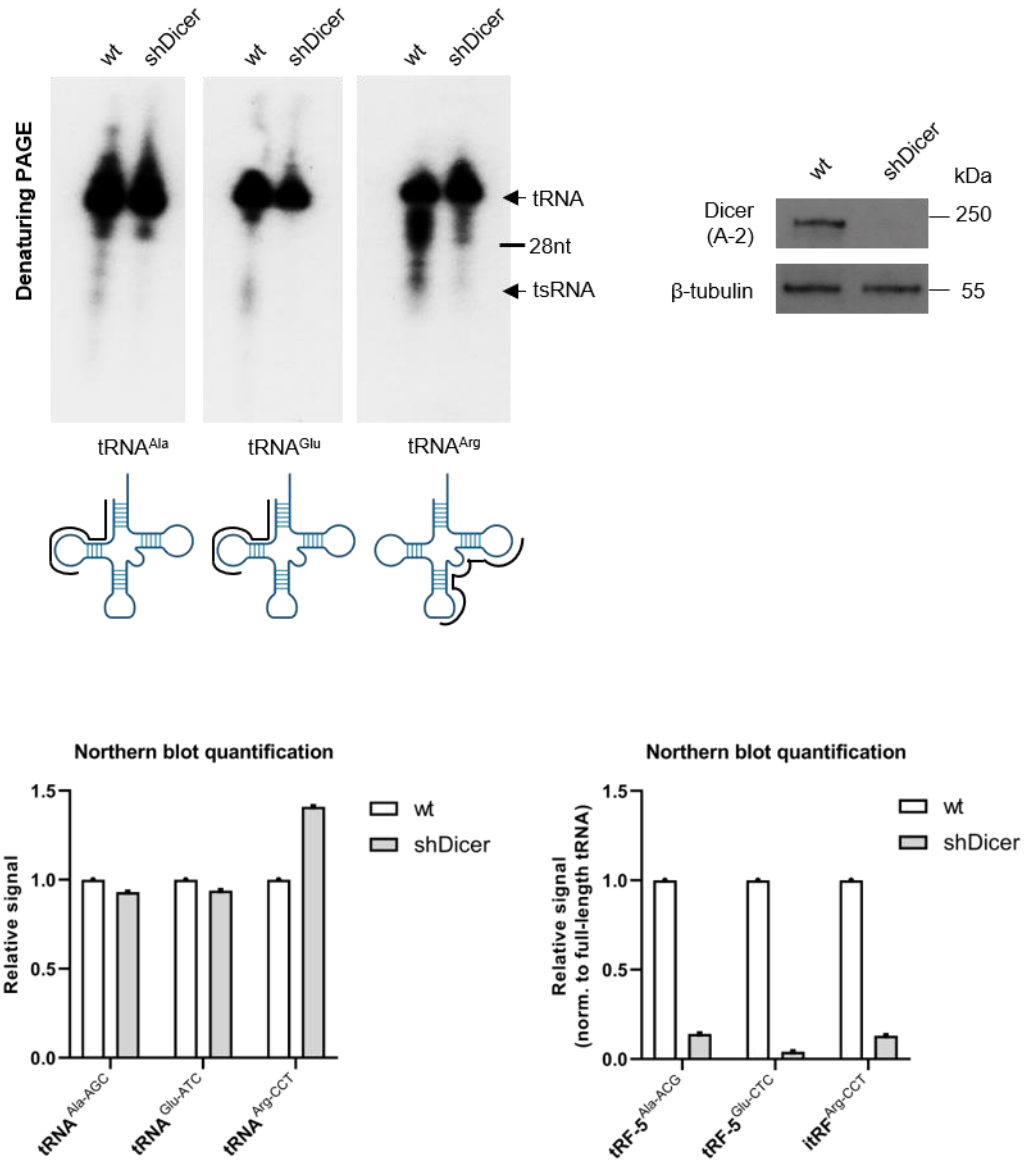


Figure 3-2 Levels of selected tsRNAs are reduced upon Dicer knockdown (Top) Images of northern blot showing signals of $tRNA^{Ala\text{-}AGC}$, $tRNA^{Glu\text{-}ATC}$, $tRNA^{Arg\text{-}CCT}$, and their corresponding tsRNAs ($tRF\text{-}5^{Ala\text{-}ACG}$, $tRF\text{-}5^{Glu\text{-}CTC}$ and $itRF^{Arg\text{-}CCT}$) in wild type HEK293T and an inducible Dicer knockdown (shDicer) cell line. The locations of DNA probes are indicated. Images of western blot showing successful depletion of Dicer in shDicer cells, with β-tubulin as a loading control. (Bottom) Bar charts showing the quantification of the signals (relative to wild type) of full-length tRNAs and their corresponding tsRNAs in the northern blots above.

3.3 Dicer cleaves tRNA *in vitro*

Previous PAR-CLIP analysis has indicated that Dicer binds directly to tsRNAs[141]. It is unknown whether Dicer can process tRNAs into sRNAs. To study this, an *in vitro* Dicer cleavage assay was performed by incubating immunoprecipitated tandem affinity purification (TAP)-tagged Dicer (Dicer-TAP) with T7-transcribed sRNAs followed by denaturing polyacrylamide gel electrophoresis (PAGE) and SYBR Gold staining. Incubation with Dicer-TAP led to the destabilisation of $tRNA^{Arg}$ and let7a precursor (pre-let7a), a well-known Dicer substrate, but did not affect the amount of snR38A, a snoRNA (Figure 3-3a). Subsequent northern analyses showed the emergence of sRNA species when $tRNA^{Arg}$ and pre-let7a, but not snR38A, were incubated with Dicer-TAP (Figure 3-3b). These data suggest that Dicer can process tRNAs into sRNAs *in vitro*.

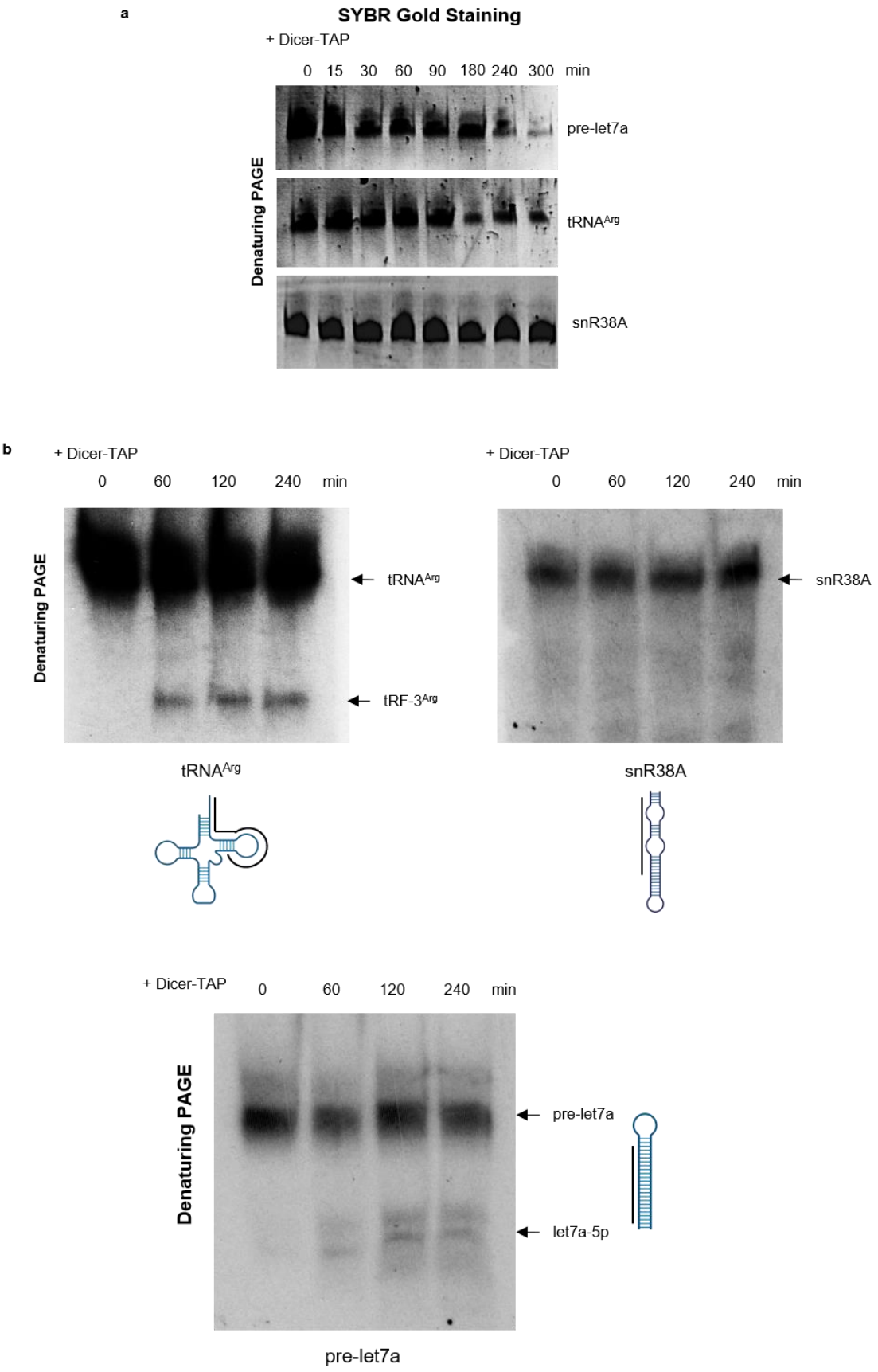

a
SYBR Gold Staining
+ Dicer-TAP
0 15 30 60 90 180 240 300 min
Denaturing PAGE
pre-let7a
tRNA^Arg
snR38A
b
+ Dicer-TAP
0 60 120 240 min
Denaturing PAGE
tRNA^Arg
tRF-3^Arg
+ Dicer-TAP
0 60 120 240 min
snR38A
tRNA^Arg
snR38A
+ Dicer-TAP
0 60 120 240 min
Denaturing PAGE
pre-let7a
let7a-5p
pre-let7a

Figure 3-3 Dicer processes tRNA into small RNA *in vitro* a) Images of polyacrylamide gel stained with SYBR Gold showing the signals of let7a precursor (pre-let7a), $tRNA^{Arg}$ and snR38A at different time points after the addition of tandem affinity purification (TAP)-tagged Dicer (Dicer-TAP). b) Images of northern blots showing signals of $tRNA^{Arg}$ and $tRF\text{-}3^{Arg}$, pre-let7a and let7a, and snR38A, respectively. The locations of probes are indicated. This experiment was performed by Adele Alagia.

3.4 Bioinformatic prediction of tsRNA target genes

Multiple reports have demonstrated that tsRNAs can act as regulators of gene expression, ranging from miRNA-like PTGS[107,120] to retrotransposon silencing[116,127]. To study the role of tsRNA in gene silencing, an unbiased bioinformatic approach was adopted. tsRNAs bound to Dicer, Ago1, Ago2, Ago3 and Ago4 were identified by employing Dicer[141] and Ago1-4[140] PAR-CLIP data. When unique tsRNA sequences were grouped into highly similar ones (e.g., one tsRNA sequence is the same as the other but one nucleotide longer), most tsRNA sequences were found in the Ago2 PAR-CLIP data (216 out of a total of 262 tsRNA sequences), indicating a strong association between Ago2 and tsRNAs.

To predict tsRNA target genes, mapped Dicer- and Ago-bound tsRNAs, which are 19 to 21 nt, were mapped to the genomic sequences of the genes that were upregulated in shDicer cells (from unpublished poly(A) RNA-seq data comparing gene expression in wild type HEK293T and shDicer cells) using miRanda, a bioinformatic tool for miRNA target prediction[47,144]. As a result, a list of genes that are predicted to be targeted by Dicer- and Ago-bound tsRNAs and their corresponding predicted target sites was obtained (Figure 3-4).

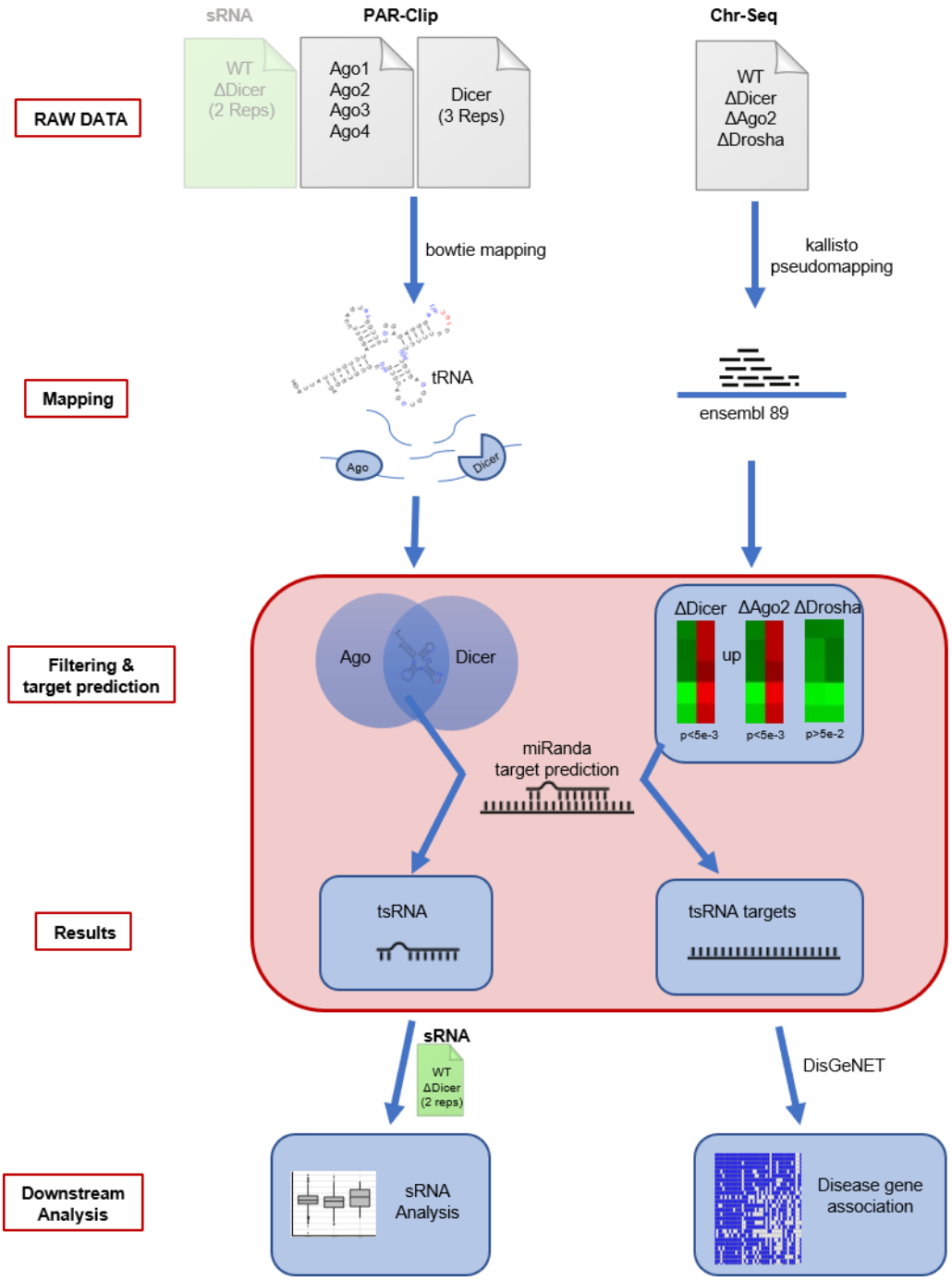


Figure 3-4 Bioinformatic prediction of tsRNA target genes Diagram summarising the bioinformatic workflow of tsRNA target gene prediction. Ago- and Dicer-associated tsRNA were identified using Dicer and Ago1, 2, 3 and 4 PAR-CLIP data. Predicted tsRNA target genes were identified by matching sequences of Ago- and Dicer-associated tsRNA with the genomic sequences of the upregulated genes in Dicer knockdown (shDicer) cells (obtained from unpublished RNA-seq dataset) by miRanda, a miRNA target screening algorithm. The analysis was done by Margarita Schlackow.

3.5 Silencing of tsRNA target genes may be miRNA-independent

Six genes, namely *NOV*, *GK*, *GUCY1A2*, *RP4-639F20.1*, *RBP7* and *SPINT1*, that were highly upregulated upon Dicer knockdown from the list of predicted target genes were randomly selected and assessed if the regulation of these genes is dependent upon miRNA. As Dicer plays a key role in miRNA biogenesis, miRNA loss is an expected consequence of Dicer knockdown and destabilisation of miRNA can result in upregulation of many genes. To study this, I transfected shRNA against *DROSHA* transiently to knock down Drosha in wild type HEK293T cells while employing cell lines with stably integrated shRNA expression cassette to downregulate Dicer and Ago2, respectively (Figure 3-5a).

qRT-PCR analysis showed that the selected tsRNAs were significantly upregulated in Dicer and Ago2 knockdown conditions but not when Drosha levels were partially reduced (Figure 3-5a). To assess whether partial Drosha knockdown has an effect on miRNAs, I measured the relative steady-state expression of *CDKN1B*, a miRNA target gene[147], and showed that the gene was significantly upregulated upon Drosha and Dicer knockdown (Figure 3-5b).

As Drosha is the most critical player in miRNA biogenesis[148], the observation indicates that the regulation of the selected tsRNA target genes may not be dependent on the miRNA pathway, but on a distinct pathway that involves both Dicer and Ago2. As *NOV* was upregulated upon Drosha knockdown, I suspect that *NOV* is regulated by both the miRNA pathway and a miRNA-independent mechanism.

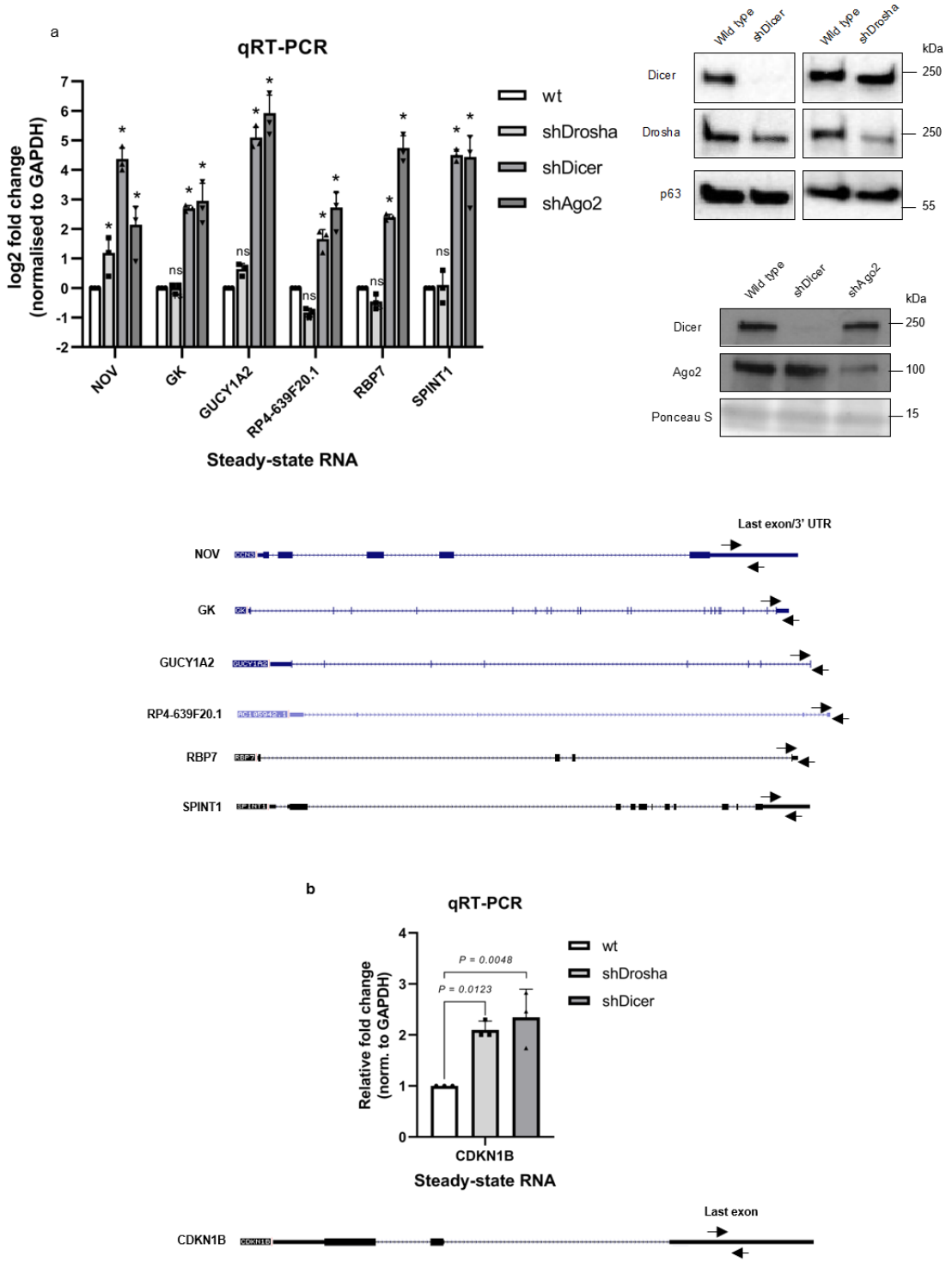


Figure 3-5 Regulation of selected tsRNA target genes may not be miRNA-dependent a) Bar chart showing the relative log2 fold change of the steady-state expression of six selected tsRNA target genes (*n = 3*), measured by qRT-PCR, upon Drosha, Dicer and Ago2 knockdown by shRNA strategies, respectively. (Right) Images of western blot showing successful knockdown of Dicer, Drosha and Ago2 with p63 and Ponceau S as loading controls. * indicates *P*-value below 0.05 (when compared with wild type). b) Bar chart showing the relative fold change of the steady-state expression of *CDKN1B*, a miRNA target gene,

(*n = 3*), measured by qRT-PCR, upon Drosha and Dicer knockdown by shRNA, respectively. The locations of qPCR primers are indicated.

3.6 Regulation of selected tsRNA target genes occurs in the nucleus

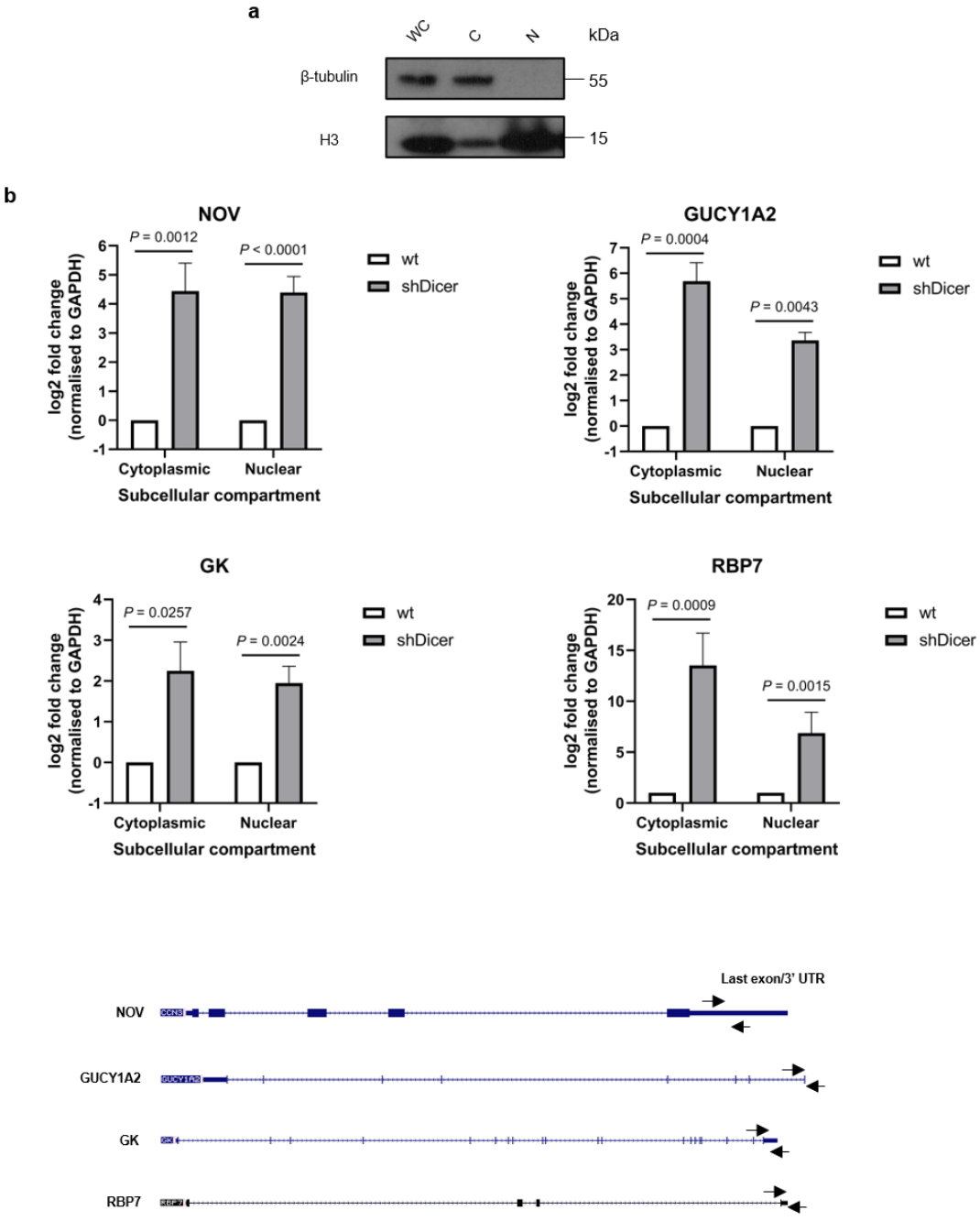
Given that the tsRNA target genes are not regulated via the miRNA pathway, the upregulation I observed might be a consequence of the disruption of TGS. To answer this question, I performed subcellular fractionation followed by qRT-PCR to test in which subcellular compartment the upregulation of the selected tsRNA target genes occurs.

To obtain a clear separation between cytoplasmic and nuclear fractions, I used hypotonic cell lysis buffer with low concentration of detergent (NP-40) to destabilise the cell membrane without breaking the nuclei followed by centrifugation. The nuclei pellet was later washed three times using the hypotonic lysis buffer without detergent before being lysed in nuclear lysis buffer. Subsequent western analysis showed that with this method, the cytoplasmic fraction was slightly contaminated by the nuclear fraction, as evidenced by detectable amount of histone H3 while the nuclear fraction is clear with undetectable levels of β-tubulin, a cytoplasmic protein. (Figure 3-6a).

I hypothesised that if the upregulation is found in both nucleus and cytoplasm, there is likelihood that the target genes are regulated at the transcriptional level. Indeed, qRT-PCR results show that the selected tsRNA target genes were significantly upregulated in both nucleus and cytoplasm (Figure 3-6b). This indicates that the selected tsRNA target genes are regulated at the nuclear level.

To eliminate the possibility of an artefact caused by cytoplasmic contamination of the nuclear fraction, I measured nascent RNA levels of the target genes by using primers that target intronic sequences in qRT-PCR. The selected tsRNA target genes, but not the control gene *ETNK1*, were significantly upregulated at the nascent RNA level upon Dicer

and Ago2 depletion (Figure 3-6c), further supporting the hypothesis that the tsRNA target genes are regulated in the nucleus.


a
WC
C
N
kDa
β-tubulin
55
H3
15
b
NOV
P = 0.0012
P < 0.0001
log2 fold change (normalised to GAPDH)
wt
shDicer
Cytoplasmic
Nuclear
Subcellular compartment
GUCY1A2
P = 0.0004
P = 0.0043
GK
P = 0.0257
P = 0.0024
RBP7
P = 0.0009
P = 0.0015
Last exon/3' UTR
NOV
GUCY1A2
GK
RBP7

c

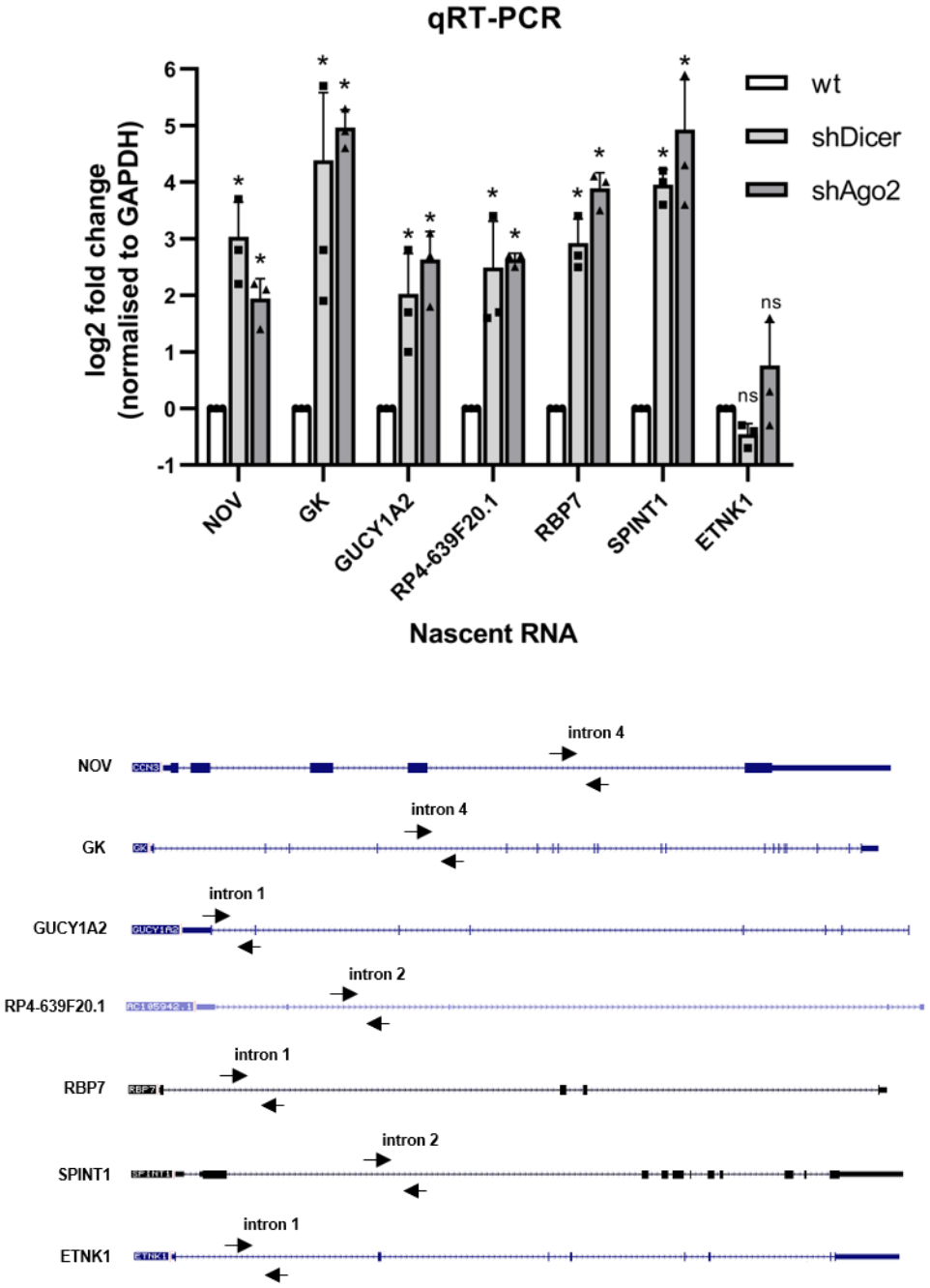


Figure 3-6 Selected tsRNA target genes are regulated in the nucleus a) Images of western blot showing protein signals of β-tubulin and H3 in whole-cell, cytoplasmic and nuclear fractions of wild type HEK293T cells. b) Bar charts showing relative log2 fold change of four selected tsRNA target genes (*n = 3*), measured by qRT-PCR, in cytoplasmic and nuclear fractions of wild type and Dicer knockdown (shDicer) cells. c) Bar chart showing the relative log2 fold change of nascent RNA levels of six selected tsRNA target genes and a control gene (*ETNK1*) (*n = 3*), measured by qRT-PCR upon Dicer (shDicer) and Ago2 (shAgo2) knockdown, respectively. * indicates *P*-value below 0.05 (when compared with wild type). The locations of qPCR primers are indicated.

3.7 Transcription of selected target genes is unaffected by Dicer knockdown

Based on the observation that the tsRNA target genes are regulated in the nucleus, I asked the question of whether they are regulated by TGS. To this end, I performed chromatin immunoprecipitation (ChIP) followed by qPCR to evaluate the interaction between total and active forms of RNA pol II, which are phosphorylated at serine position 2 or 5, and the promoter, exonic region and 3' UTR of the selected tsRNA target genes. ChIP analyses revealed that the levels of total and active RNA pol II across all the tested regions remained unchanged upon Dicer depletion, suggesting that Dicer knockdown has no effect on the transcription of the selected tsRNA target genes despite an upregulation at the nascent and steady-state RNA levels (Figures 3-5a, 3-6c and 3-7a). This observation implies that there is an additional layer of regulation which occurs between the regulation of RNA pol II and nuclear export of the transcript.

I also explored the possibility that heterochromatin is established at the selected tsRNA target gene loci. Using primers against the predicted target sites of the tsRNA target genes in ChIP-qPCR assays, I studied the levels of histone H3 lysine 9 dimethylation (H3K9me2), a potent repressive epigenetic mark[149], on the target sites of selected tsRNA target genes. The results revealed that levels of H3K9me2 at the predicted target sites were similar to that of the highly expressed control gene, *GAPDH*, and were unaffected by Dicer knockdown (Figure 3-7b).

The next question I asked was whether the upregulation of the selected tsRNA target genes upon Dicer knockdown is caused by a global dysregulation of transcription and heterochromatin. Western analysis showed that levels of total and active RNA pol II and H3K9me2 (together with H3) in wild type HEK293 cells were identical to those in

shDicer cells, suggesting that the upregulation I observed is not caused by a non-specific, global dysregulation of transcription and epigenetic repression by H3K9me2.

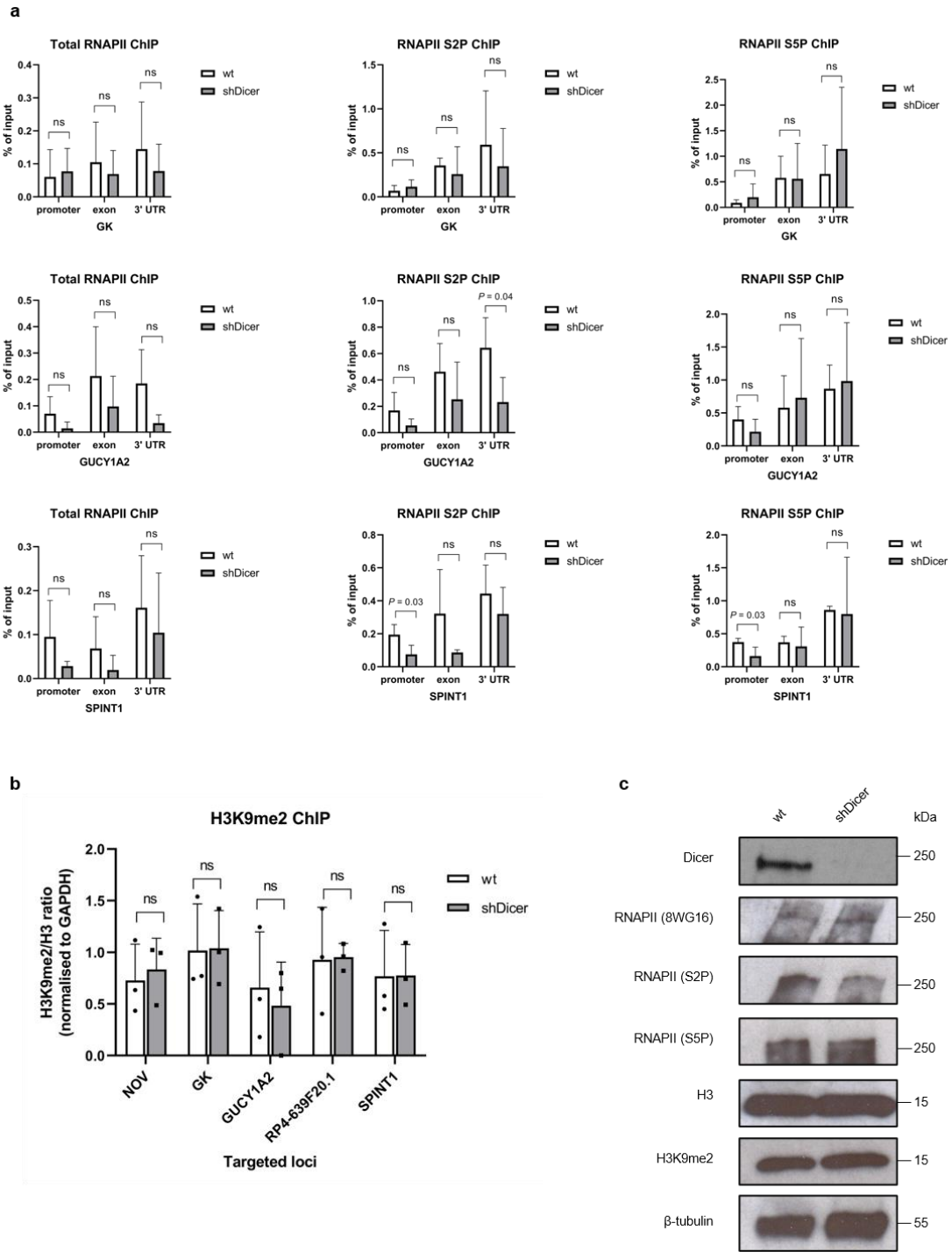


Figure 3-7 Transcription of tsRNA target genes is unaffected by Dicer knockdown a) Bar charts showing percentage (%) of input, measured by chromatin immunoprecipitation (ChIP) followed by qPCR, of total and active forms (S2P and S5P) of RNA polymerase II (RNAPII) at the promoter, exonic region and 3' UTR of three selected tsRNA target genes in wild type and Dicer knockdown (shDicer) cells. b) Bar

charts showing the ratio of histone H3 lysine 9 dimethylation (H3K9me2) to H3, measured by ChIP-qPCR, at the corresponding predicted target site of five selected tsRNA target genes in wild type and shDicer cells. c) Images of western blot representing protein signals of Dicer, total RNAPII, active forms of RNAPII, H3, H3K9me2 and β-tubulin (loading control) in wild type and shDicer cells.

3.8 Hypothetical model of nascent RNA silencing

Taken together, the data compelled me to hypothesise that tsRNA can mediate gene silencing in a manner that is distinct from that of PTGS and TGS. This hypothesis is developed based on three main findings:

1. The regulation of tsRNA target genes may be independent of the miRNA pathway.
2. Selected target genes were upregulated at the nascent level when Dicer levels are reduced.
3. Transcription of tsRNA target genes remained unchanged upon Dicer knockdown despite the upregulation of tsRNA target genes.

As the pathway is dependent upon Dicer and Ago2, but not Drosha, I speculate that tsRNAs are generated from Dicer processing of alternatively folded tRNAs and subsequently loaded onto Ago2 to form a silencing complex, similar to RISC. The Ago2-containing silencing complex is guided by the tsRNA to target RNA co-transcriptionally, resulting in the destabilisation of nascent RNA, in a sequence-specific manner. The term “nascent RNA silencing”, or NRS, was coined for this novel gene silencing mechanism.

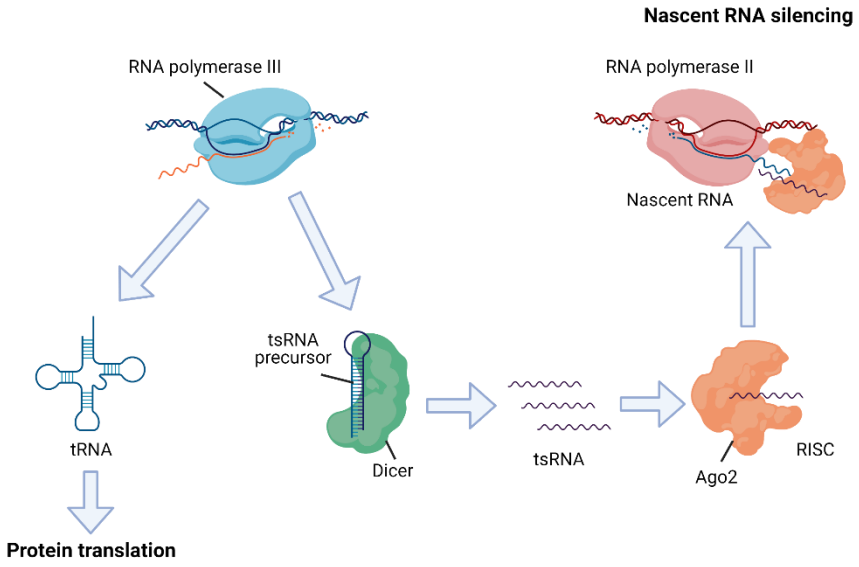


Figure 3-8 Hypothetical model of nascent RNA silencing Diagram illustrating the proposed model for a novel gene silencing mechanism mediated by tsRNA. Transfer RNAs (tRNAs) are transcribed by RNA polymerase (pol) III. Some of the tRNAs that are folded in a structure recognisable by Dicer are processed into tRNA-derived small RNAs (tsRNAs), which are subsequently loaded onto Argonaute 2 (Ago2) proteins to form a silencing complex. Guided by tsRNA, the Ago2-containing silencing complex finds its target RNA through sequence recognition and destabilises the RNA while it was being transcribed by RNA polymerase II (RNA pol II). This hypothetical mechanism is termed nascent RNA silencing (NRS).

3.9 Discussion

In the first part of this chapter, I asked the question of whether Dicer is responsible for tsRNA biogenesis. Previously, Dicer has been shown to be bound to structural RNAs, such as tRNAs, snoRNAs and vault RNAs, by PAR-CLIP[141]; the depletion of Dicer also led to a slight increase in tRNA levels, measured by qPCR using primers that amplify the full-length RNA[141]. In other words, the absence of Dicer can stabilise a small group of tRNAs. Another report also suggested that tRNAs might fold into shRNA-like structures that could be processed by Dicer[102]. Hasler *et al.* showed that Lupus autoantigen La can help stabilise the canonical tRNA clover-leaf structure and prevent tRNA from folding into shRNA and being processed by Dicer. However, the authors also found tRNA precursors that generate tsRNA in the absence of La[150]. Based on these findings, I

speculate that a proportion of tRNAs are folded into shRNA and serve as a source of tsRNA, which act as regulators of gene expression. This hypothesis is currently being investigated by my colleagues in the Gullerova Lab.

I observed that Dicer depletion led to the loss of tsRNAs, including tRF-3s, tRF-5s and itRFs, using northern blotting and next generation sequencing methods. However, it is possible that the decrease of tsRNAs is caused by an indirect effect of Dicer knockdown. To circumvent this, TAP-tagged Dicer was shown to cleave $tRNA^{Arg}$ *in vitro*, generating tsRNA. This demonstrates a direct effect of Dicer has on tRNAs. Importantly, our data echoes the report by Cole *et al.* in which the authors observed marked decrease of a tsRNA derived from $tRNA^{Gln}$ using a different Dicer knockdown strategy in another cell line – siRNA knockdown in HeLa cells – and by incubating recombinant Dicer, as opposed to immunoprecipitating endogenous Dicer, with the same tRNA and detected generation of tsRNA[106].

The biogenesis of itRF still remains unknown. In our sRNA-seq and northern blot analyses, levels of many itRFs were significantly reduced upon Dicer knockdown, indicating that Dicer might play a role in generating itRF. Further structural analyses are required to explore whether it is physically possible for Dicer to generate itRFs.

In the second part, a bioinformatic method was employed to explore the role of tsRNA in gene regulation in an unbiased manner. Re-analysis of Dicer and Ago PAR-CLIP data confirmed the association between tsRNA and the two proteins. Our data also demonstrated that tsRNAs were highly associated with Ago2, a finding which compelled us to focus on Ago2 for further investigation of the proposed gene silencing mechanism.

miRanda takes into account of three criteria when predicting an sRNA::target pair: matching of seed sequence, binding free energy and conservation of both binding site and

position. Importantly, miRanda considers the sRNA::target match along the entire sRNA sequence[47]. By using miRanda as a predictor, we assumed that 19-22 nt tsRNA matches its target sequences in a miRNA-like manner. Instead of feeding the algorithm with only 3' UTR of upregulated genes, full genomic sequences were considered as the initial idea was to test for TGS in *trans*. This approach was also motivated by a report on RNA degradation induced by siRNA originated from an intronic sequence in plants[151] and another study in which intronic sequences were detected in nuclear Ago binding[70].

The test for the miRNA-dependency of the predicted target genes were based on the fact that Drosha is the most crucial player in miRNA biogenesis. The protein is responsible for the generation of a vast majority (96.5%) of miRNA in HCT116 cell line[148]. If the expression of selected target genes were not responsive to Drosha knockdown, there is a high possibility that they are not regulated by miRNAs. However, as I was only able to achieve a partial knockdown of Drosha, further investigations are required to ascertain the miRNA dependency of tsRNA target genes. For instance, miRNA levels can be quantified by northern blotting or sRNA-seq; RNA-seq can also be performed to measure differential expression of miRNA target genes to confirm that the partial knockdown of Drosha can lead to upregulation of miRNA targets. Intriguingly, the regulation of selected target genes was dependent upon the other two major players in the canonical miRNA pathway, Dicer and Ago2. I therefore hypothesise that tRNA precursors can directly interact with Dicer, bypassing the first step of the two-step processing for miRNA.

Further investigations into the regulatory mechanism of the tsRNA target genes revealed that the regulation occurs in the nucleus. Subcellular fractionation shed light on the distribution of the target transcripts at steady state. It is important to note that when quantifying the relative change of expression, the take-off values of the transcripts were normalised to that of a housekeeping gene based on the assumption that at steady state

the target gene transcripts are distributed across the subcellular compartments in a similar manner to the housekeeping gene. Intron-based qRT-PCR resolves the issue of imperfect fractionation and the eliminates the possibility that the upregulation I observed was a deviation from how transcripts are normally distributed in the cell.

Previously, intron-based qRT-PCR had been employed to measure transcription of genes[152]. The observation that selected tsRNA target genes were upregulated upon Dicer and Ago2 depletion in intron-based qRT-PCR had therefore prompted me to hypothesis that tsRNA might be mediating TGS in *trans*. However, careful examination of the transcription of the target genes by ChIP analyses led to a confounding result: transcription of the genes was not increased, despite an upregulation at the steady-state level.

Altogether, these data formed the basis for the hypothetical model for a novel nuclear gene silencing mechanism.

4 Molecular mechanism of tsRNA-mediated gene silencing

4.1 Introduction

In Chapter 3, I proposed that Dicer-dependent tsRNAs can promote gene silencing of target genes at the nascent level through a novel pathway I call NRS (Section 3.8). Here, I intend to elaborate on the molecular mechanisms underlying NRS by using a broad range of experimental strategies.

4.2 Genome-wide analysis of chromatin-associated RNA reveals upregulation of many genes in Dicer and Ago2 knockdown

To examine NRS genome-wide, I performed chromatin-associated RNA sequencing (chrRNA-seq) comparing the levels of chrRNA, as a proxy for nascent RNA, in wild type, shDrosha, shDicer and shAgo2 conditions. The knockdown of Drosha, Dicer and Ago2 were verified by qRT-PCR prior to sequencing (Figure 4-1a). Differential gene expression analysis revealed an overlap of 2143 upregulated genes between Dicer and Ago2 knockdown samples, whilst only 256 genes were upregulated in Drosha knockdown cells (Figure 4-1b).

This huge overlap of upregulated genes between Dicer and Ago2 suggests a Dicer- and Ago2-dependent gene silencing mechanism which regulates genes at the nascent level. This mechanism, however, is independent of Drosha, which correlates with my previous results in Chapter 3.

The RNA-seq data in our bioinformatic tsRNA target gene prediction pipeline (refer to Section 3.4) was substituted with the chrRNA-seq data to eliminate genes that were not upregulated at the nascent level, yielding a revised list of 1789 predicted tsRNA target

genes, including both protein-coding genes and long non-coding RNAs. By employing miRTarBase[147,153], 72% of these predicted tsRNA target genes were found to have no or weak experimental evidence as miRNA targets (Figure 4-1c; Appendix 2).This suggests that NRS regulates a different group of genes than those regulated primarily by the miRNA pathway.

Careful examination of the location of the predicted target sites within the genes showed that a large proportion of the genes that were upregulated in both Dicer and Ago2 knockdown cells were predicted to be targeted in either the intronic or exonic regions; very few were targeted exclusively in their exons (Figure 4-1d). This observation supports the notion that NRS occurs in the nucleus.

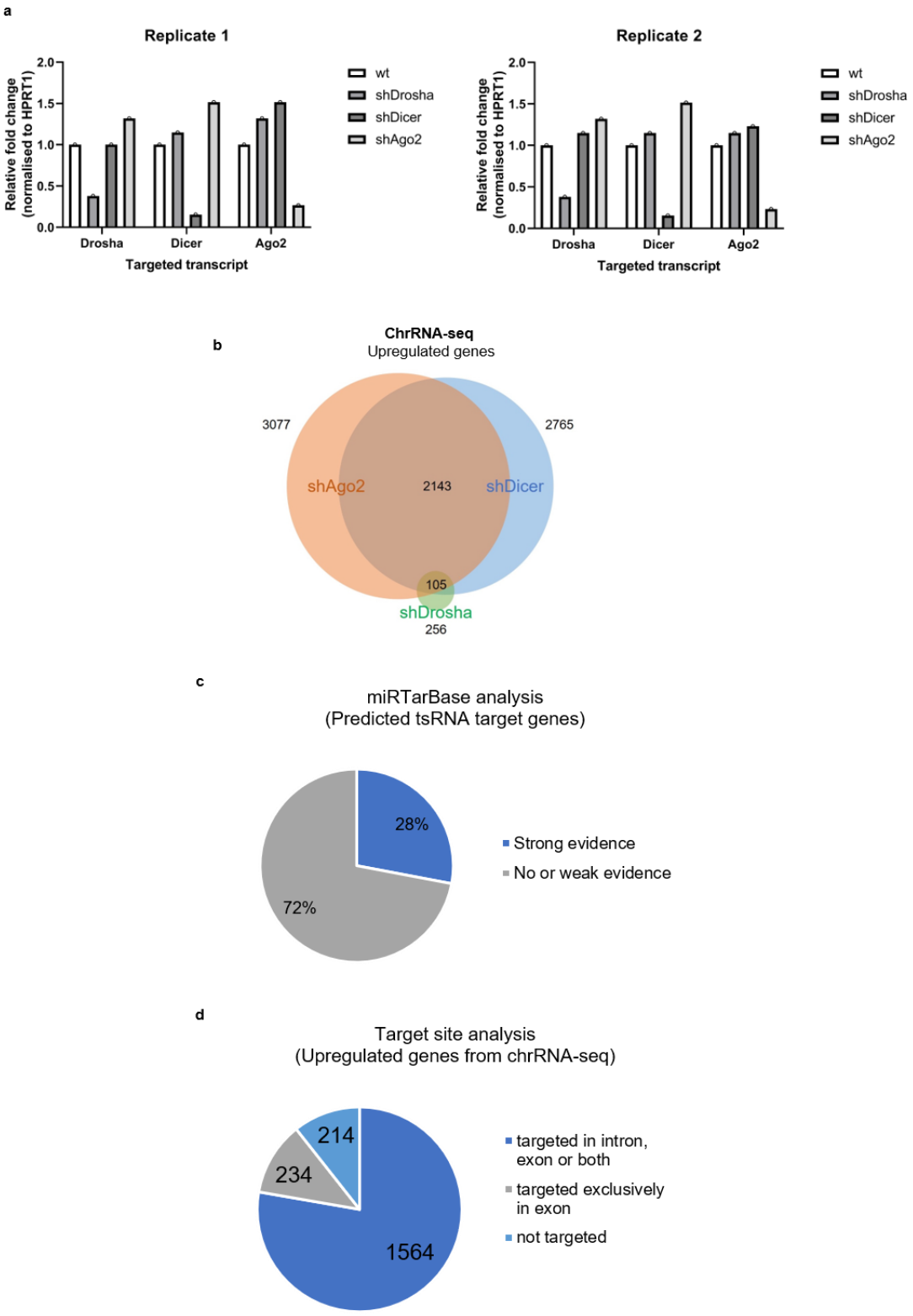
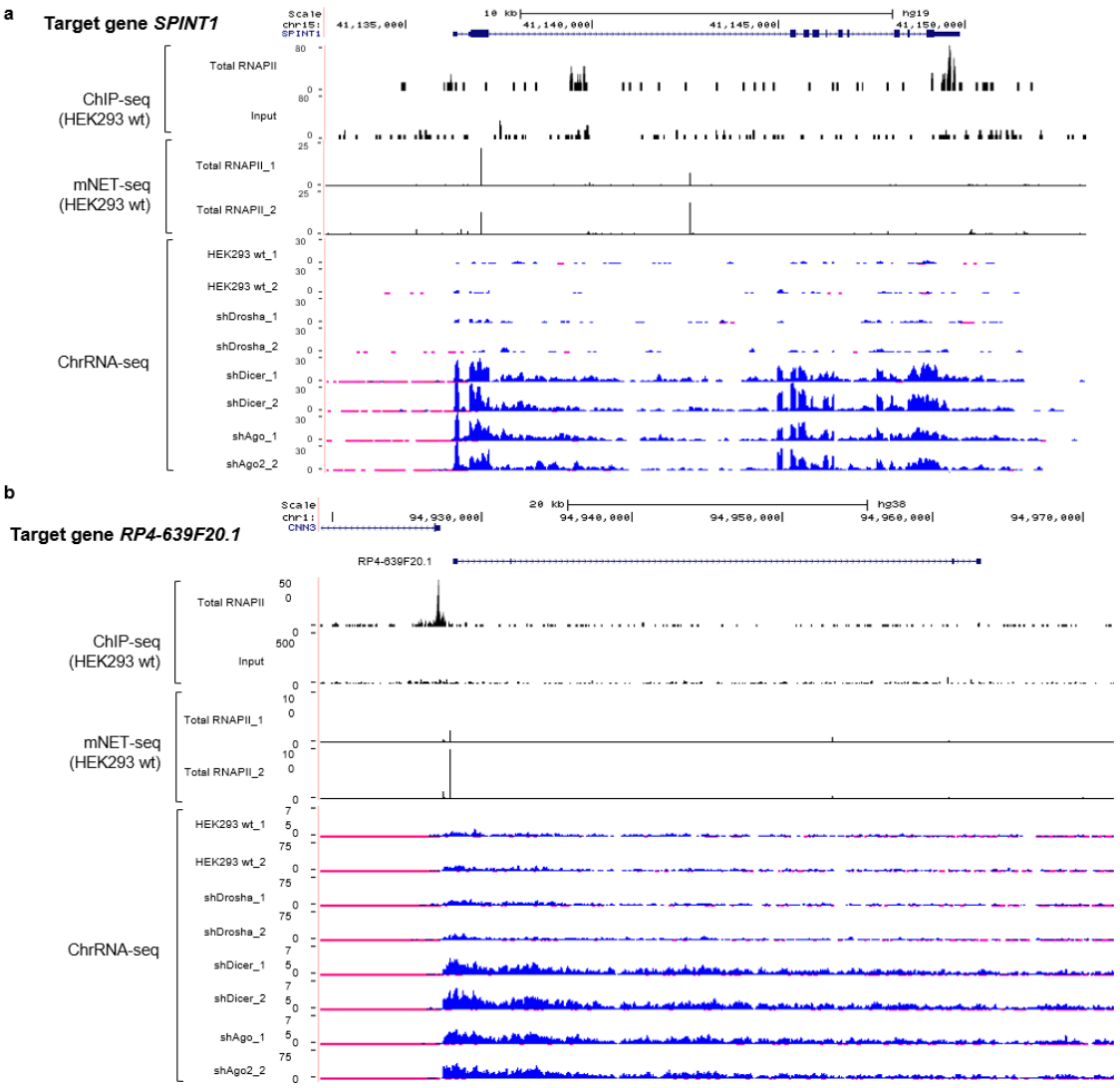


Figure 4-1 Summary of chrRNA-seq results a) Bar charts showing relative fold change of expression of steady-state *DROSHA*, *DICER* and *AGO2* transcripts, measured by qRT-PCR, in Drosha (shDrosha), Dicer Dicer (shDicer) and Ago2 knockdown (shAgo2) cells, respectively. b) Venn diagram representing the overlap of the upregulated genes between shDicer (*P*-value < 0.005), shDrosha (*P*-value < 0.05) and

shAgo2 (*P*-value < 0.005) cells. Differential gene expression analysis was performed followed by whole-genome sequencing of chromatin associated RNA (chrRNA-seq). c) Pie chart representing the breakdown of tsRNA target genes by evidence for being miRNA targets in miRTarBase. d) Pie chart representing the distribution of the position of target sites among tsRNA target genes.

Predicted protein-coding target gene *SPINT1* and non-coding target gene *RP4-639F20.1* were significantly elevated in Dicer and Ago2 knockdown cells but showed very low expression in wild type and Drosha knockdown conditions. On the contrary, the expression of non-target gene *GAPDH* was constant across all four conditions. The expression of a miRNA target gene *PTEN*[154] was unaffected by Drosha, Dicer and Ago2 knockdown. To ensure that the low expression we detected for these two selected target genes was not due to the absence of transcription, RNA pol II ChIP-seq data and mNET-seq data[155] were aligned with the chrRNA-seq data to show that target genes *SPINT1* and *RP4-639F20.1* were actively transcribed, as evidenced by signals of RNA pol II on the genes (Figure 4-2).


a
Target gene SPINT1
Scale
chr15:
SPINT1
41,135,000
10 kb
41,140,000
41,145,000
hg19
41,150,000
ChIP-seq
(HEK293 wt)
Total RNAPII
Input
mNET-seq
(HEK293 wt)
Total RNAPII_1
Total RNAPII_2
ChrRNA-seq
HEK293 wt_1
HEK293 wt_2
shDrosha_1
shDrosha_2
shDicer_1
shDicer_2
shAgo_1
shAgo2_2
b
Target gene RP4-639F20.1
Scale
chr1:
CNN3
94,930,000
20 kb
94,940,000
94,950,000
94,960,000
hg38
94,970,000
RP4-639F20.1
ChIP-seq
(HEK293 wt)
Total RNAPII
Input
mNET-seq
(HEK293 wt)
Total RNAPII_1
Total RNAPII_2
ChrRNA-seq
HEK293 wt_1
HEK293 wt_2
shDrosha_1
shDrosha_2
shDicer_1
shDicer_2
shAgo_1
shAgo2_2

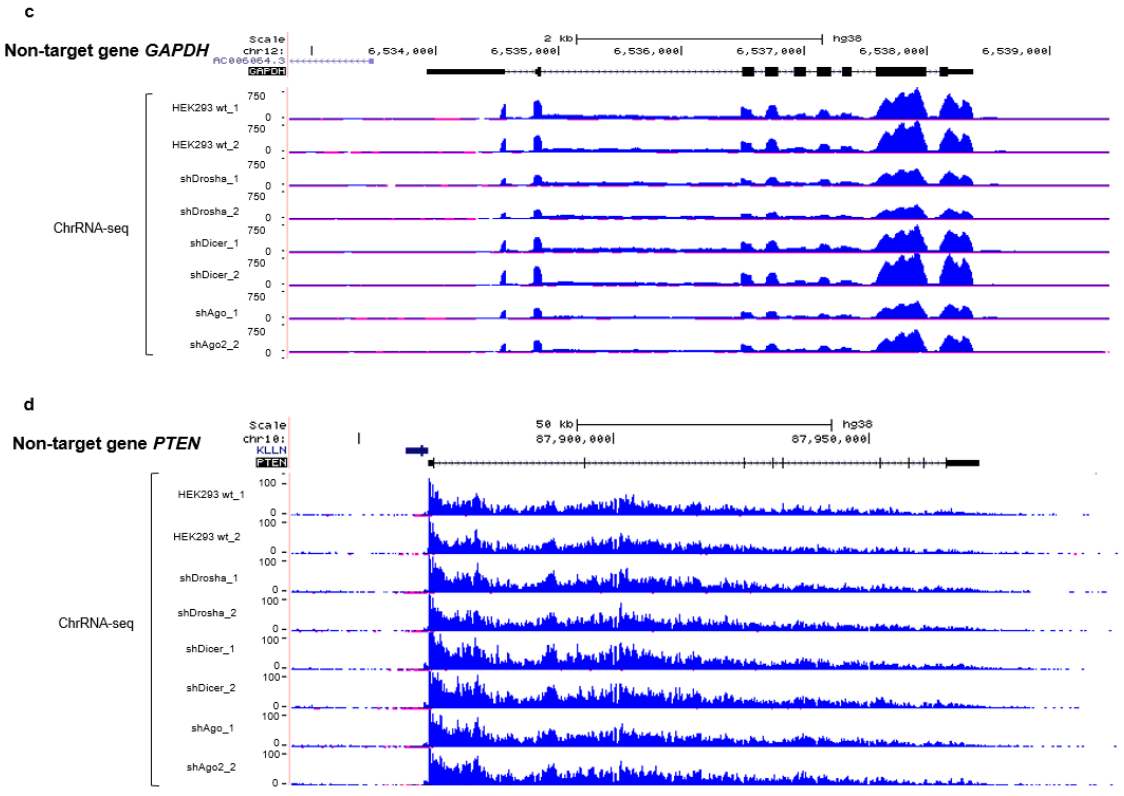


Figure 4-2 chrRNA-seq profiles of target versus non-target genes Combined snapshots of RNA pol II ChIP-seq, mNET-seq and chrRNA-seq profiles across tsRNA a) protein-coding target gene *SPINT1* b) non-coding target gene *RP4-639F20.1* c) non-target gene *GAPDH* (*n = 2*) (d) non-target gene *PTEN* (n = 2). Normalised read counts are indicated.

4.3 tsRNA are present in both the nucleus and cytoplasm

To further explore the notion that NRS takes place in the nucleus, I performed subcellular fractionation followed by northern blotting to assess whether tsRNA are present in the nucleus. The results showed that both tRF-5$^{Glu\text{-}CTC}$ and itRF$^{Arg\text{-}CCT}$, which were previously found to be significantly downregulated upon Dicer knockdown in sRNA-seq (Section 3.2), were present in both nucleus and cytoplasm at steady state (Figure 4-3a). However, this result could not distinguish between the subcellular compartment in which tsRNAs are generated and carry out their functions.

To find out where tsRNAs execute their functions, I chemically labelled *in vitro* transcribed siRNA against *EGFR* (siEGFR) and tsRNAs, which were predicted to target

EGFR (tsEGFR, also tRF-5$^{Glu\text{-}CTC}$), with fluorescent dye Alexa Fluor® 488 and transfected them into wild type HEK293T cells. The cells were imaged by fluorescence microscopy 24 hours post-transfection. Confocal images showed that the labelled siEGFRs were detected only within the cytoplasm whilst tsEGFRs were found in both the nucleus and cytoplasm. It was expected that siRNAs are localised in the cytoplasm because they primarily carry out gene silencing through PTGS in the cytoplasm. Altogether the data suggest that, unlike siRNA tsRNA can localise to both the nucleus and cytoplasm and has a potential nuclear role.

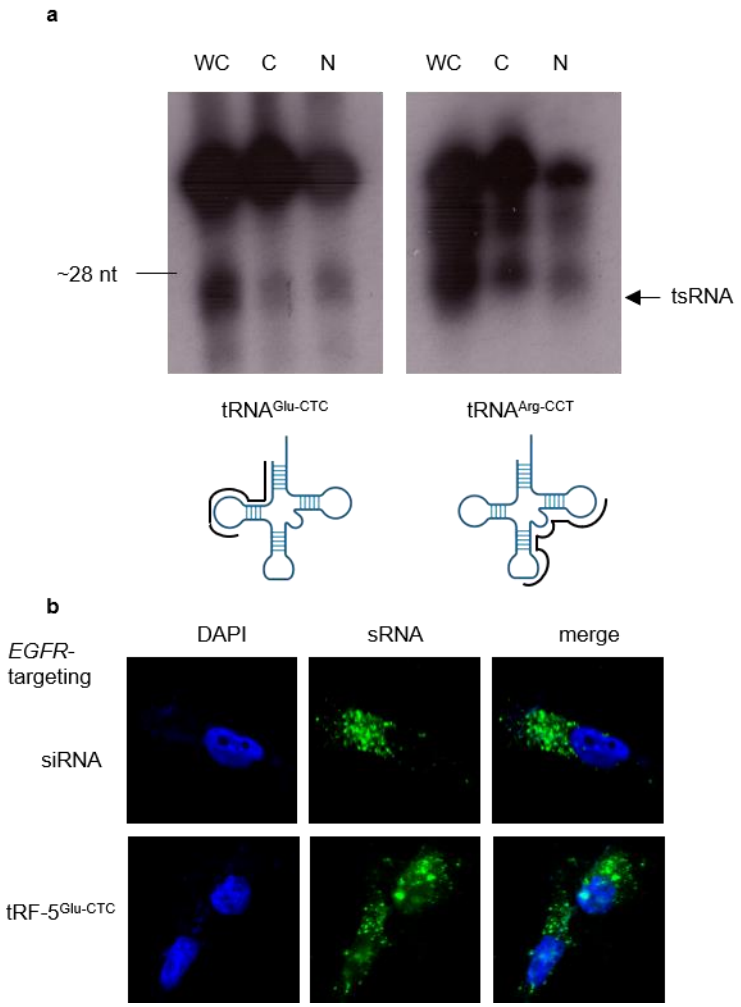


Figure 4-3 Cellular localisation of tsRNA a) Image of northern blot showing signals of tRNA$^{Glu\text{-}CTC}$, tRNA$^{Arg\text{-}CCT}$ and their corresponding tsRNAs, tRF-5$^{Glu\text{-}CTC}$ and itRF$^{Arg\text{-}CCT}$ from whole-cell (WC), cytoplasmic (C) and nuclear (N) RNA fractions from wild type HEK293T cells. b) Representative confocal

images showing the localisation of fluorescently labelled tRF-5$^{Glu\text{-}CTC}$ and siEGFR (in green) after transfection into BT549 cells. DAPI signal is indicated in blue.

4.4 Transfection of tsRNA mimics leads to silencing of selected target gene *SPINT1*

To explore the nuclear role of tsRNA in the context of NRS, I transfected unmodified, single-stranded mimics of the tsRNA predicted to target intron 2 of *SPINT1* (tsSPINT1, also tRF-3$^{Gln\text{-}CTG}$) into wild type HEK293T cells, isolated total RNA from the cells 48 hours post-transfection and quantified the relative change of expression of *SPINT1* by qRT-PCR.

In this experiment, I used two primers which amplified a region in intron 2 of *SPINT1* containing the predicted target site and an intronic site which is adjacent (around 250bp downstream) to the target site, respectively. I observed a significant reduction in expression of both targeted sites upon transfection of synthetic tsSPINT1 in comparison to mock transfection, while the expression of non-target gene *HPRT1* was unaffected.

In the condition where *SPINT1* was upregulated due to Dicer depletion, transfection of synthetic tsSPINT1 resulted in an attenuation of the upregulation. The transfection of another synthetic tsRNA which was predicted to target *GK* (tsGK, also itRF$^{Gln\text{-}CTG}$) did not affect the upregulation of *SPINT1*. This indicates that the effect intron-targeting tsSPINT1 had on *SPINT1* was specific.

In summary, transfection of tsRNA mimics, that is increasing cellular levels of tsRNAs, can lead to silencing of its target genes in a sequence-specific manner.

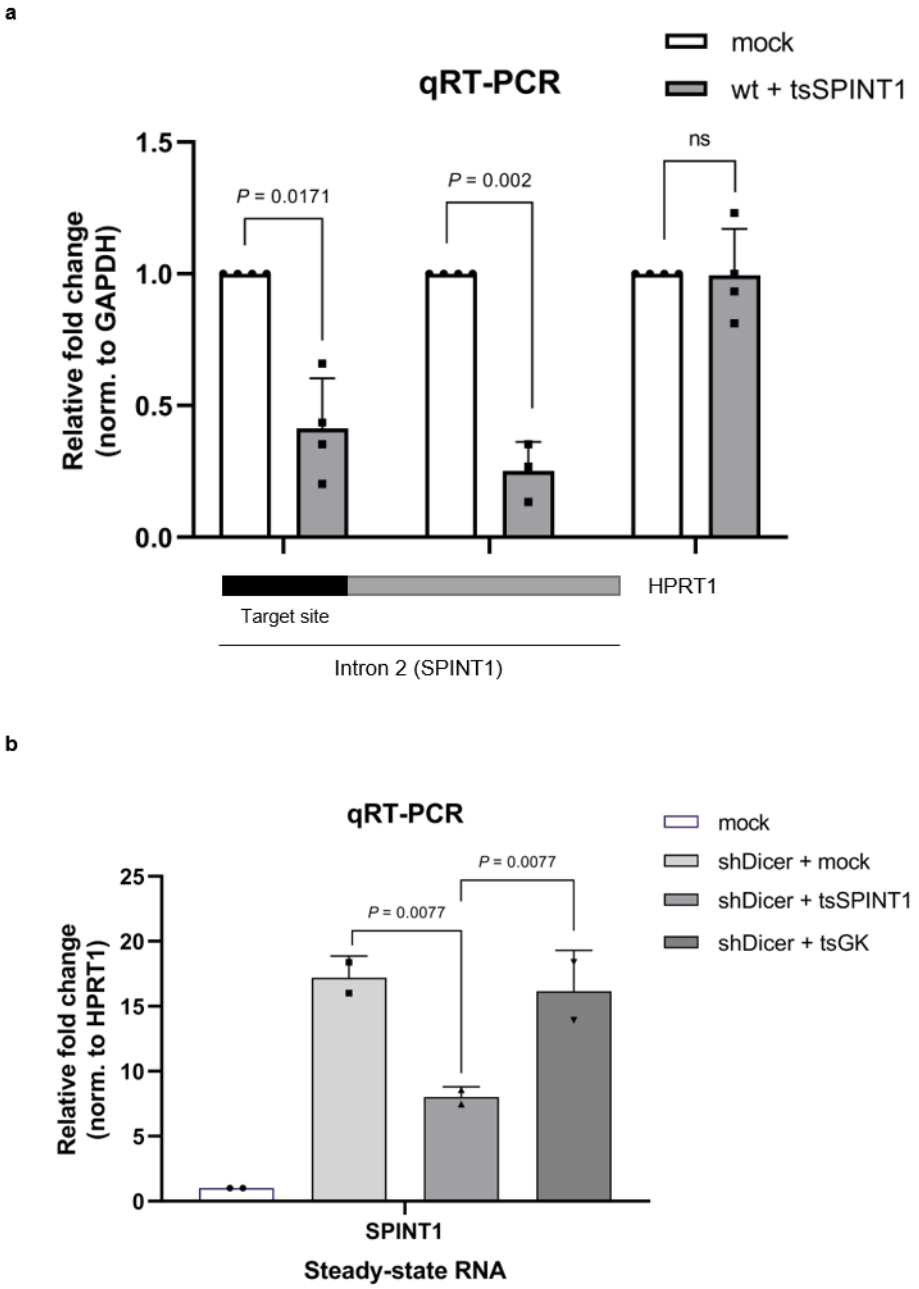


Figure 4-4 Transfection of tsRNA mimics silences target gene *SPINT1* a) Bar charts showing the relative fold change of expression of the predicted tsSPINT1 (also tRF-3$^{Gln\text{-}CTG}$) target site and the intronic site adjacent to it within intron 2 of *SPINT1* (*n = 3*), measured by qRT-PCR, with *HPRT1* as negative control. b) Bar charts showing the relative fold change of steady-state expression of *SPINT1* (*n = 2*), measured by qRT-PCR, in wild type HEK293T, shDicer and shDicer transfected with tsSPINT1 and tsRNA predicted to target *GK* (tsGK, also itRF$^{Gln\text{-}CTG}$), respectively.

4.5 CRISPR-Cas9 knockout of a predicted tsRNA target site results in elevated levels of the target gene

Previous experimental strategies I used to study tsRNA-mediated NRS had been focused on deregulating tsRNA levels. By knocking down Dicer, I studied how the loss of tsRNA can affect the expression of tsRNA target genes. I later tested whether an increase in tsRNA levels by transfection of synthetic tsRNAs can result in the silencing of the target genes. In this section, I will outline another strategy I used to validate tsRNA-mediated NRS, which is by removing the predicted target site of tsSPINT1 (Figure 4-5a). I hypothesise that the removal of *SPINT1* target site will lead to an increase in its expression.

I generated a HEK293T-based cell line where a genomic region containing the predicted intronic target site within the *SPINT1* gene is deleted using CRISPR-Cas9 gene editing technology. The 18-bp target site is located within intron 2 of *SPINT1*. The CRIPSR-Cas9 strategy is outlined in Figure 4-5b.

Screening of mutants involved the amplification of a region of 579 bp in intron 2 of *SPINT1* which encompasses the 135-bp region targeted for deletion. A PCR product of 579 bp was expected for cells with a wild type copy of the *SPINT1* gene while cells harbouring the mutant form of the gene would yield a shorter (444-bp) PCR product. I selected a heterozygous ($SPINT1^{+/-}$) mutant and a homozygous ($SPINT1^{-/-}$) mutant for further analysis (Figure 4-5c). The corresponding PCR products were Sanger sequenced to confirm the deletion.

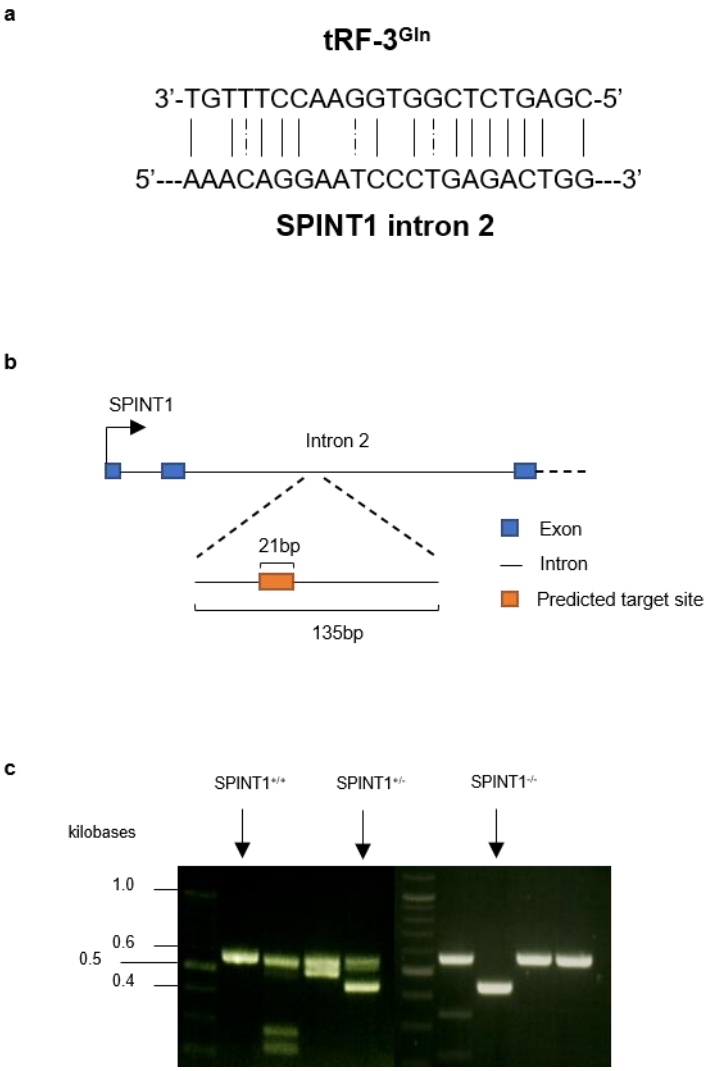


Figure 4-5 CRISPR-Cas9 knockout of *SPINT1* target site a) Diagram representing the matching between tRF-3^{Gln} and the predicted target site in intron 2 of *SPINT1*. b) Diagram illustrating the CRISPR-Cas9 strategy for generating a mutant HEK293T-based cell line in which the tsSPINT1 target site within intron 2 of *SPINT1* and its flanking sequences are deleted. c) PCR screening for mutants followed by CRISPR-Cas9 gene editing and single cell sorting. Image of DNA gel showing signals of PCR products from single-cell clones. 579-bp band corresponds to wild type allele, while 444-bp band corresponds to mutant allele.

qRT-PCR analysis showed that an increase in *SPINT1* expression at both nascent and steady-state levels in SPINT1$^{-/-}$ cells, but not in SPINT1$^{+/-}$ cells (Figure 4-6a, b). Such elevation is also reflected at the protein level by western analysis (Figure 4-6c). To understand why *SPINT1* expression was unaffected in SPINT1$^{+/-}$ cells, I performed RT-PCR to examine the expression of *SPINT1* in wild type, SPINT1$^{+/-}$ and SPINT1$^{-/-}$ cells,

respectively. The results showed that in SPINT1$^{+/-}$ cells the wild type copy of *SPINT1* transcript was predominantly expressed while in SPINT1$^{-/-}$ cells where *SPINT1* is significantly upregulated, only the mutant form of *SPINT1* is expressed. (Figure 4-6d). The RT-PCR products were subsequently sequenced and genotyped.

Deletion of a genomic region containing the predicted target site of *SPINT1* leading to increase of the *SPINT1* strongly suggests that the target site plays a role in regulating the expression of the gene. The elevated expression occurs at both nascent and steady-state levels, through to the protein level, indicating that the regulation happens at nascent level.

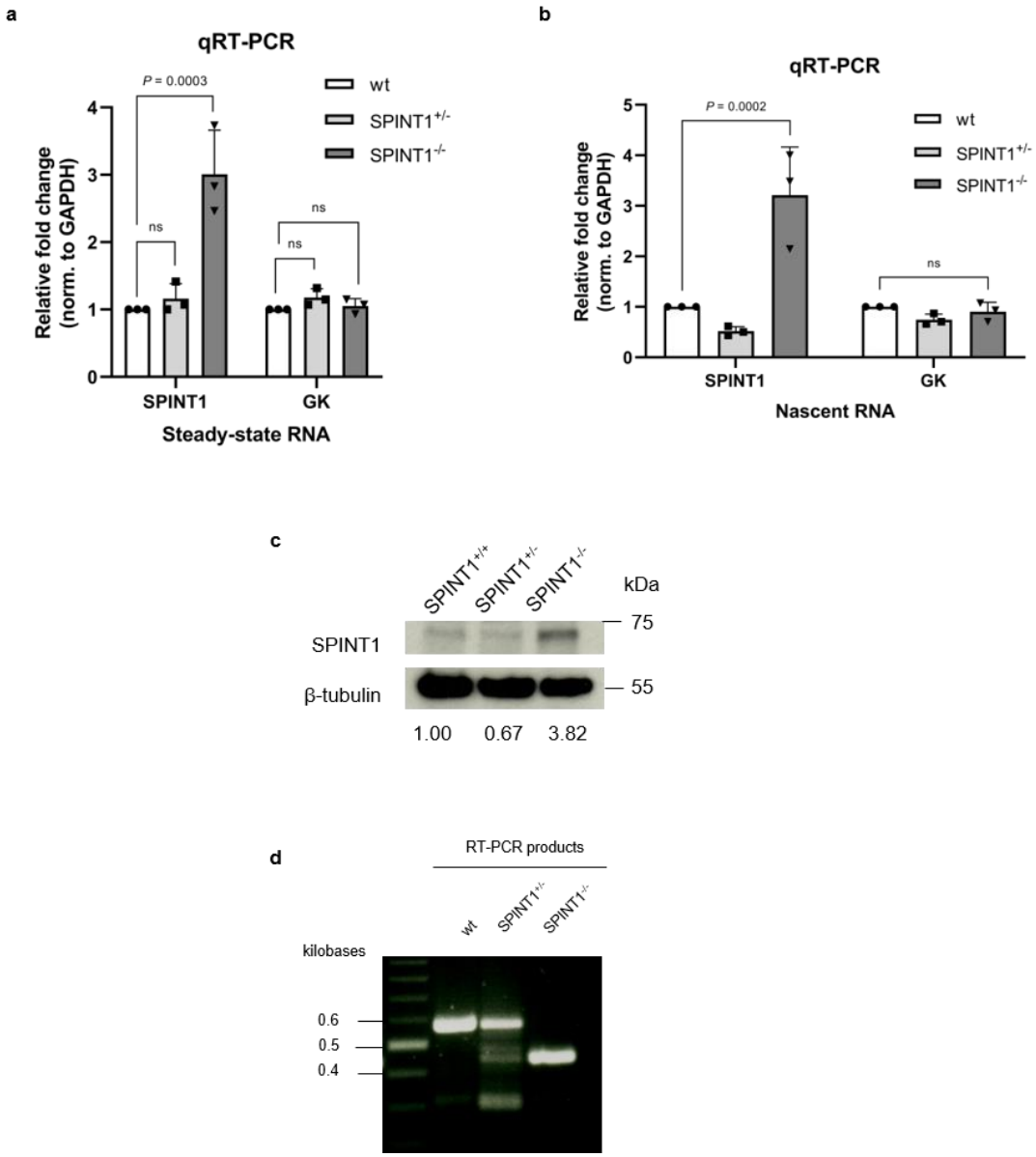


Figure 4-6 *SPINT1* expression is elevated in homozygous knockout mutant Bar charts showing the relative fold change of expression of a) steady-state and b) nascent level of *SPINT1* in wild-type HEK293, mutant cell lines heterozygous (SPINT1$^{+/-}$) and homozygous (SPINT1$^{-/-}$) for *SPINT1* predicted tsRNA target site. c) Images of western blot showing protein signals of SPINT1 in wild type, SPINT1$^{+/-}$ and SPINT1$^{-/-}$ cells. d) Image of DNA gel showing signals of RT-PCR products corresponding to the predicted tsRNA target site of *SPINT1*.

4.6 Role of Ago2 in tsRNA-mediated gene silencing

Previously, I found that selected tsRNA target genes were upregulated upon Ago2 knockdown (Section 3.5) and genome-wide analysis of chrRNA-seq showed that around

2700 genes were upregulated at the chromatin level when Ago2 was depleted (Section 4.2). Hence, I hypothesise that Ago2 has an important role in the model of NRS.

To validate the importance of Ago2 in NRS, I treated wild type HEK293T cells with α-amanitin, an inhibitor of transcription[156,157] for 24 hours, followed by subcellular fractionation, and examined the protein levels of Ago2 in the cytoplasmic and chromatin fractions by western blot. Levels of Ago2 were drastically reduced in the chromatin fraction when cells were treated by α-amanitin. Cytoplasmic Ago2 levels were comparable in both conditions. This suggests that the association between Ago2 and the chromatin is dependent upon transcription.

In the previous chapter, I discussed that tsRNA sequences were found in Ago2 PAR-CLIP data, indicating that tsRNA are associated with tsRNA. To verify this data, I selected itRF$^{Arg\text{-}CCT}$, which was significantly downregulated upon Dicer depletion, and tested whether it is physically associated with Ago2. Immunoprecipitation of Ago2 followed by northern blotting revealed that tsRNA^{Arg} was associated with Ago2, supporting the notion that tsRNA is bound by Ago.

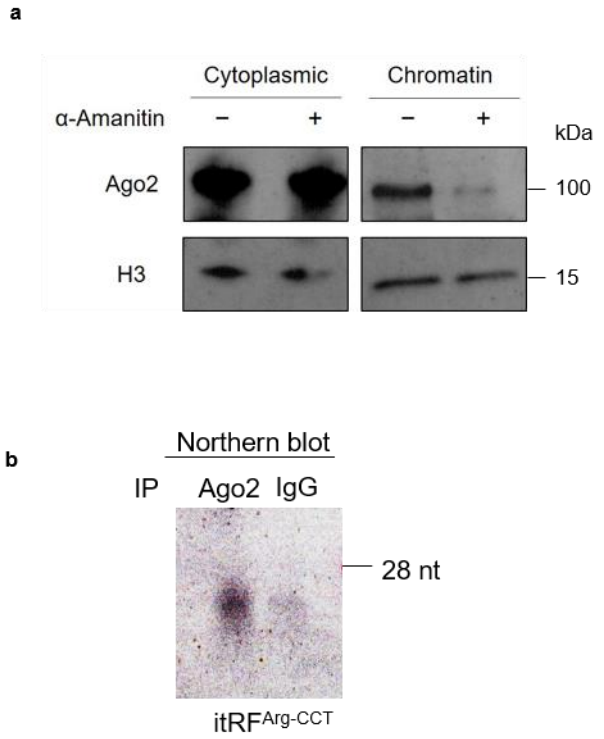


Figure 4-7 Role of Ago2 in tsRNA-mediated gene silencing a) Images of western blot showing the protein signals of Argonaute 2 (Ago2) in the cytoplasmic and chromatin fractions of wild type HEK293T cells, with H3 as a loading control. b) Image of northern blot showing the signals of itRF$^{Arg\text{-}CCT}$ in samples from Ago2 and IgG immunoprecipitation, respectively.

4.7 Ago2 slicer activity in nascent RNA silencing

Next, I investigated the destabilisation mechanism of nascent RNA in NRS. In siRNA mediated PTGS, Ago2, as part of the RISC, is guided by siRNA to target mRNA for silencing. The silencing is promoted by the slicer activity of Ago2, which leads to mRNA cleavage[38].

Ago2 *in vitro* cleavage assay was employed to study whether the slicer activity of Ago2 is required for NRS. Recombinant Ago2, which was pre-incubated with tsSPINT1, was incubated with purified 450-bp RNA substrates, whose sequences were derived from intron 2 of *SPINT1* containing the predicted target site, in the presence of a cocktail of ATP and Mg^{2+}. Samples were collected at different timepoints for further analysis. Northern blotting data showed the accumulation of a 150-bp RNA product which

corresponds to the smaller RNA segment which was expected from a specific cleavage at the predicted target site. The absence of tsSPINT1, Ago2 or ATP/Mg^{2+} prevented the cleavage of RNA substrate (Figure 4-8a). This suggests that tsRNA-bound Ago2 can cleave tsRNA target transcript *in vitro* in a sequence-specific manner.

Plasmids containing wild type and ($hAgo^{wt}$) mutant human Ago2 ($hAgo2^{D669A}$) respectively, the latter of which lacks endonuclease activity but can still bind sRNA and carry out translational inhibition[38], were constructed to test whether the slicer activity of Ago2 is essential for NRS. To this end, I transfected plasmids expressing $hAgo^{wt}$ and $hAgo2^{D669A}$ into an Ago2-knockout ($Ago2^{-/-}$) mouse embryonic fibroblasts (MEF) cell line[135] and measured the nascent RNA level expression of *Egfr* by intron-based qRT-PCR 48 hours post-transfection. Results showed that expression of $hAgo2^{wt}$ restored the silencing effect on *Egfr* at the nascent RNA level. In contrast, $hAgo2^{D669A}$ had no effect on the nascent expression of *Egfr* (Figure 4-8b). This data corroborates the idea that the slicer activity is important for NRS *in vivo*. The expression of the plasmids was confirmed by qRT-PCR using primers targeting the region spanning hAgo2 and FLAG-tag, with the plasmid expressing $hAgo2^{D669A}$ being expressed at a higher level (Figure 4-8c).

Subsequent western analysis of $Ago2^{-/-}$ MEFs transfected with plasmids expressing $hAgo2^{wt}$ and $hAgo2^{D669A}$ demonstrated that both $hAgo2^{wt}$ and $hAgo2^{D669A}$ led to decreased levels of Egfr proteins. This observation demonstrated that despite having no effect on *Egfr* at the nascent RNA level, slicer-inactivated $hAgo2^{D669A}$ restored miRNA mediated PTGS in the cells.

In summary, in the presence of tsRNA, Ago2 can cleave target RNA specifically *in vitro* and the slicer activity of Ago2 may be involved in gene silencing by NRS.

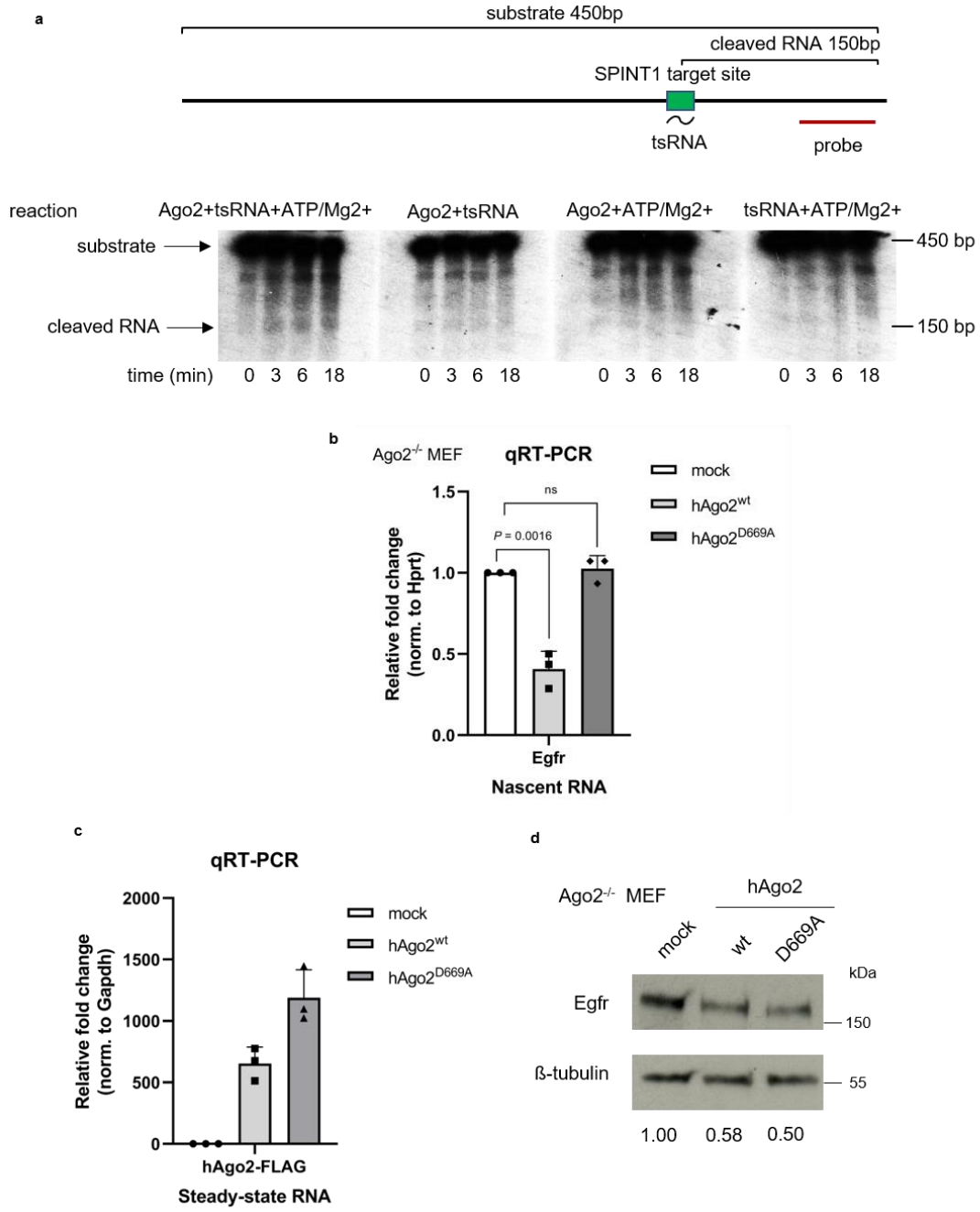


Figure 4-8 Ago2 slicer activity in NRS a) Images of northern blots showing signals of the tested RNA substrate derived from the genomic sequences of *SPINT1* containing its predicted tsRNA target site and flanking sequences, and the resulting cleaved RNA at different timepoints after the addition of components indicated above each image. The experimental strategy is illustrated on top. b) Bar charts showing the relative fold change of the nascent RNA level of *EGFR* (*n = 3*), measured by qRT-PCR, in Ago knockout ($Ago^{-/-}$) mouse embryonic fibroblasts (MEFs) transfected with plasmids expressing wild type human Ago2 ($hAgo2^{wt}$) and mutant (D669A) human Ago2 ($hAgo2^{D669A}$). c) Bar charts showing the relative fold change of the steady-state level of human *Ago2-FLAG* (*n = 3*), measured by qRT-PCR, in mock transfected $Ago2^{-/-}$ MEFs and $Ago2^{-/-}$ MEFs transfected with plasmids expression $hAgo2^{wt}$ and $hAgo^{D669A}$, respectively. d) Images of western blot showing protein signals of Egfr, with β-tubulin as loading control, in mock

transfected Ago2$^{-/-}$ MEFs and Ago2$^{-/-}$ MEFs transfected with plasmids expression hAgo2^{wt} and hAgo^{D669A}, respectively.

4.8 Discussion

In this chapter, I studied the molecular mechanism of NRS in more detail.

Using genome-wide quantification of chrRNA, I showed that similar genes were upregulated in Dicer and Ago2 knockdown cells. The largest class of nascent RNAs that are retained on the chromatin are protein-coding pre-mRNAs through the association with RNA pol II in the process of transcription[158]. The readout of chrRNA-seq was therefore used as a proxy for the nascent RNA expression of the tsRNA target genes. As a result, I was able to narrow down the list of predicted target genes to those which are only upregulated at the nascent RNA level.

Using subcellular fractionation followed by northern blotting, I showed that tsRNA are present in both cytoplasm and nucleus. Detection of exogenous, fluorescently labelled tsRNA mimics in the nucleus implies that tsRNA has role in the nucleus. This, however, does not mean that tsRNA has a 'functional' role in the nucleus. The wide-spreading foci formed by fluorescently labelled *EGFR*-targeting tsRNA suggests that the tsRNA have more than one target, which is consistent with our bioinformatic prediction.

Transfection of *SPINT1* targeting single-stranded tsRNA mimics into wild type HEK293T cells led to the silencing of nascent *SPINT1* RNA. The same tsRNA mimics specifically alleviated the upregulation of *SPINT1* caused by Dicer depletion, showing that increasing tsRNA *in vivo* can have an effect on the expression of its target gene. In this case, this tsSPINT1 (also tRF-3$^{Gln\text{-}CTG}$) was predicted to target intron 2 of *SPINT1*(hg38; chr15: 40,848,231-40,848,252). As the nascent expression of *SPINT1* was affected, the interaction must occur in the nucleus. However, this experiment cannot

prove that the silencing I observed is the outcome of the interaction between the tsRNA and the intronic target site.

To demonstrate that the predicted target site has a regulatory role in the expression of *SPINT1*, I deleted the genomic region of *SPINT1* containing the predicted target site using CRISPR-Cas9. The result evidently showed that the intronic target site is crucial for the silencing of *SPINT1*. Careful examination demonstrated that the upregulation is of *SPINT1* upon deletion of this intronic target site is marginal in comparison to the upregulation caused by Dicer and Ago2 knockdown. I hypothesise that there might be multiple target sites for *SPINT1,* and the ablation of a single target site will not have an effect as large as the depletion of the entire Dicer-dependent tsRNA population.

Based on the well-studied nuclear localisation of Ago2[70,159] and the observation that predicted tsRNA target genes are upregulated significantly upon Ago2 depletion, I hypothesised that Ago2 is guided by Dicer-dependent tsRNA to carry out NRS. The de-association of Ago2 from the chromatin as a consequence of transcription repression implies that a large proportion of Ago2 is only recruited to the chromatin during active transcription. This observation supports the hypothesis that they are guided by Dicer-dependent tsRNA to target nascent RNA for silencing.

The role of Ago2 in NRS is further elucidated by the Ago2 *in vitro* cleavage assay and rescue experiment in which wild-type and mutant versions of hAgo2 were expressed in $Ago2^{-/-}$ MEFs. The slicer activity of Ago2 is the driver behind mRNA target cleavage in miRNA and siRNA mediated PTGS[37,38]. Even though it was previously reported that only extensive matching between regulatory sRNA and target mRNA can lead to cleavage *in vivo*[160], recent high-throughput analyses showed that Ago2 cleavage can be induced by limited complementarity[161]; insertions of up to 8 nucleotides can be tolerated by the

RISC[162]. This might explain why specific cleavage was observed in the Ago2 *in vitro* cleavage assay. It is also possible that tsRNAs, having originated from heavily modified tRNAs, have structural features that are absent in miRNA and siRNA, thereby allowing the activation of Ago2 slicer activity.

The silencing of *Egfr* at the nascent transcript level in $Ago2^{-/-}$ MEFs was rescued by $hAgo2^{wt}$ but not $hAgo2^{D669A}$. This experiment shows that the slicer activity of Ago2 may be important in NRS. However, it is important to note that this experiment is conducted in mouse cells based on the assumption that *Egfr* is a target for tsRNA-mediated NRS in the organism. Further analyses are required to show that *Egfr* is a *bona fide* target of tsRNA in MEFs and whether the pairing between *Egfr* target site and tsRNA allows for the activation of Ago2 slicer activity. Apart from the slicer activity, other degradation machineries may be responsible for tsRNA-mediated NRS.

5 Therapeutic potential of tsRNAs

5.1 Introduction

In Chapter 5, I intend to explore the potential of tsRNA mimics being used for therapeutic purposes. As Dicer-dependent tsRNAs can carry out NRS, they could be a great addition to the growing repertoire of small RNA therapeutics. Parts of the data I show here are included in a patent application[163] supported by Cancer Research UK.

5.2 Association between tsRNA target genes and diseases

First, to understand the biological relevance of NRS, a disease-gene association analysis was conducted by employing DisGeNET[145,146], a database for human disease-related genes. tsRNA target genes were found to be significantly disease-associated, relative to non-target genes (Appendix 3). Close inspection of the data revealed that the target genes were preferentially associated with cancers, but also with diseases of the nervous and cardiovascular systems.

a

Contingency table	Target genes	Non-target genes
Disease-associated	1225	15320
Non disease-associated	339	19759

$P < 2.2 \times 10^{-16}$
Fisher's exact test

b

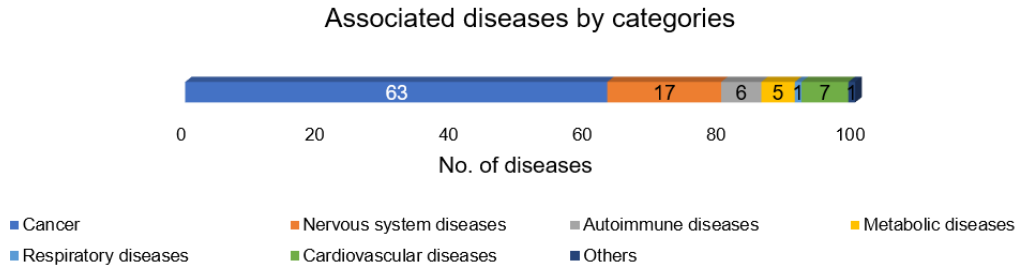


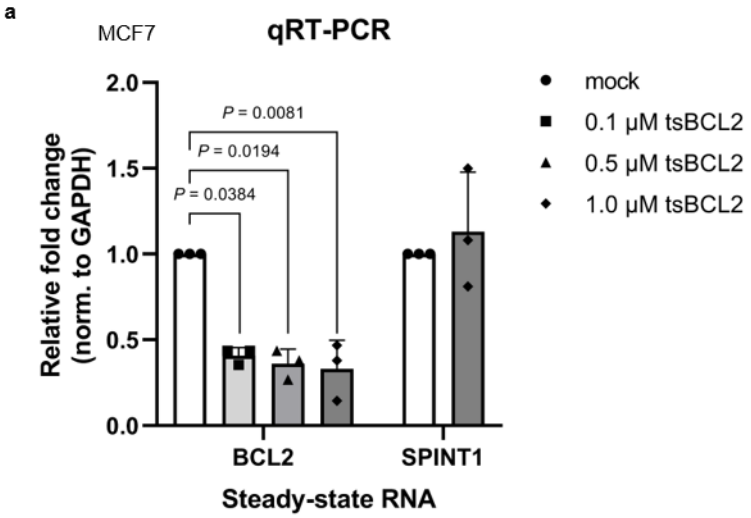
Figure 5-1 Disease-tsRNA target gene association analysis a) Contingency table summarising the relationship between the identity of a gene being a tsRNA target or non-target gene and their association with diseases. b) Horizontal stacked bar chart representing disease categories related to the 100 top diseases that are associated with tsRNA target genes (Appendix 3). This analysis was performed by Margarita Schlackow.

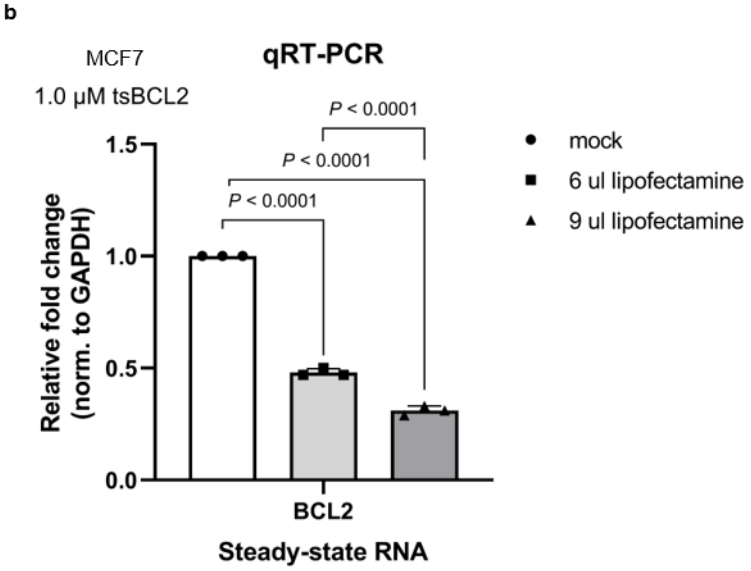
5.3 Silencing of proto-oncogenic genes using tsRNA mimics

To study whether tsRNA mimics can be used to silence proto-oncogenic genes, unmodified, single-stranded mimics of tsRNAs predicted to target intron 2 of *BCL2* (tsBCL2, also itRF$^{Arg\text{-}CCT}$) and intron 21 of *EGFR* (tsEGFR; tRF-5$^{Glu\text{-}CTC}$) were transfected into human breast cancer cell lines MCF7 and BT549, respectively.

Transfection of tsBCL2 into MCF7 cells led to the silencing of *BCL2* at the steady-state level. The silencing effect of tsBCL2, however, did not appear to be dose-dependent, which could be a consequence of lipofectamine saturation. Varying amount of lipofectamine were tested under the same tsBCL2 concentration, showing a stronger silencing effect when more lipofectamine was used.

As BCL2 is an anti-apoptotic factor, I tested whether transfection of tsBCL2 could induce cell death in MCF7 cells. Western analysis revealed that BCL2 protein levels were reduced by increasing concentration of tsBCL2 (starting from 0.1 μM). Levels of cleaved Caspase-9, which signifies activation of apoptosis, were markedly increased at concentrations 0.2 μM and 0.4 μM.


a
MCF7
qRT-PCR
Relative fold change (norm. to GAPDH)
2.0
1.5
1.0
0.5
0.0
P = 0.0081
P = 0.0194
P = 0.0384
BCL2
SPINT1
Steady-state RNA
mock
0.1 μM tsBCL2
0.5 μM tsBCL2
1.0 μM tsBCL2


b
MCF7
1.0 μM tsBCL2
qRT-PCR
Relative fold change (norm. to GAPDH)
1.5
1.0
0.5
0.0
P < 0.0001
P < 0.0001
P < 0.0001
BCL2
Steady-state RNA
mock
6 ul lipofectamine
9 ul lipofectamine

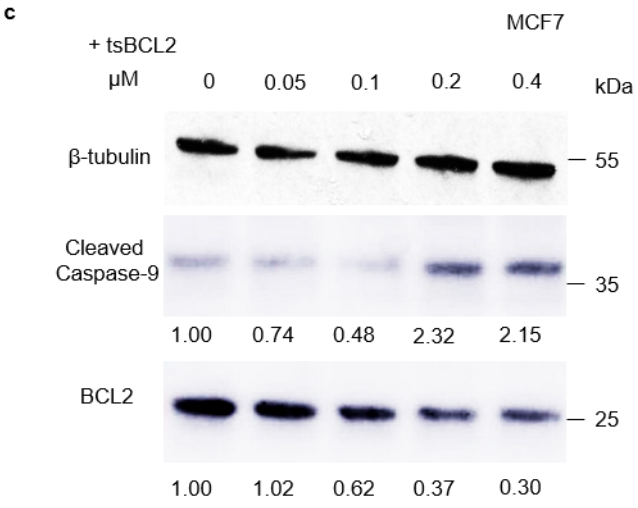

c
MCF7
+ tsBCL2
μM
0
0.05
0.1
0.2
0.4
kDa
β-tubulin
55
Cleaved Caspase-9
35
1.00
0.74
0.48
2.32
2.15
BCL2
25
1.00
1.02
0.62
0.37
0.30

Figure 5-2 *BCL2* is silenced by tsBCL2 transfection a) Bar charts showing the relative fold change of steady-state level of *BCL2* and *SPINT1* (negative control) (*n = 3*), measured by qRT-PCR, in mock transfected MCF7 cells and MCF7 cells transfected with varying concentrations (0.1, 0.5 and 1.0 μM) of tsRNA against *BCL2* (tsBCL2, also itRF$^{Arg\text{-}CCT}$). b) Bar charts showing the relative fold change of steady-state level of *BCL2* (*n = 3*), measured by qRT-PCR, in mock transfected MCF7 cells and MCF7 cells transfected with 1.0 μM tsBCL2 using varying amount (6 and 9 μl) of lipofectamine RNAiMAX, c) Images of western blot showing protein signals of BCL2, cleaved Caspase-9 in MCF7 cells transfected with increasing concentrations of tsBCL2. The qRT-PCR experiments were performed by Kaiyang Song under my supervision.

I compared the silencing effect of tsEGFR with that of siEGFR by transfecting both sRNAs into BT549 cells, respectively. qRT-PCR analysis showed that expression of *EGFR* at the steady-state level were significantly reduced upon transfection of tsEGFR and siEGFR, with the latter being more effective. Such silencing effect was observed at the protein level, as evidenced by western blotting.

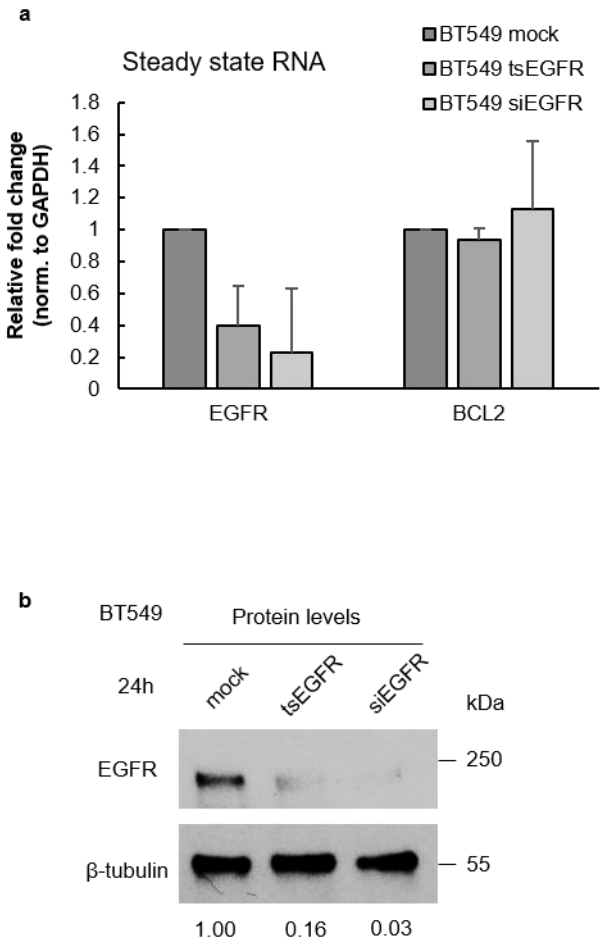


Figure 5-3 *EGFR* is silenced by tsEGFR transfection a) Bar charts showing the relative fold change of steady-state levels of *EGFR* and *BCL2* (negative control) (*n = 3*), measured by qRT-PCR, in mock transfected BT549 cells and BT549 cells transfected with tsRNA (tsEGFR, also tRF-5$^{Glu\text{-}CTC}$) and siRNA against *EGFR* (siEGFR), respectively. b) Images of western blot showing protein signals of EGFR in mock transfected BT549 cells and BT549 cells transfected with tsEGFR and siEGFR, respectively.

5.4 Effect of RNA modifications on the silencing of tsRNAs

A previous report showed that 5'-phosphorylated, single-stranded siRNA has a better silencing effect than non-phosphorylated versions[164]. Endogenous Dicer processing of siRNA also generates sRNA duplexes with 5' monophosphates[20–23].

qRT-PCR analysis showed that the silencing effect of non-phosphorylated tsBCL2 is weaker in comparison to the 5'-phosphorylated version, consistent with previous findings.

Transfection of unmodified, double-stranded (or duplex) tsBCL2 into MCF7 cells was shown to result in a markedly weaker silencing effect on *BCL2* than unmodified, single-stranded tsBCL2.

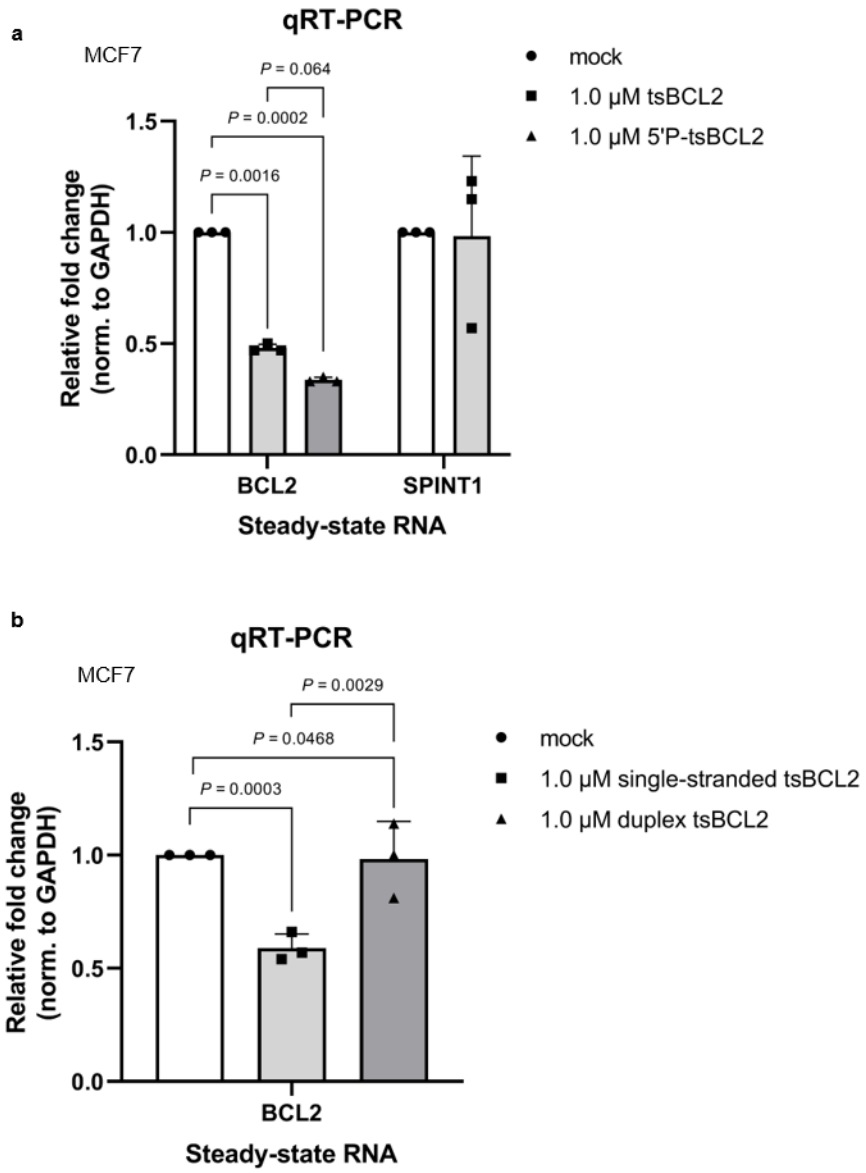


Figure 5-4 Silencing effects of 5' phosphorylated and duplex tsBCL2 a) Bar charts showing the relative fold change of steady-state level of *BCL2* and *SPINT1* (negative control) (*n = 3*), measured by qRT-PCR, in mock transfected MCF7 cells and MCF7 cells transfected with 1.0 μM unmodified and 5' phosphorylated tsRNA against *BCL2* (tsBCL2, also itRF$^{Arg\text{-}CCT}$), respectively. c) Bar charts showing the relative fold change of steady-state level of *BCL2* (*n = 3*), measured by qRT-PCR, in mock transfected MCF7 cells and MCF7 cells transfected with 1.0 μM single-stranded and duplex tsBCL2, respectively. These experiments were performed by Kaiyang Song under my supervision.

5.5 Intron- versus exon-targeting tsRNA

In previous chapters, I have been studying NRS as an intron-targeting mechanism. However, in the bioinformatic prediction of target genes, exon target sites were also identified. To test this, I sought a tsRNA candidate which was predicted to target exon 2 of BCL2 (exonic tsBCL2, also tRF-5$^{Ala\text{-}AGC}$) and compared its silencing effect with that of intronic tsBCL2, which was used previously (Section 5.3 & 5.4). Transfection of exonic and intronic tsBCL2 into MCF7 cells followed by qRT-PCR revealed that both versions could reduce *BCL2* expression significantly. Such reduction occurs at both nascent and steady-state RNA levels. The difference between the silencing effect induced by exonic and intronic tsBCL2 was not significant.

This demonstrates that NRS is not limited to intron-targeting. Unmodified tsRNA that targets exons can also lead to gene silencing through NRS.

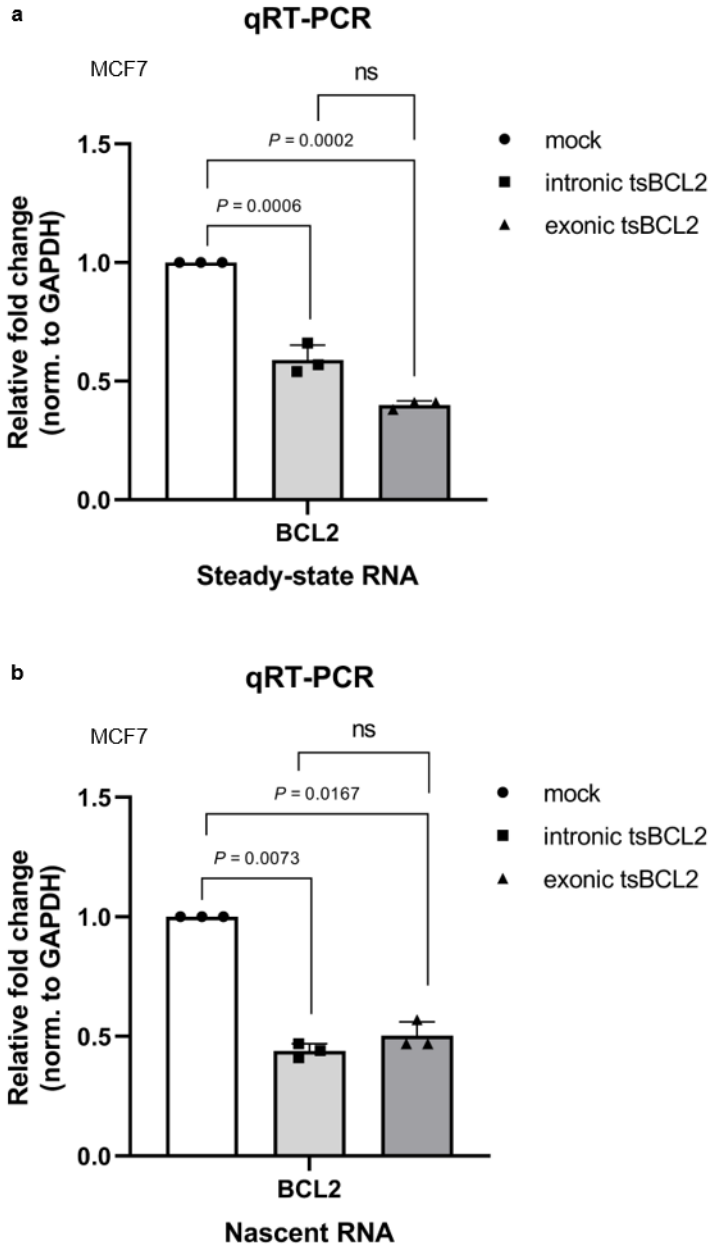


Figure 5-5 Comparison between intron- and exon-targeting tsBCL2 Bar charts showing the relative fold change of a) steady-state and b) nascent levels of *BCL2* (*n = 3*), measured by qRT-PCR, in mock transfected MCF7 cells and MCF7 cells transfected with 1.0 μM intron- and exon-targeting tsRNA against *BCL2* (tsBCL2), respectively. These experiments were performed by Kaiyang Song under my supervision.

5.6 Silencing of *TTR*

The first commercial-based RNAi therapeutic approved by the United States FDA and the European Commission (EC) was ONPATTRO® (patisiran) developed by Alnylam

Pharmaceuticals. Patisiran is used for the treatment of hATTR by reducing the expression of *TTR* through siRNA mediated PTGS.

The GalNAc-siRNA conjugate delivery platform has allowed effective delivery of siRNA to the hepatocytes[128]. I am therefore keen on testing whether tsRNA can be an addition to the repertoire of sRNA therapeutics against disease-associated genes in the liver context.

As TTR was not on the original list of predicted target genes in HEK293T cells, thresholds for the bioinformatic prediction were lowered to identify potential tsRNA candidates for TTR. From the predicted list, I selected 10 tsRNA candidates (details in Section 2.6.5) based on the following criteria:

1. They were associated with Dicer and Ago proteins in PAR-CLIP.
2. They were significantly downregulated upon Dicer depletion (in at least two biological repeats) in sRNA-seq.
3. They could be found in MINTbase, an online repository of nuclear and mitochondrial tsRNAs from multiple human tissues[165,166].
4. They were predicted to target at least one intronic site of *TTR*

The potential target sites of the 10 tsRNA candidates, named according to their origins, are displayed in Figure 5-6.

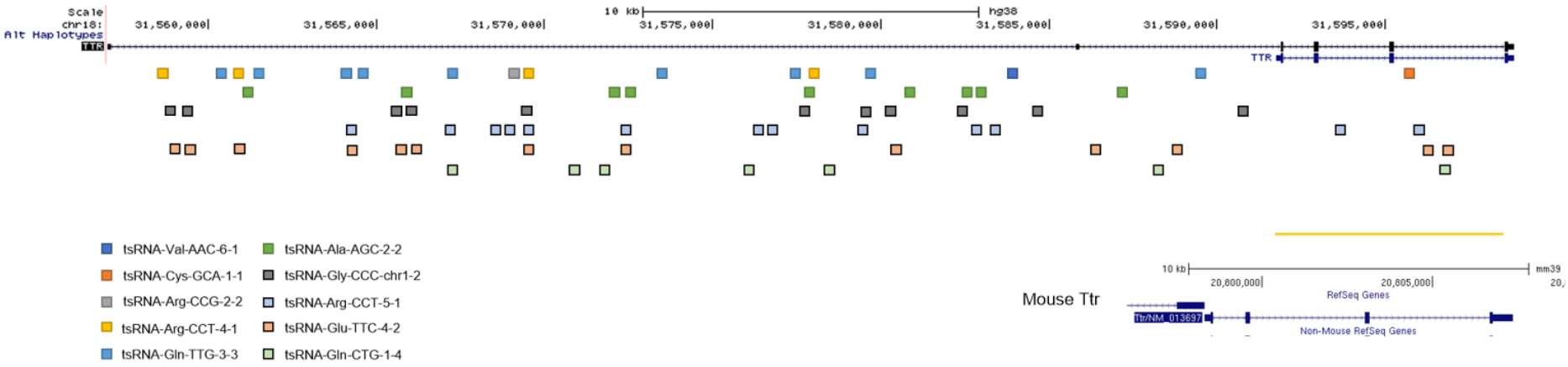

Scale
chr18:
Alt Haplotypes
TTR
31,560,000
31,565,000
31,570,000
10 kb
31,575,000
31,580,000
hg38
31,585,000
31,590,000
31,595,000
TTR
tsRNA-Val-AAC-6-1
tsRNA-Cys-GCA-1-1
tsRNA-Arg-CCG-2-2
tsRNA-Arg-CCT-4-1
tsRNA-Gln-TTG-3-3
tsRNA-Ala-AGC-2-2
tsRNA-Gly-CCC-chr1-2
tsRNA-Arg-CCT-5-1
tsRNA-Glu-TTC-4-2
tsRNA-Gln-CTG-1-4
Mouse Ttr
10 kb
mm39
20,800,000
20,805,000
RefSeq Genes
Ttr/NM_013697
Non-Mouse RefSeq Genes

Figure 5-6 Predicted target sites of tsTTR candidates Diagram illustrating the predicted target sites of tsRNA against *TTR* (tsTTR). tsTTRs are labelled based on their origins and colour coded. Yellow bar indicates the region of human *TTR* which corresponds to the mouse *Ttr* gene.

To test the silencing effect of these 10 tsTTR candidates, I employed the human hepatocellular carcinoma cell line, HepG2. Interestingly, transfection of all 10 tsTTR candidates led to the silencing of *TTR* to varying degrees, with 4 candidates (tRF-3$^{Gln-CTG}$, itRF$^{Glu-TTC}$, tRF-5$^{Val-AAC}$ and itRF$^{Arg-CCG}$) resulting in more than 50% reduction of steady-state *TTR* expression (Figure 5-7a).

To further the translational development of the tsRNA gene silencing technology, pre-clinical research data would be required. To that end, I took the decision to test the tsRNA candidates for *TTR* silencing in the mouse BWGT3 hepatoma cell line. It is important to note that the mouse *Ttr* gene is similar to the shorter isoform for human *TTR*. I have therefore narrowed down the list to 4 tsRNA candidates which were predicted to target both the shorter and longer isoforms of *TTR* (indicated by the yellow line in Figure 5-6).

Transfection of these 4 tsRNA candidates (tRF-5$^{Cys-GCA}$, itRF$^{Glu-TTC}$, itRF$^{Arg-CCT}$ and tRF-3$^{Gln-CTG}$) and another tsRNA which targets the large isoform of *TTR* (tRF-5$^{Val-AAC}$) into mouse hepatoma BWGT3 cells followed by qRT-PCR revealed that itRF$^{Glu-TTC}$ could significantly silence *Ttr* expression at the steady-state level in BWGT3 cells (Figure 5-7b), without affecting the expression of *Gapdh*, a non-target gene in this case. This result indicates that human tsRNA could be employed to silence *Ttr* in mouse cells.

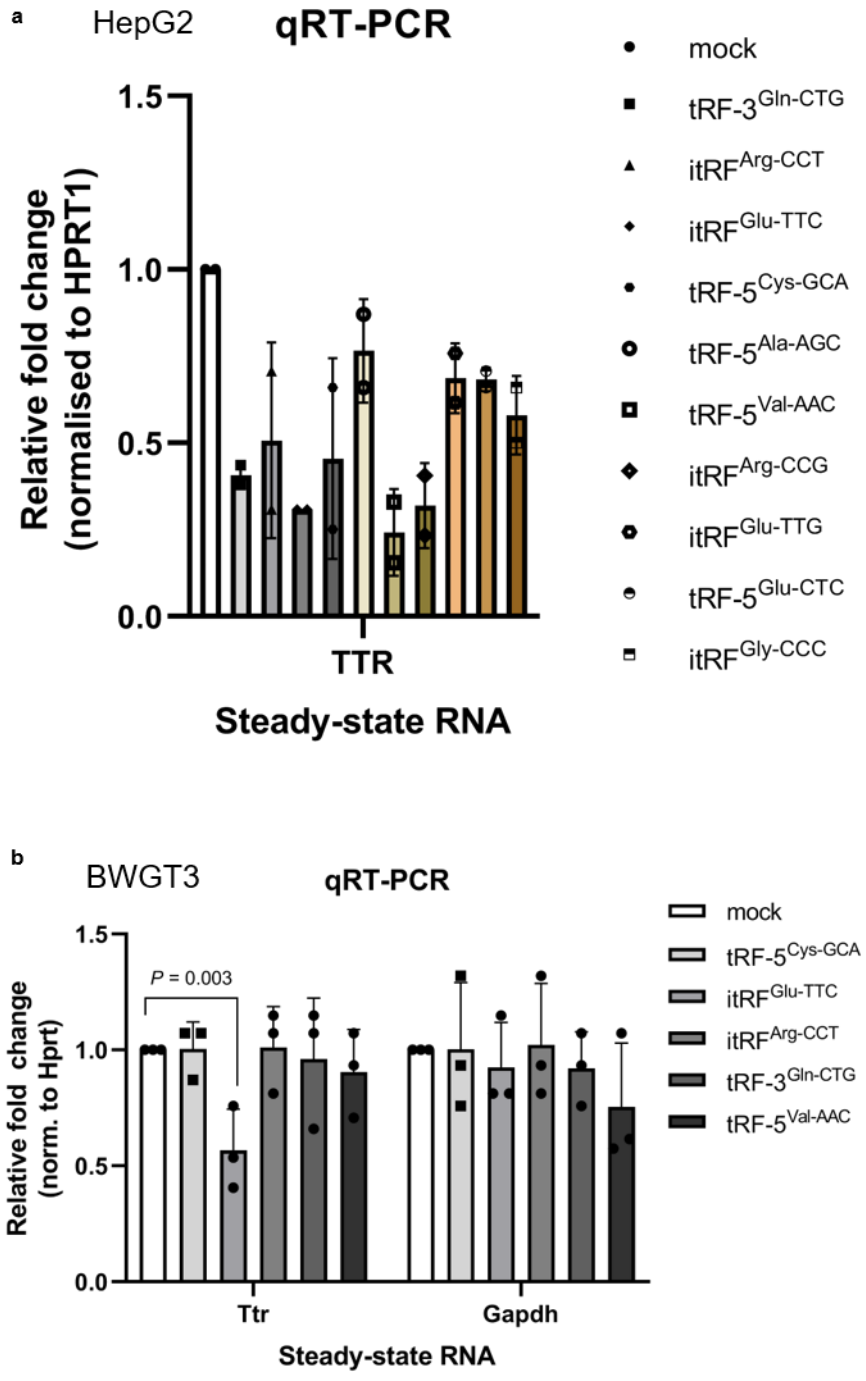


Figure 5-7 *TTR* is silenced by transfection of tsTTR candidates Bar chart showing the relative fold change of a) human *TTR* (*n = 2*) and b) mouse *Ttr* and *Gapdh* (negative control) (*n = 3*), measured by qRT-PCR, in mock transfected a) human hepatoma HepG2 cells and b) murine hepatoma BWGT3 cells and BWGT3 cells transfected with 1.0 µM tsRNA candidates predicted to target human *TTR* (tsTTR).

5.7 Discussion

In this chapter, I explored the therapeutic potential of tsRNA in the context of cancer, by employing multiple cancer cell lines.

Disease-gene association analysis showed that tsRNA target genes were highly associated with diseases, suggesting the importance of tsRNA mediated NRS. The limitation of this analysis lies in the fact that DisGeNET does not have information on whether the diseases were caused by an upregulation of the associated gene. To circumvent this caveat, I focussed on genes that are upregulated in cancer cell lines to test the silencing efficacy of tsRNAs.

Inhibition of anti-apoptotic factor BCL2 can activate cell death and has been used a strategy to treat cancer, especially leukemia[167–169]. This motivated me to test whether tsRNA could be used to silence *BCL2* through NRS in MCF7 cells. Unmodified, single-stranded tsBCL2 mimics effectively silenced *BCL2* by more than 50%; similar effects were observed at the protein levels at high concentrations. The activation of apoptosis was reflected by increasing amount of cleaved Caspase-9.

Overexpression of *EGFR* is associated with tumorigenesis and poor prognosis of various cancers[170,171]. EGFR-targeted therapies have been developed for cancer treatment[172]. In BT549 cells where EGFR is highly expressed[173], transfection of tsEGFR led to a decrease in *EGFR* expression and in turn reduction of EGFR protein levels. The silencing effect of tsEGFR is comparable to that of commercial-based siEGFR.

Another area I explored is whether 5'-phosphorylation and duplex formation can enhance the silencing effect of tsBCL2. As expected, and similar to siRNA, 5'-phosphorylation increased the silencing effect of single-stranded tsBCL2 but duplex form of tsBCL2 failed to induce silencing. It is important to note that the duplex form of tsBCL2 is formed

by hybridising two fully complementary tsRNA sequences without any modification or overhangs. I suspect that such stable duplex formation might prevent strand separation and the formation of active RISC. It is important to note that the transfection of duplex tsBCL2 had caused significant cell death, which I believe is due to the activation of interferon response, which is commonly induced by the introduction of unmodified, exogenous dsRNA agents.

Finally, I tested the efficacy of tsTTR in the silencing of *TTR* in both human and mouse cell systems. More data is required to understand whether TTR expression is reduced through NRS thereby allowing it to have an additive effect on current siRNA drugs used in the treatment of hATTR.

As I discussed in Section 1.5, RNAi therapeutics are expected to move beyond siRNA in the near future. tsRNA can therefore be an excellent addition to the current collection of RNAi therapeutics.

6 Conclusions and perspectives

6.1 Summary of thesis

Throughout this thesis, I have provided evidence for the role of tsRNAs in the regulation of gene expression.

First, I showed that some tsRNAs were dependent upon Dicer using northern blot and whole-genome sequencing of sRNA molecules. *In vitro* transcribed $tRNA^{Arg}$ was later shown to be cleaved by endogenous TAP-tagged Dicer, yielding a specific sRNA product.

Using an unbiased bioinformatic approach, genes targeted by Dicer- and Ago-associated tsRNAs were identified. Subsequent analyses demonstrated that Dicer-dependent tsRNA genes are regulated co-transcriptionally in an Ago2-dependent manner without affecting transcription. These observations had led to the hypothetic model of a novel gene silencing mechanism. NRS, as the name suggests, targets nascent transcripts for destabilisation, resulting in low expression of the target genes.

In Chapter 4, I showed a high overlap of elevated genes at the chromatin between Ago2 and Dicer depletion using chrRNA-seq. Drosha knockdown had a marginal effect on the nascent expression of genes in comparison. This supports a common mechanism between Dicer and Ago2 on nascent RNA regulation. Transfection of tsRNA labelled with fluorescent dye led to foci formation in both cytoplasm and nucleus, suggesting a nuclear role for tsRNA. Perhaps most interestingly, removal of the predicted target site embedded in intron 2 of *SPINT1* significantly increased the expression of the gene at nascent, steady-state and protein levels, providing evidence for intron being a regulatory platform for pre-mRNA. The association between Ago2 and the chromatin was markedly reduced when transcription was inhibited, further corroborating the NRS model. Interaction

between a selected tsRNA candidate with Ago2 has also been confirmed. Lastly, the slicer activity of Ago2 was shown to be essential for the effect of NRS.

tsRNA mimics were employed to silence disease-related genes in Chapter 5. *BCL2* and *EGFR*, both proto-oncogenic genes, were successfully knocked down by transfection of tsRNA mimics, selected based on bioinformatic prediction and other criteria. Interestingly, the silencing effect of tsEGFR is comparable to that of siRNA in BT549 cells, despite the difference in mode of action. 5' phosphorylation of tsBCL2 displayed a slightly stronger silencing effect than the unmodified version while duplex form of tsBCL2, without 5' phosphorylation and 3' overhangs, failed to silence *BCL2*. The silencing of *TTR* was tested in human and mouse liver cancer cells as a pilot experiment for further pre-clinical investigation.

6.2 Future work

More systematic work is required to better understand the role of Dicer in tsRNA biogenesis. *In silico* analyses have shown that some tRNAs can fold into long hairpin structures, which can be recognised by Dicer[102], but there is insufficient *in vitro* and *in vivo* structural data to corroborate this. The additional layer of RNA modifications should also be taken into account when predicting or resolving the alternative structure of tRNAs, as studies have found that tRNA modifications can determine the fate of a tRNA[121–123]. Furthermore, it would be interesting to know the proportion of tRNAs that are misfolded and prone to Dicer cleavage. I suspect only a very small portion of tRNAs will be folded into this alternative structure as most will still be used in translation.

Even though I have shown that selected tsRNAs can be found in the nucleus and cytoplasm and tsRNA may potentially carry out their functions in both subcellular compartments. Little is known about where they are being generated. Are the

alternatively folded tRNAs directly processed in the nucleus by nuclear Dicer or are they exported into the cytoplasm for processing? It is also unknown whether Dicer and other protein factors are required for the loading of tsRNA onto Ago proteins and where the loading of tsRNA takes place.

There are still many questions left to be answered about NRS. I have shown by transfection of tsRNA targeting intron of a selected target gene and deletion of the predicted target site embedded in the intron of the same target gene that introns can serve as a platform for RNAi. However, due to the limitation of time and resources, only one selected gene was investigated in detail. To capture the tsRNA interactome in the context of NRS, subcellular fractionation followed by CLASH-seq can be performed. Subcellular fractionation would allow the enrichment of nuclear and chromatin-associated transcripts while CLASH can identify *bona fide* tsRNA::pre-mRNA interactions.

Another important question to ask is about the degradation mechanism of NRS. Since many of the tsRNA targets were predicted through a miRNA target prediction algorithm, most of the target sites are not fully complementary to the tsRNAs, suggesting that the slicer activity of Ago2 may not be involved. To elucidate the degradation mechanism of NRS, immunoprecipitation of Ago2 in the chromatin fraction followed by mass spectrometry can be performed to identify at the binding partners of Ago2 at the chromatin.

In terms of bioinformatic prediction of sRNA targets, most previous studies have been strictly focused on exonic and 3’ UTR regions of target genes, completely disregarding potential interactions within the nucleus. I hope my observations would encourage others to broaden their scope and consider the potential interactions between sRNA and nuclear transcripts in the future.

As for the future for using tsRNA as a therapeutic, more research needs to investigate the loading mechanism of tsRNA to understand whether it is necessary to increase the amount of tsRNA being transported into the nucleus. This might explain why the silencing effect of tsRNA is slightly weaker than siRNA. Large scale, high throughput testing is required to understand the full therapeutic potential of tsRNAs. It would also be exciting to know whether tsRNA can be used in combination with other sRNAs to enhance silencing effects on target genes.

7 Bibliography

1. Carninci, P. *et al.* Molecular biology: The transcriptional landscape of the mammalian genome. *Science (80-.).* **309**, 1559–1563 (2005).

2. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

3. Palazzo, A. F. & Lee, E. S. Non-coding RNA: What is functional and what is junk? *Front. Genet.* **5**, (2015).

4. Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M. B. Annotating non-coding regions of the genome. *Nat. Rev. Genet.* **11**, 559–571 (2010).

5. Brosnan, C. A. & Voinnet, O. The long and the short of noncoding RNAs. *Current Opinion in Cell Biology* vol. 21 416–425 (2009).

6. Jawdekar, G. W. & Henry, R. W. Transcriptional regulation of human small nuclear RNA genes. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* vol. 1779 295–305 (2008).

7. Reichow, S. L., Hamma, T., Ferré-D'Amaré, A. R. & Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res.* **35**, 1452–1464 (2007).

8. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015).

9. Wu, X. *et al.* The Biogenesis and Functions of piRNAs in Human Diseases. *Molecular Therapy - Nucleic Acids* vol. 21 108–120 (2020).

10. Williams, Z. *et al.* Discovery and Characterization of piRNAs in the Human Fetal Ovary. *Cell Rep.* **13**, 854–863 (2015).

11. Keam, S. P. & Hutvagner, G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life (Basel, Switzerland)* **5**, 1638–1651 (2015).

12. Li, S., Xu, Z. & Sheng, J. tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes (Basel).* **9**, (2018).

13. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998).

14. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).

15. Caplen, N. J., Parrish, S., Imani, F., Fire, A. & Morgan, R. A. Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9742–9747 (2001).

16. Press release: The 2006 Nobel Prize in Physiology or Medicine - NobelPrize.org. https://www.nobelprize.org/prizes/medicine/2006/press-release/.

17. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).

18. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901–906 (2000).

19. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).

20. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).

21. Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science (80-.).* **293**, 2269–2271 (2001).

22. Provost, P. *et al.* Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.* **21**, 5864–5874 (2002).

23. Zhang, H., Kolb, F. A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).

24. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).

25. Kwon, S. C. *et al.* Structure of Human DROSHA. *Cell* **164**, 81–90 (2016).

26. Nguyen, T. A. *et al.* Functional anatomy of the human microprocessor. *Cell* **161**, 1374–1387 (2015).

27. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).

28. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).

29. Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* **14**, 447–459 (2013).

30. Noland, C. L., Ma, E. & Doudna, J. A. SiRNA Repositioning for Guide Strand Selection by Human Dicer Complexes. *Mol. Cell* **43**, 110–121 (2011).

31. Gredell, J. A., Dittmer, M. J., Wu, M., Chan, C. & Walton, S. P. Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry* **49**, 3148–3155 (2010).

32. Masliah, G. *et al.* Structural basis of si RNA recognition by TRBP double-stranded RNA binding domains . *EMBO J.* **37**, (2018).

33. Iwasaki, S. *et al.* Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39**, 292–299 (2010).

34. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* **19**, 145–151 (2012).

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

36. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).

37. Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).

38. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).

39. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of argonaute and its implications for RISC slicer activity. *Science (80-.).* **305**, 1434–1437 (2004).

40. Park, M. S. *et al.* Human Argonaute3 has slicer activity. *Nucleic Acids Res.* **45**, 11867–11877 (2017).

41. Ipsaro, J. J. & Joshua-Tor, L. From guide to target: Molecular insights into eukaryotic RNA-interference machinery. *Nature Structural and Molecular Biology* vol. 22 20–28 (2015).

42. Lai, E. C. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**, 363–364 (2002).

43. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of Mammalian MicroRNA Targets. *Cell* **115**, 787–798 (2003).

44. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

45. Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: Targets and expression. *Nucleic Acids Res.* **36**, (2008).

46. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **11**, (2010).

47. Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome Biol.* **5**, (2003).

48. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **18**, 1165–1178 (2004).

49. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).

50. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of

Drosophila microRNA targets. *PLoS Biol.* **1**, (2003).

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

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

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

54. Wahle, E. & Winkler, G. S. RNA decay machines: Deadenylation by the Ccr4-Not and Pan2-Pan3 complexes. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* vol. 1829 561–570 (2013).

55. Jonas, S. & Izaurralde, E. The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes and Development* vol. 27 2628–2641 (2013).

56. Iwakawa, H. oki & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends in Cell Biology* vol. 25 651–665 (2015).

57. Wilson, R. C. & Doudna, J. A. Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* **42**, 217–239 (2013).

58. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics* vol. 14 100–112 (2013).

59. Martienssen, R. & Moazed, D. RNAi and heterochromatin assembly. *Cold Spring*

Harb. Perspect. Biol. **7**, a019323 (2015).

60. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science (80-.).* **297**, 1833–1837 (2002).

61. Reinhart, B. J. & Bartel, D. P. Small RNAs correspond to centromere heterochromatic repeats. *Science (80-.).* **297**, 1831 (2002).

62. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).

63. Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886 (2006).

64. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).

65. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).

66. Zhang, K., Mosch, K., Fischle, W. & Grewal, S. I. S. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat. Struct. Mol. Biol.* **15**, 381–388 (2008).

67. Gullerova, M. & Proudfoot, N. J. Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S. pombe. *Cell* **132**, 983–995 (2008).

68. Zeng, Y. & Cullen, B. R. RNA interference in human cells is restricted to the cytoplasm. *RNA* **8**, 855–860 (2002).

69. Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A. & Corey, D. R. RNAi factors are

present and active in human cell nuclei. *Cell Rep.* **6**, 211–221 (2014).

70. Sarshad, A. A. *et al.* Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Mol. Cell* **71**, 1040-1050.e8 (2018).

71. Guan, L., Karaiskos, S. & Grigoriev, A. Inferring targeting modes of Argonaute-loaded tRNA fragments. *RNA Biol.* 1–11 (2019) doi:10.1080/15476286.2019.1676633.

72. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).

73. Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754–765 (1988).

74. Carrillo Oesterreich, F. *et al.* Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165**, 372–381 (2016).

75. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* **28**, 1008–1019 (2018).

76. Nojima, T. *et al.* RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell* **72**, 369-379.e4 (2018).

77. Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol. Cell* **77**, 985-998.e8 (2020).

78. Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U. & Rasko, J. E. J. The

changing paradigm of intron retention: Regulation, ramifications and recipes. *Nucleic Acids Res.* **47**, 1149711513 (2019).

79. Boutz, P. L., Bhutkar, A. & Sharp, P. A. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* **29**, 63–80 (2015).

80. Rodnina, M. V. & Wintermeyer, W. The ribosome as a molecular machine: The mechanism of tRNA-mRNA movement in translocation. *Biochemical Society Transactions* vol. 39 658–662 (2011).

81. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184-189 (2016).

82. Phizicky, E. M. & Hopper, A. K. tRNA biology charges to the front. *Genes and Development* vol. 24 1832–1860 (2010).

83. Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3′ end formation. *Cell* **117**, 311–321 (2004).

84. Lipowsky, G. *et al.* Coordination of tRNA nuclear export with processing of tRNA. *RNA* **5**, 539–549 (1999).

85. Ibba, M. & Soll, D. Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry* vol. 69 617–650 (2000).

86. Holley, R. W. *et al.* Structure of a ribonucleic acid. *Science (80-.).* **147**, 1462–1465 (1965).

87. Kim, S. H. *et al.* Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science (80-.).* **185**, 435–440 (1974).

88. Lorenz, C., Lünse, C. E. & Mörl, M. Trna modifications: Impact on structure and thermal adaptation. *Biomolecules* vol. 7 (2017).

89. Cantara, W. A. *et al.* The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, (2011).

90. Robertus, J. D. *et al.* Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546–551 (1974).

91. Helm, M. *et al.* The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res.* **26**, 1636–1643 (1998).

92. F, V.-H. *et al.* A methyl group controls conformational equilibrium in human mitochondrial tRNA(Lys). *J. Am. Chem. Soc.* **129**, (2007).

93. Alexandrov, A. *et al.* Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* **21**, 87–96 (2006).

94. Lyons, S. M., Fay, M. M. & Ivanov, P. The role of RNA modifications in the regulation of tRNA cleavage. *FEBS Letters* vol. 592 2828–2844 (2018).

95. Kirchner, S. & Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* **16**, 98–112 (2015).

96. Levitz, R. *et al.* The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. *EMBO J.* **9**, 1383–1389 (1990).

97. Lee, S. R. & Collins, K. Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. *J. Biol. Chem.* **280**, 42744–42749 (2005).

98. Yamasaki, S., Ivanov, P., Hu, G. F. & Anderson, P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* **185**, 35–42 (2009).

99. Fu, H. *et al.* Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett.* **583**, 437–442 (2009).

100. Thompson, D. M. & Parker, R. Stressing out over tRNA cleavage. *Cell* **138**, 215–219 (2009).

101. Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P. & Anderson, P. Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Mol. Cell* **43**, 613–623 (2011).

102. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).

103. Calabrese, J. M., Seila, A. C., Yeo, G. W. & Sharp, P. A. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18097–18102 (2007).

104. Kawaji, H. *et al.* Hidden layers of human small RNAs. *BMC Genomics* **9**, (2008).

105. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649 (2009).

106. Cole, C. *et al.* Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* **15**, 2147–2160 (2009).

107. Haussecker, D. *et al.* Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**, 673–695 (2010).

108. Pederson, T. Regulatory RNAs derived from transfer RNA? *RNA* **16**, 1865–1869 (2010).

109. Kumar, P., Kuscu, C. & Dutta, A. Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends Biochem. Sci.* **41**, 679–689 (2016).

110. Soares, A. R. & Santos, M. Discovery and function of transfer RNA-derived fragments and their role in disease. *Wiley Interdiscip. Rev. RNA* **8**, (2017).

111. Green, D., Fraser, W. D. & Dalmay, T. Transfer RNA-derived small RNAs in the cancer transcriptome. *Pflugers Arch. Eur. J. Physiol.* **468**, 1041–1047 (2016).

112. Su, Z., Wilson, B., Kumar, P. & Dutta, A. Noncanonical Roles of tRNAs: tRNA Fragments and Beyond. *Annu. Rev. Genet.* **54**, annurev-genet-022620-101840 (2020).

113. Wang, Q. *et al.* Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Mol. Ther. J. Am. Soc. Gene Ther.* **21**, 368–379 (2013).

114. Maute, R. L. *et al.* tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1404–1409 (2013).

115. Soares, A. R. *et al.* Conserved and highly expressed tRNA derived fragments in zebrafish. *BMC Mol. Biol.* **16**, 22 (2015).

116. Martinez, G., Choudury, S. G. & Slotkin, R. K. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res.* **45**, 5142–5152 (2017).

117. Kumar, P., Anaya, J., Mudunuri, S. B. & Dutta, A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* **12**, 78 (2014).

118. Alves, C. S. *et al.* Genome-wide identification and characterization of tRNA-

derived RNA fragments in land plants. *Plant Mol. Biol.* **93**, 35–48 (2017).

119. Li, Z. *et al.* Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res.* **40**, 6787–6799 (2012).

120. Kuscu, C. *et al.* tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* **24**, 1093–1105 (2018).

121. Guzzi, N. *et al.* Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. *Cell* **173**, 1204-1216.e26 (2018).

122. Cosentino, C. *et al.* Pancreatic -cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes. *Nucleic Acids Res.* **46**, 10302–10318 (2018).

123. Reinsborough, C. W. *et al.* BCDIN3D regulates tRNAHis 3' fragment processing. *PLoS Genet.* **15**, e1008273 (2019).

124. Sobala, A. & Hutvagner, G. Small RNAs derived from the 5′ end of tRNAs can inhibit protein translation in human cells. *RNA Biol.* **10**, 553–563 (2013).

125. Kim, H. K. *et al.* A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature* **552**, 57–62 (2017).

126. Goodarzi, H. *et al.* Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell* **161**, 790–802 (2015).

127. Schorn, A. J., Gutbrod, M. J., LeBlanc, C. & Martienssen, R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* **170**, 61-71.e11 (2017).

128. Setten, R. L., Rossi, J. J. & Han, S. ping. The current state and future directions of RNAi-based therapeutics. *Nature Reviews Drug Discovery* vol. 18 421–446

(2019).

129. Adams, D. *et al.* Trial design and rationale for APOLLO, a Phase 3, placebo-controlled study of patisiran in patients with hereditary ATTR amyloidosis with polyneuropathy. *BMC Neurol.* **17**, (2017).

130. Adams, D. *et al.* Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *N. Engl. J. Med.* **379**, 11–21 (2018).

131. Nair, J. K. *et al.* Impact of enhanced metabolic stability on pharmacokinetics and pharmacodynamics of GalNAc-siRNA conjugates. *Nucleic Acids Res.* **45**, 10969–10977 (2017).

132. Kamerkar, S. *et al.* Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. *Nature* **546**, 498–503 (2017).

133. Rupaimoole, R. & Slack, F. J. MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases. *Nature Reviews Drug Discovery* vol. 16 203–221 (2017).

134. Schmitter, D. *et al.* Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res.* **34**, 4801–4815 (2006).

135. O'Carroll, D. *et al.* A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. *Genes Dev.* **21**, 1999–2004 (2007).

136. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).

137. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).

138. Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, (2014).

139. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

140. Hafner, M., Lianoglou, S., Tuschl, T. & Betel, D. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods* **58**, 94–105 (2012).

141. Rybak-Wolf, A. *et al.* A variety of dicer substrates in human and C. elegans. *Cell* **159**, 1153–1167 (2014).

142. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).

143. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

144. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2**, (2004).

145. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).

146. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855 (2020).

147. Huang, H. Y. *et al.* MiRTarBase 2020: Updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* **48**, D148–D154 (2020).

148. Kim, Y.-K., Kim, B. & Kim, V. N. Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**,

E1881-1889 (2016).

149. Poleshko, A. *et al.* Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* **171**, 573-587.e14 (2017).

150. Hasler, D. *et al.* The Lupus Autoantigen La Prevents Mis-channeling of tRNA Fragments into the Human MicroRNA Pathway. *Mol. Cell* **63**, 110–124 (2016).

151. Hoffer, P. *et al.* Posttranscriptional gene silencing in nuclei. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 409–414 (2011).

152. Ponzio, T. A., Yue, C. & Gainer, H. An intron-based real-time PCR method for measuring vasopressin gene transcription. *J. Neurosci. Methods* **164**, 149–154 (2007).

153. Chou, C.-H. *et al.* miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302 (2018).

154. Ghafouri-Fard, S. *et al.* Regulatory role of microRNAs on PTEN signaling. *Biomedicine and Pharmacotherapy* vol. 133 110986 (2021).

155. Mayer, A. *et al.* Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).

156. Chafin, D. R., Guo, H. & Price, D. H. Action of α-amanitin during pyrophosphorolysis and elongation by RNA polymerase II. *J. Biol. Chem.* **270**, 19114–19119 (1995).

157. Rudd, M. D. & Luse, D. S. Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J. Biol. Chem.* **271**, 21549–21558 (1996).

158. Li, X. & Fu, X. D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nature Reviews Genetics* vol. 20 503–519 (2019).

159. Huang, V. & Li, L. C. Demystifying the nuclear function of Argonaute proteins. *RNA Biology* vol. 11 18–24 (2014).

160. Chen, S. *et al.* Ago-2-Mediated slicer activity is essential for Anti-Flaviviral efficacy of RNAi. *PLoS One* **6**, (2011).

161. Jung, E., Seong, Y., Jeon, B., Song, H. & Kwon, Y. S. Global analysis of AGO2-bound RNAs reveals that miRNAs induce cleavage of target RNAs with limited complementarity. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860**, 1148–1158 (2017).

162. Becker, W. R. *et al.* High-Throughput Analysis Reveals Rules for Target RNA Binding and Cleavage by AGO2. *Mol. Cell* **75**, 741-755.e11 (2019).

163. GULLEROVA, M. MEDIATORS OF GENE SILENCING. (2020).

164. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).

165. Pliatsika, V., Loher, P., Telonis, A. G. & Rigoutsos, I. MINTbase: A framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. *Bioinformatics* **32**, 2481–2489 (2016).

166. Pliatsika, V. *et al.* MINTbase v2.0: A comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all the Cancer Genome Atlas projects. *Nucleic Acids Res.* **46**, D152–D159 (2018).

167. Roberts, A. W. *et al.* Targeting BCL2 with Venetoclax in Relapsed Chronic

Lymphocytic Leukemia. *N. Engl. J. Med.* **374**, 311–322 (2016).

168. Stilgenbauer, S. *et al.* Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study. *Lancet Oncol.* **17**, 768–778 (2016).

169. Jones, J. A. *et al.* Venetoclax for chronic lymphocytic leukaemia progressing after ibrutinib: an interim analysis of a multicentre, open-label, phase 2 trial. *Lancet Oncol.* **19**, 65–75 (2018).

170. Wang, Z. ErbB receptors and cancer. in *Methods in Molecular Biology* vol. 1652 3–35 (Humana Press Inc., 2017).

171. Arteaga, C. L. & Engelman, J. A. ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* vol. 25 282–303 (2014).

172. Xu, M. J., Johnson, D. E. & Grandis, J. R. EGFR-targeted therapies in the post-genomic era. *Cancer Metastasis Rev.* **36**, 463–473 (2017).

173. Irwin, M. E., Mueller, K. L., Bohin, N., Ge, Y. & Boerner, J. L. Lipid raft localization of EGFR alters the response of cancer cells to the EGFR tyrosine kinase inhibitor gefitinib. *J. Cell. Physiol.* **226**, 2316–2328 (2011).

8 Appendices

All appendices are submitted as “Additional Materials” alongside the digital copy of this thesis. The description of each file is as follows:

Appendix 1.xlsx contains the list of all the tsRNA target genes, their corresponding log2 fold change upon Dicer depletion (detected by sRNA-seq), and hits in Dicer and Ago1-4 PAR-CLIP data. Target genes which have no or weak evidence for being miRNA targets (by comparative analysis using miRTarBase) are displayed in separate sheets.

Appendix 2.txt contains the raw output of miRanda software, which was used to match Dicer- and Ago-associated tsRNA (identified from Dicer and Ago1-4 PAR-CLIP) with the genomic sequences of genes upregulated upon Dicer and Ago2 depletion (detected by chrRNA-seq) for tsRNA target gene prediction.

Appendix 3.jpg is an image containing the list of the 100 top ranked genes that are associated with diseases in DisGeNET. The heatmap was plotted by constructing a binary matrix for tsRNA target gene – disease association. The matrix was ordered first by column sums (genes), then by row sums (diseases).