Biogenesis of Y RNA–derived small RNAs

Adam Edward Hall

A thesis submitted for the degree of
Doctor of Philosophy

University of East Anglia
School of Biological Sciences

February 2013

'Nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under thy observation in life.'

Marcus Aurelius, Roman Emperor, c. 161 A.D.

Abstract

Small non-coding RNAs (sRNAs) constitute a significant portion of the transcriptome in eukaryotes. Many of these sRNAs regulate gene expression. Next-generation sequencing (NGS) has revealed a plethora of previously uncharacterised sRNAs with potential biological function, a number of which originate from longer RNAs. Here, the biogenesis of sRNAs derived from the non-coding Y RNAs (YsRNAs) was characterised as a model for understanding this emerging class of sRNA fragments. Y RNAs are highly conserved, 100 nt long molecules involved in DNA replication which bind to the autoimmune proteins Ro60 and La.

YsRNAs are produced in cells undergoing apoptosis. Here, it was demonstrated that YsRNAs are generated from the 5' and 3' ends of all four Y RNAs in stressed and unstressed cells. Furthermore, production of these fragments was observed in both cancerous and non-cancerous cells. Although YsRNAs have been proposed to have gene silencing activity, experiments done here found that YsRNAs do not enter the microRNA pathway and are not generated by the gene silencing-related protein Dicer. Furthermore, experiments established that the enzyme which produces fragments from tRNAs, angiogenin, was also not responsible for YsRNA generation.

Using mammalian cultured cells along with gene knockout and RNA interference (RNAi) technology, it was determined that RNase L contributed to YsRNA generation. Furthermore, the Y RNA binding protein Ro60 was shown to be essential for YsRNA production through a model of RNase protection. Analysis of deep sequencing data in Ro60 knockout cells revealed that many other sRNAs are also dependent on Ro60.

Finally, a 'high definition' (HD) protocol to improve NGS detection of sRNAs was tested. The HD protocol was found to be better at detecting sRNAs than current methods. This will facilitate more efficient detection of novel sRNAs in the future.

Table of Contents

Introduction

Y RNAs are fragmented into Y RNA-derived sRNAs in a range of cell types under stressed and non-stressed conditions

Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway

YsRNA generation is dependent on the autoimmune protein Ro60

Next-generation sequencing of a novel class of small RNAs

Summary and general discussion

List of figures

Chapter 6

Preface

This thesis is solely the result of my own work and contains nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. It describes experiments performed in the laboratory of Professor Tamas Dalmay at the School of Biological Sciences, University of East Anglia, between October 2009 and February 2013. No part of this thesis has been accepted, or is currently being submitted, for a degree or qualification at any other university. The length of this thesis does not exceed the permitted length set by the Academic Director of Research Degree Programmes, University of East Anglia.

Adam E. Hall

Norwich, February 2013

Acknowledgements

First and foremost, I would like to thank Professor Tamas Dalmay, my supervisor and group leader. His great expertise and experience in the small RNA field as well as his patience and untiring attention have provided the essential foundations for the successful laboratory in which I have been delighted to work. I am also very appreciative for his advice and comments during the writing of this thesis. Thanks also to my secondary supervisor Professor Andrea Munsterberg for her valuable input to this project.

During the course of my research, many other members of the laboratory, both past and present, have provided support, help and advice. In particular, I would like to thank Dr Karim Sorefan for engaging me in thoughtful discussions about my project and the scientific impact my research might have in science generally. I am also grateful to Dr Tibor Csorba, whose technical skill and enthusiastic approach 'at the bench' has not only inspired me but has added a more skilled and practical edge to my scientific curiosity. This thesis has also benefitted immeasurably from the help of three of the best bioinformaticians in the field; Dr Helio Pais, Matt Beckers and Dr Irina Mohorianu.

Finally, I would like to thank my family for their continuing love and support during my academic development. Specifically, I am indebted to my grandmother and fellow molecular biologist, Dr Milica Wass, who has been a continuing voice of encouragement in my pursuit of a scientific career, and whose first author *Nature* paper (Wass *et al.*, 1978) remains a benchmark I aspire to.

Publications

Some of the data and information contained in this thesis and generated during work done for the PhD has subsequently been published in the following:

Nicolas*, F.E., **Hall*, A.E.**, Csorba, T., Turnbull, C., & Dalmay, T. (2012) Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway. FEBS letters 586: 1226-1230.

Sorefan, K., Pais, H., **Hall, A. E.**, Kozomara, A., Griffiths-Jones, S., Moulton, V., & Dalmay, T. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence 3(4): 1-11.

Hall*, A.E., Turnbull*, C., & Dalmay T. (2013) Y RNAs: recent developments. Biomolecular Concepts (*In press*).

* Joint first authors

Abbreviations

AGO – Argonaute protein

ANG – Angiogenin ribonuclease

cDNA – Complementary DNA

CDS – Coding DNA sequence

DNA – Deoxyribonucleic acid

ES cell – Embryonic stem cell

G-phase – Gap phase of cell cycle

HD adapter – High definition adapter

hnRNP – Heterogeneous nuclear ribonucleoprotein

LNA – Locked nucleic acid

lncRNA – Long non-coding RNA

MBT – Midblastula transition

MEF – Mouse embryonic fibroblast

mES – Mouse embryonic stem cell

miRBase – Online database of microRNA sequences

miRISC – MicroRNA guided RNA-induced silencing complex

miRNA – MicroRNA

mRNA – Messenger RNA

ncRNA – Non-coding RNA

NGS – Next-generation sequencing

ORC – Origin recognition complex

PCR – Polymerase chain reaction

piRNA – Piwi-interacting RNAs

Poly(I :C) – Polyinosinic:polycytidylic acid

pre-miRNA – Precursor microRNA

pre-RC – Pre-replicative complex

pri-miRNA – Primary microRNA

qPCR – Quantitative PCR

RISC – RNA-induced silencing complex

RNA – Ribonucleic acid

RNAi – RNA interference

RNAPII – RNA polymerase II

RNase – Ribonuclease

RNH1 – Ribonuclease inhibitor of angiogenin

RNP – Ribonucleoprotein

RoBPI – Ro RNP binding protein

ROP-1 – Ro60 homologue in *Caenorhabditis elegans*

rosRNA – Ro60-dependent small RNA

Ro RNP – Ro60 ribonucleoprotein

RPM – Reads per million

RRM – RNA recognition motif

rRNA – Ribosomal RNA

Rsr – Ro60 related protein

RT-PCR – Reverse transcriptase PCR

sdRNA – SnoRNA-derived RNA

siRNA – Small interfering RNA

SLE – Systemic lupus erythematosus

snoRNA – Small nucleolar RNA

snRNA – Small nuclear RNA

S-phase – Synthesis phase of cell cycle

sRNA – Small RNA (<200 nt)

SS – Sjögren's syndrome

svRNA – Small vault-derived RNA

tiRNA – Stress-induced tRNA-derived RNA (tRNA halves)

tRF – Transfer RNA regulatory fragment

tRNA – Transfer RNA

U RNA – Small non-coding nUclear RNA

UTR – Untranslated region

UV – Ultraviolet light

vRNA – Vault RNA

Y RNA – CYtoplasmic RNA (Y1, Y3, Y4, Y5)

Y RNP – Y RNA ribonucleoprotein

YsRNA – Y RNA-derived small RNA

ZBP1 – Zipcode-binding protein 1

Chapter 1

Introduction

1.1 Overview: A genome governed by RNA

One of the most striking discoveries in recent times was the discovery of just how few genes there are in the human genome (Lander & International Human Genome Sequencing Consortium, 2001; Collins & International Human Genome Sequencing Consortium, 2004). Compared to the predicted 100,000 protein coding genes originally estimated, only 20,687 were finally identified recently following the completion of the ENCODE (Encyclopaedia of DNA Elements) project (Bernstein & ENCODE Project Consortium, 2012). Paradoxically, the human proteome is likely to hold an estimated 500,000 unique protein variants based on investigations into the yeast and fruit fly genomes (Ghaemmaghami *et al.*, 2003; Brunner *et al.*, 2007). The complexity of the proteome relative to the modest repertoire of *bona fide* genes is now becoming clear.

There is an overlapping arrangement of genetic information which is achieved by several mechanisms (Tuck & Tollervey, 2011). Firstly, variable transcription start sites for the same locus can give rise to transcripts which share a common sequence but are functionally independent. Secondly, genes can be differentially 'spliced' to give rise to varied isoforms. Thirdly, a single transcript can contain several genetic units which are differentially processed, and serve separate biological roles, (such as miRtrons - microRNAs which are present within the introns of protein-coding genes). And finally, transcripts with one function can be cleaved to give rise to smaller transcripts with distinct functions. This is notably demonstrated by small RNAs derived from longer RNAs.

RNA acts as an intermediary in the 'central dogma' of gene expression (the expression of genes from DNA to RNA, then to protein). However, the vast majority of cellular RNA is non-coding. Indeed, most of the genome in complex organisms is transcribed into non-coding RNA rather than protein coding transcripts (Mattick & Makunin, 2006). Only 1.5% of the human genome is protein coding yet around 80% is transcribed into RNA (Birney & ENCODE Project Consortium, 2007), making non-coding RNA a dominant feature in the cellular landscape. It has become increasingly apparent that the role of RNA in cellular processes is much more significant than first thought (Sharp, 2009). There are a vast number of distinct families which make up the 'RNA world' (Gilbert, 1986), having diverse roles such as gene silencing, intermediates during gene expression, structure formation, non-protein based enzymes and splicing.

Two characteristic features of these RNA families are the length and secondary structure of the RNA molecules. The structure of an RNA molecule is important to its function and depends on intramolecular sequence complementarity between nucleotide bases. It is sometimes misleading to think of each RNA family as a discrete group of molecules, as there can be significant interchangeability between groups. For instance, some transfer RNAs (tRNAs) and small nucleolar RNAs (snoRNAs) can enter small interfering RNA (siRNA)-like pathways (Pederson, 2010; Brameier *et al.*, 2011).

Y RNAs are one such family of non-coding RNAs. Y RNAs can be considered as a model of firstly, how a non-coding RNA can have a significant biological role in cellular biology, and secondly, how RNA fragmentation can

significantly increase the spectrum of distinct genetic units within the cell, giving rise to an overlapping arrangement of genetic information.

1.2   Evolution, structure and biogenesis of Y RNAs

Y RNAs were discovered following two investigations into the autoimmune disorder Systemic Lupus Erythematosus (SLE) (Lerner & Steitz, 1979; Lerner *et al.*, 1981), which is characterised by the inflammation of connective tissue. Lerner and colleagues aimed to uncover nuclear and cytoplasmic targets of the antibodies involved in this autoimmune response. They achieved this by immunoprecipitating cell extracts from mouse derived Ehrlich ascites tumour cells using serum from SLE patients containing antibodies to the Ro and La proteins. First, they investigated using only nuclear cell extracts from Ehrlich ascites, and identified a group of small non-coding nuclear RNA (U RNA) containing ribonucleoproteins (RNPs) as auto-antibody targets (Lerner & Steitz, 1979). In a later study in 1981, they used total extracts from the Ehrlich ascites along with serum containing antibodies to the Ro and La proteins (Lerner *et al.*, 1981). It was in this second set of experiments that they identified small cYtoplasmic RNAs contained in RNPs to which they designated the prefix Y, to distinguish them from their nUclear U RNA counterparts (Lerner *et al.*, 1981).

There are four principal Y RNA molecules (Figure 1.1) denoted Y1, Y3, Y4 and Y5 (Y2 is now considered to be a truncated form of Y1) (Wolin & Steitz, 1983). Y RNAs show high evolutionary conservation within the animal kingdom, and there is evidence that orthologs also exist in prokaryotes (Chen *et al.*, 2000;

Mosig *et al.*, 2007; Perreault *et al.*, 2007), suggesting that these genes may have an important function. Y3 shows the highest conservation in vertebrate species (Farris *et al.*, 1995). It is thought that the Y RNA family evolved from a common ancestral gene, which, through gene duplication, gave rise to Y1/Y3 in one lineage and Y4/Y5 in the other (Mosig *et al.*, 2007).

Many animal species do not contain the full Y RNA complement. For example, many rodents only possess Y1 and Y3 (Pruijn *et al.*, 1993), although both the mouse and rat genomes have a redundant 'fossil' Y5 RNA gene that is no longer expressed (Perreault *et al.*, 2007). In frogs, the Y5 gene has been further duplicated to give Yα (Mosig *et al.*, 2007). Interestingly, Y RNA genes have not been detected in the genome of the fruit fly *Drosophila melanogaster*, although putative Y RNA genes have been predicted in the mosquito genome (Perreault *et al.*, 2007). The first identifiable Y RNA in invertebrates was the single, ceY RNA of *Caenorhabditis elegans* (Vanhorn *et al.*, 1995). Interestingly the human Ro protein could bind to ceY RNA, but the equivalent Ro homologue in the nematode could not bind the human Y RNAs (Vanhorn *et al.*, 1995). Stem-bulge RNAs (sbRNAs) in the *Caenorhabditis elegans* have also been shown to be highly homologous to vertebrate Y RNAs, and contain a protein-binding motif for the Y RNA-associated protein Ro60 (Boria *et al.*, 2010). The Ro protein has not been detected in plants or fungi, and presumably these organisms also lack Y RNAs (Perreault *et al.*, 2007). However, a Ro ortholog (Rsr – Ro sixty related) has been documented in the eubacterium *Deinococcus radiodurans* and has been shown to bind an RNA which resembles a Y RNA (Chen *et al.*, 2000). As the Ro ortholog is a rare occurrence in prokaryotes it is likely that the gene has been acquired via horizontal gene transfer

from a eukaryotic organism (Chen & Wolin, 2004), with the ortholog having the

ability to bind to an endogenous, Y RNA-like molecule.
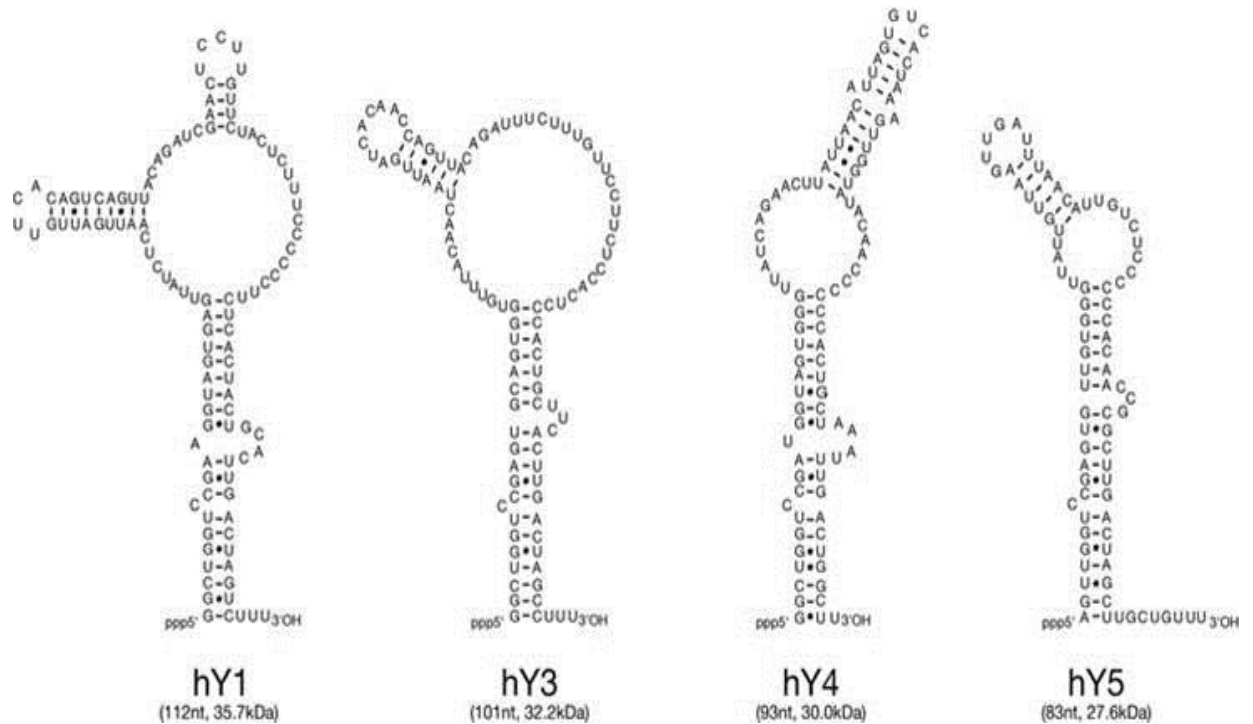
Figure 1.1 Predicted secondary structure of the four human Y RNAs.

All Y RNAs have at least two main stems separated by a large internal pyrimidine-rich loop. The most conserved region between the Y RNAs is the stem at the 3'/5' end. This is where the Ro60 protein binds, and involves the important Cytosine bulge which is crucial for this interaction (Pruijn *et al*., 1991). Both enzymatic cleavage and chemical modification experiments have been used to determine these structures shown (Wolin & Steitz, 1984; van Gelder *et al*., 1994). Figure reproduced from Christov *et al*., 2006.

Y RNAs are controlled by RNA polymerase III promoters (Wolin & Steitz, 1983) and are transcribed into short, 84-113 nt RNAs which fold back on themselves to form distinct hairpin-containing structures (Figure 1.1). These structures largely reflect the minimal-free energy calculated predictions, but important structural features have also been determined in different studies via both enzymatic cleavage and nucleotide chemical modification experiments (Wolin & Steitz, 1984; van Gelder *et al.*, 1994). Different RNases were used to differentiate between single stranded and double stranded (such as RNase VI) regions (van Gelder *et al.*, 1994). Dimethlysulfate, CMCT and kethoxal treatments were used to chemically modify nucleotides in order to determine the exact positions of the bases, using primer extension assays to detect the modified nucleotides (van Gelder *et al.*, 1994). As a general trend, the Y RNA secondary structures have a minimum of 2 stems each. Another structural characteristic is a large single stranded pyrimidine-rich loop which separates the stems (Wolin & Steitz, 1983). This central loop is present in all studied Y RNAs and shows high structural stability despite being the region with the least sequence conservation (Teunissen *et al.*, 2000). The region with the highest sequence conservation between the Y RNAs is the main stem at the 5'/3' end. Y RNAs have a 3' poly (U) tail and a 5' triphosphorylated guanine residue (pppG), except for Y5 which has pppA (Hendrick *et al.*, 1981).

Each of the four Y RNA genes are independently expressed, although, they tend to be clustered within the genome. For instance, in humans, the four Y RNA genes (hY1, hY3, hY4 and hY5) are all located on chromosome 7 (Maraia *et al.*, 1994b), with Y1 being expressed from the opposite strand (Mosig *et al.*, 2007).

Unusually for pol III transcripts, the Y RNA genes are found as single copies in the human genome (Wolin & Steitz, 1983).

Although Y RNAs interact with a range of proteins, which will be discussed in detail later, a defining trait of Y RNAs is their ability to bind the 60 kDa protein Ro60 (Hendrick *et al.*, 1981; Lerner *et al.*, 1981; Wolin & Steitz, 1984) and the 47 kDa La protein (Hendrick *et al.*, 1981). The typical Y RNA life cycle begins with the Y RNA, and two key genes (*Ro60* and *La*) being transcribed in the nucleus. The *Ro60* and *La* transcripts are exported to the cytoplasm where they are translated into proteins, which are subsequently imported back into the nucleus. As Y RNA transcription is terminated, the newly synthesised transcript associates with La, followed by Ro60 (Peek *et al.*, 1993). The entire Ro RNP complex is then transported to the cytoplasm. It has been shown however, that Ro60 binding is not always necessary for Y RNAs to be functionally active in some pathways (Langley *et al.*, 2010).

Y RNAs are expressed in all human cells (Pruijn *et al.*, 1993), although the cellular composition of the different Y RNAs varies between cell types. Interestingly, red blood cells only possess hY1 and hY4 and lack hY3 and hY5 (O'Brien & Harley, 1990). As erythrocytes are anucleated cells, this indicates different functions for Y1/Y4 and Y3/Y5. Differential Y RNA expression between human tissue types has also been documented. Y RNAs are present in high levels in heart and brain tissues, and lower levels in the liver (Wolin & Steitz, 1984).

It has been calculated via Northern blot that approximately 59% of the Y RNA population in HeLa cells (immortalised cervical cancer cells) is made up of Y5, 26% of Y4, 12% of Y3, and around 3% is composed of Y1 (Gendron *et al.*,

2001). More recent analysis of Y RNA abundance via qPCR in HeLa cell extract gave a similar result (Christov *et al.*, 2006), although the relative abundance of Y4 RNA was estimated to be around 3-4 fold lower than that reported by Gendron *et al,* with Y1 and Y3 showing comparatively higher expression levels than Y4 RNA. The strikingly high expression levels of hY5 RNA and the relatively modest levels of hY4 RNA reported (Gendron *et al.*, 2001; Christov *et al.*, 2006) is not consistent with a report which showed that the hY5 gene promoter is not as efficient as the hY4 gene promoter (Maraia *et al.*, 1996).

As well as having four actively expressed Y RNAs, the human genome also has around 1,000 pseudogenes derived from these Y RNAs which are distributed over all 23 human chromosomes (Perreault *et al.*, 2005). The number of Y RNA pseudogenes in humans and chimpanzees is much greater than that of rodents. These pseudogenes are capable of retrotransposition via the L1 (Long interspersed nuclear element-1) transposition machinery (Perreault *et al.*, 2005). The biological role of these pseudogenes is unknown.

1.3   Y RNA binding proteins

Y RNAs were first discovered owing to their association with proteins recognised by autoantibodies from immune sera of patients with systemic lupus erythematosus (SLE) (Lerner *et al.*, 1981). Y RNAs form complexes within the cell with either single or multiple proteins, forming Y RNA ribonucleoproteins (Y RNPs). When Ro60 is bound to Y RNAs this complex may also be referred to as the Ro ribonucleoprotein complex (Ro RNP), and the two terms are often

interchangeable. The most researched Y RNA binding proteins to date are the 60 kDa protein Ro60 and the 47 kDa La.

1.3.1   General anatomy of the Y RNA Ribonucleoprotein (Y RNP)

Almost as soon as Y RNAs were discovered, it was demonstrated that two key proteins targeted by autoimmune patients, Ro60 and La, both could bind to Y RNAs within the cell (Figure 1.2) (Hendrick *et al.*, 1981). Both Ro60 and La have an RNA binding motif of around 80 amino acids (Pruijn *et al.*, 1991). Reconstitution experiments demonstrated that Ro particles also possessed the La determinant and that RNAs which bind Ro60 and La were generated by RNA polymerase III (Hendrick *et al.*, 1981). In this study, it was also shown that the La protein interacts with all pol III transcripts following transcription. The interaction of La with pol III transcripts was later shown to be due to binding to the 3' polyuridylate stretch on these nascent transcripts (Stefano, 1984; Pruijn *et al.*, 1991; Slobbe *et al.*, 1992). La also binds to the 3' polyuridine tail of Y RNAs (Figure 1.2) (Pruijn *et al.*, 1991), which unlike most pol III transcripts, is retained following maturation. This allows the retention of La by the Y RNA, whereas most pol III transcripts lose this tail during maturation (Mamula *et al.*, 1989; Boire & Craft, 1990). Preliminary studies indicated that La RNPs were primarily localised to the nucleus and that Ro particles were mainly cytoplasmic (Hendrick *et al.*, 1981). Each mammalian cell was found to have around 10^5 copies of the Ro RNP, a level which equates to approximately to 1% of the number of ribosomes (Hendrick *et al.*, 1981).
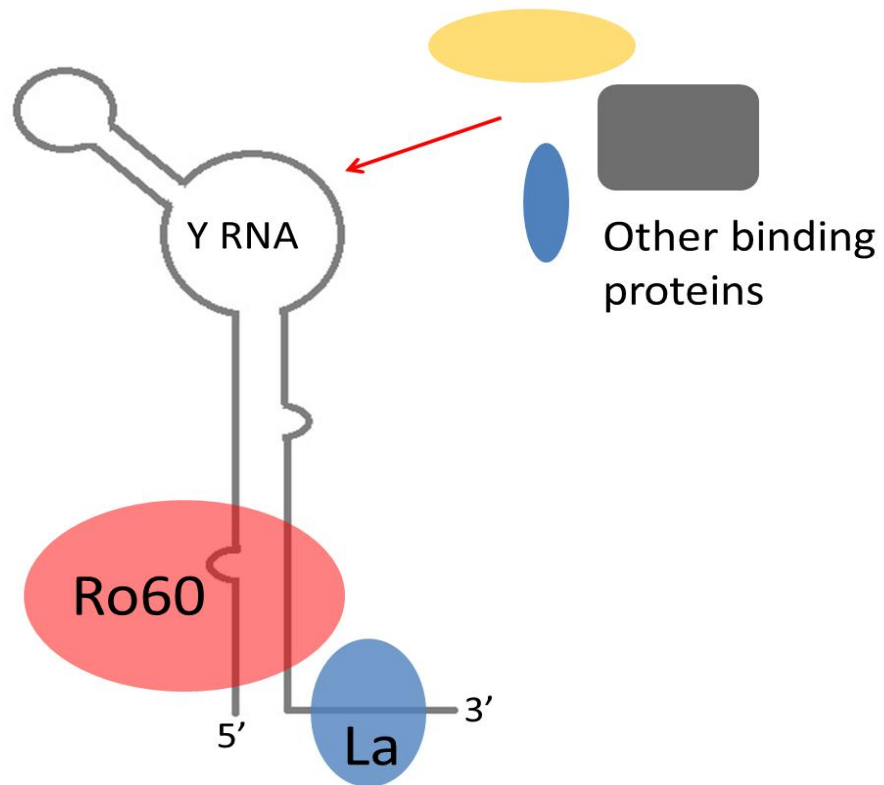
Figure 1.2. Molecular depiction of the Y RNP showing typical Y RNA-protein interactions.

The Ro60 protein, which binds to the most conserved region (the 3'/5' stem) on the Y RNA (Wolin & Steitz, 1984) is a major component of the complex. The bulged cytidine residue essential for Ro60 binding is shown as a kink in the lower 5' stem. Another key protein of the Y RNP, the 47 kDa La, binds to the 3'-poly(U) tail (Pruijn *et al*., 1991). Other reported binding proteins include Nucleolin, hnRNP I & K and L5 protein which primarily bind to the large internal loop of the Y RNA (Fabini *et al*., 2000).

The Ro60 binding site on the Y RNAs is the lower double stranded stem region where the 3' and 5' ends of the RNA meet (Figure 1.2) (Wolin & Steitz, 1984). This stem is characterised by a bulged cytidine residue which is essential for Ro60 binding (Wolin & Steitz, 1984; Pruijn *et al.*, 1991). The binding site was determined via RNase protection assays followed by immunoprecipitation of digested particles using anti-Ro60 antibodies. However, it was noted that the RNase protection experiments would not necessarily reveal other Ro60 binding sites which are easily released following RNase treatment, as very high concentrations of pancreatic ribonulcease were used (Wolin & Steitz, 1984). Deletion mutant experiments later showed the bulged C-residue to be essential for Ro60 binding (Pruijn *et al.*, 1991). An intact Ro60 binding site on the Y RNAs is required for nuclear export of the Y RNAs (Simons *et al.*, 1996).

The size of the Ro RNPs was initially estimated to be between 93-150 kDa (Wolin & Steitz, 1984) but was later found to be between either 230-350 kDa (Boire & Craft, 1990) or 150-550 kDa (Fabini *et al.*, 2000). In all studies, it was acknowledged that this size was greater than the sum of the molecular masses of Ro60, La and a 33 kDa RNA molecule, hinting that other molecular determinants were present in Y RNPs. The affinity of Y RNAs for Ro60 is disrupted at a relatively low ionic strength using 500 mM salt, which is similar for most nucleic acid binding proteins (Boire & Craft, 1990). La binding is slightly more resilient and dissociates after exposure to around 700 mM salt (Boire & Craft, 1990). Y RNPs also elute at different salt concentrations during anion exchange experiments, with hY5 RNP being eluted at lower salt concentrations of around 200 mM NaCl

(Boire *et al.*, 1995) and hY RNPs 1-4 being eluted at higher salt concentrations nearer 300 mM (Boire & Craft, 1990; Boire *et al.*, 1995).

The heterogeneity of Y RNPs is further demonstrated by the presence of autoantibodies in sera from lupus patients which specifically recognised hY5 RNPs (Boire & Craft, 1989). Indeed, further experimentation identified the presence of antibodies in these sera which targeted the RNA component of Y5 RNPs rather than the bound proteins (Boulanger *et al.*, 1995). Immunoprecipitation experiments of *in vitro* transcribed hY5 showed that the region next to the Ro60 binding site on the 5'/3' end of the RNA was the region possessing the epitope recognised by IgG α-hY5 antibodies. Interestingly, one study has shown that hY5 RNP has been found to be primarily nuclear compared with the exclusively cytoplasmic hY RNAs 1-4, and that the signal for nuclear retention resides in the Y5 RNA itself (Gendron *et al.*, 2001).

The literature is divided on the percentage of Y RNAs that are complexed with the Ro protein. The first demonstration that most Y RNAs are bound by Ro60 was demonstrated in mouse cells (Wolin & Steitz, 1983). Extracts from mouse cells were immunoprecipitated with anti-Ro antibodies and it was found that all of the mY RNAs were depleted from the supernatant. Later, this was repeated in human cells (Peek *et al.*, 1993), which gave similar results supporting that most Y RNAs are stably associated with Ro60. However, recent experiments which used qPCR to detect remaining Y RNAs following RoRNP immunoprecipitation have argued that 50% of the Y RNAs are not bound by either Ro60 or La (Langley *et al.*, 2010). This discrepancy might be due to the varying specificity of the antibodies used by these groups and the differential sensitivity of the techniques employed

(Northern blot versus qPCR) to test the amount of remaining Y RNA following immunoprecipitation.

A different investigation found that all Y RNAs are bound by La *in vivo*, but not necessarily all Y RNAs are bound by Ro60 (Fabini *et al.*, 2000). Furthermore, coprecipitation of recent, novel Y RNA binding proteins was more efficient with α-La antibodies than with α-Ro60 antibodies (Fabini *et al.*, 2000). On the other hand, the level of endogenous Y RNAs is dependent on the presence of the Ro60 protein. For example, there is a significant decrease in the level of the single Y RNA present in *Caenorhabditis elegans* (the 105nt long ceY RNA) in ROP-1 (the Ro60 homologue) mutants (Labbe *et al.*, 1999). This decrease in ceY RNA levels is rescued by transgenic expression of ROP-1. Indeed, the levels of mY1 and mY3 are significantly diminished in Ro60$^{-/-}$ mouse embryonic stem cells (Chen *et al.*, 2003).

Although Ro60 may bind a significant proportion of the Y RNAs in the cell, the Moloney murine leukemia virus (MLV) has been shown to selectively encapsidate host mY1 and mY3 RNAs, even in a Ro60$^{-/-}$ background (Garcia *et al.*, 2009). The function of this selective uptake of Y RNAs is unknown, but implies that Ro60 is not necessarily required for Y RNA function, which was further confirmed in chromosomal DNA-replication studies (Langley *et al.*, 2010), discussed in detail later.

Subcellular localization experiments involving both antibody staining and enucleation analysis were used to determine Y RNP and Y RNA distribution (Peek *et al.*, 1993). Antibody based immunofluorescence gave conflicting results depending on the method of fixation used. When either Hep-2 or HeLa cells were

fixed using methanol/acetone, it was shown that monoclonal Ro60 and La antibody staining was speckled in the nucleus but was stronger in the cytoplasm. Conversely, paraformaldehyde fixation gave the opposite result (Peek *et al.*, 1993). However, nuclear and cytoplasmic fractioning experiments resolved this issue, and revealed that the Y RNAs are specifically localised to the cytoplasm, whereas the La and Ro60 proteins are present in both the nucleus and the cytoplasm (Peek *et al.*, 1993). In the nucleus, neither Ro60 or La was found coupled to Y RNAs, whereas most Y RNAs were found protein bound in the cytoplasm. Approximately 30-50 % of free, non-RNA bound Ro60 and La was found to be in the nucleus depending on the cell line analysed (Peek *et al.*, 1993).

Microinjection experiments in *Xenopus laevis* using either *in vitro* transcribed Y RNAs (denoted xY3, xY4, xY5 and xYα) or *in vitro* translated Ro60 and La showed that Y RNAs accumulate in the cytoplasm (Simons *et al.*, 1994). Most of the microinjected La went to the nucleus and a substantial amount of Ro60 was also imported to the nucleus, where both proteins were shown to associate with the Y RNAs (Simons *et al.*, 1994).

1.3.2   Ro60

Ro60 is an important autoantigen in SLE and SS (Sjögren's Syndrome) patients (Ohlsson *et al.*, 2002) and is evolutionarily conserved (Deutscher *et al.*, 1988; Slobbe *et al.*, 1991). Mice lacking Ro60 develop symptoms characteristic of autoimmune disorders, including photosensitivity and skin lesions (Xue *et al.*,

2003). Ro60 is essential for the stability of Y RNAs, as there is a thirty-fold reduction in mY1 and mY3 in Ro60 knockout cells (Xue *et al.*, 2003).

The first identifiable function of Ro60 was its involvement in non-coding RNA quality control (O'Brien & Wolin, 1994; Shi *et al.*, 1996). It was shown that in *Xenopus laevis*, 5S rRNA variants which had 8-10 nt additions at their 3' ends were bound by Ro60. These extensions are a result of RNA polymerase III failing to recognise the first of three transcription termination signals. Variant 5S rRNAs fold differently compared to their wild-type counterparts, forming an alternative helix. It is this specific, altered helix which allows Ro60 to recognise and bind to variant 5S rRNA (Shi *et al.*, 1996). As well as having extensions, variant 5S rRNAs have internal mutations (Shi *et al.*, 1996). These isoforms are inefficiently processed and are eventually degraded. When mutant and wild-type 5S rRNAs are injected into *Xenopus* oocytes, Ro60 specifically associates with the mutant variety (O'Brien & Wolin, 1994). It was later shown that Ro60 binds several RNAs with single stranded 3' tails (which are normally removed during maturation) which include pre-tRNAs and U6 snRNA (Fuchs *et al.*, 2006) as well as misfolded U2 snRNAs (Chen *et al.*, 2003). Although the process of how mutant 5S rRNA associated with Ro60 are eventually degraded is not clear, it appears to be an important quality control pathway. ROP-1 (the Ro60 homologue) mutants in *Caenorhabditis elegans* display a five-fold increase in the level of variant 5S rRNA compared to wild-type strains (Labbe *et al.*, 1999).

Ro60 also plays a role in cell survival following high ultraviolet light (UV) exposure. Solar ultraviolet radiation overexposure can be extremely damaging to cells, and has severe health effects on target organs including the eyes, skin and

immune system (Lucas *et al.*, 2006). In some cases UV damage results in the development of terminal melanoma (Narayanan *et al.*, 2010).

It has been demonstrated that the prokaryote *Deinococcus radiodurans*, which shows high resilience during contact with DNA mutagens, has an ortholog of Ro60 called Rsr (Ro sixty-related) (White *et al.*, 1999). Furthermore, it has been shown that when this bacterium is exposed to damaging UV irradiation, Rsr is upregulated in the cell, and binds to a Y RNA-like molecule which also shows a corresponding increase in expression (Chen *et al.*, 2000). Bacterial strains lacking the Rsr gene show decreased survival rates following UV irradiation than their wild-type counterparts. Another study demonstrated that mouse embryonic stem (ES) cells with a Ro60$^{-/-}$ genotype also show higher mortality rates following UV exposure compared to wild-type ES cells (Chen *et al.*, 2003). It was found that mouse Ro60 and Y RNAs accumulate in the nucleus following UV treatment. Taken together, these findings imply an important role for Y RNAs and the associated Ro60 in survival following UV damage. This role might involve entry into a DNA damage response pathway, or preventing newly transcribed RNAs from misfolding or crosslinking with nuclear proteins during UV exposure (Chen *et al.*, 2003). Indeed, the dynamic subcellular distribution of Ro60 during cellular stress (for example oxidative stress or UV irradiation) is thought to be modulated by Y RNA binding (Sim *et al.*, 2009), which in turn makes Y RNAs key molecules in the first line of response to damaging stress stimuli.

X-ray crystallography studies revealed the structure of Ro60 (Stein *et al.*, 2005). Ro60 is a toroid shaped protein with two domains; a Willebrand factor A domain and an RNA binding α-helical HEAT-repeat ring domain. The inner ring of

the HEAT-repeat domain is positively charged and is 10-15 Å across, and binds to the single stranded 3' ends of misfolded RNAs. The outer ring binds the double stranded RNA portion of the misfolded RNA, but also binds Y RNAs (Stein *et al.*, 2005). The outer binding site for Y RNAs and the duplex portion of misfolded RNAs overlap meaning that there is a competition for the same binding site. This suggests that Y RNAs regulate access of misfolded RNAs to Ro60. The RNA binding surface is extensive and Ro60 was shown to scavenge for not only misfolded, but also correctly folded non-coding RNAs which failed to complex with appropriate binding proteins (Fuchs *et al.*, 2006).

As well as regulating Ro60's interaction with other small cytoplasmic RNAs, Y RNAs also control the subcellular localisation of Ro60 by governing its access to the nucleus (Sim *et al.*, 2009). Ro60 mutants which are unable to bind to Y RNAs are shuttled to the nucleus, as disassociation with Y RNAs reveals a nuclear localisation signal on the surface of the protein (Sim *et al.*, 2009). In the nuclei, Ro60 can then access misfolded RNAs (O'Brien & Wolin, 1994; Chen *et al.*, 2003). Interestingly, it was also found that cellular stresses such as UV light exposure or oxidative stress (such as hydrogen peroxide exposure) result in nuclear accumulation of Ro60 in mouse cells (Sim *et al.*, 2009). Therefore, a key function of the Y RNAs is making Ro60 available for RNA quality control following events which induce cellular damage.

In the eubacterium *Deinococcus radiodurans*, the Ro60 ortholog Rsr shows increased expression following nutrient starvation (Wurtmann & Wolin, 2010). Furthermore, the exoribonuclease polynucleotide phosphorylase (PNP) associates with Rsr during cellular stress in order to mediate rRNA decay (Wurtmann &

Wolin, 2010). However neither an increase in Ro60 levels during cellular stress, nor an association with exoribonucleases have been reported in eukaryotes. Indeed, the most recent tandem affinity purifications of Ro60 failed to identify any ribonuclease association (Sim *et al.*, 2012).

1.3.3   The La protein

The La autoantigen is a very abundant, primarily nuclear, protein which is highly evolutionarily conserved, and is composed of 408 amino acids (Hendrick *et al.*, 1981; Chambers *et al.*, 1988; Scherly *et al.*, 1993; Topfer *et al.*, 1993; Bai *et al.*, 1994). The La protein is around 50-fold more abundant then Ro60 (Peek *et al.*, 1993), and binds RNAs using its N-terminal domain (Pruijn *et al.*, 1991). This binding is regulated through phosphorylation of the carboxy-terminal domain which decreases the affinity of La for the 3' polyuridine tail (Francoeur *et al.*, 1985; Pfeifle *et al.*, 1987). La binds to all RNA polymerase III transcripts, including pre-tRNAs, pre-5S RNA (Rinke & Steitz, 1982) and also some viral encoded RNAs such as the Epstein-Barr virus-encoded ribonucleic acids EBER 1 and EBER 2 and the adenovirus-associated ribonucleic acids VAI and VAII (Rosa *et al.*, 1981). La is thought to function by releasing nascent pol III RNAs from the DNA template allowing multiple rounds of transcription to take place (Gottlieb & Steitz, 1989; Maraia *et al.*, 1994a) .

The La binding site favours the nuclear retention of Y RNAs, as demonstrated by the creation of hY1 deletion mutants, and their injection into Xenopus oocytes (Simons *et al.*, 1996). Here, La binding to hY1 was shown to

slow down export of the Y RNA to the cytoplasm by approximately six hours. Human Y RNAs remain bound to La once in the cytoplasm (Boire & Craft, 1990; Peek *et al.*, 1993), however in *Xenopus* association of La is lost following nuclear export, perhaps due to 3' end shortening which is known to occur in pre-5S rRNA (Simons *et al.*, 1994; Simons *et al.*, 1996). Y RNA binding to La also appears to be important for Y RNA stabilisation. Experiments in *C. elegans* have shown that RNAi depletion of La results in a gradual decrease in Y RNA levels over a 72 hour period (Aftab *et al.*, 2012). As mentioned, Ro60 binding has also been shown to be important for Y RNA stability (Chen *et al.*, 2000; Xue *et al.*, 2003). This makes Y RNA retention by these two proteins in the cytoplasm highly likely.

La has had many implicated biological functions, many of which are controversial or lack significant functional evidence. For example, it was proposed that La acts as a transcription termination factor and facilitates the re-initiation of pol III transcription (Gottlieb & Steitz, 1989; Maraia *et al.*, 1994a). However, in yeast and vertebrate systems where La was depleted, pol III function was unaffected (Yoo & Wolin, 1994; Lin-Marq & Clarkson, 1998). A more established role for La is in protection of pol III transcripts from exonuclease digestion prior to maturation, after which La is displaced (Kufel *et al.*, 2000; Xue *et al.*, 2000). La appears to act as a molecular chaperone for nascent RNAs as it only binds targets temporarily and assists correct assembly and fate of these immature transcripts (Xue *et al.*, 2000).

The La protein also has a role in pre-tRNA maturation, specifically in processing the 5' end of pre-tRNAs (Fan *et al.*, 1998). Indeed, this is a further example of the protein's RNA chaperone activity. La stabilises pre-tRNAs and

arranges the RNA in a specific structure which allows endonucleolytic cleavage and ultimately, 3' end maturation (Yoo & Wolin, 1997). During tRNA maturation, 5' end processing precedes 3' end processing. The N-terminal region of La binds to the 3' poly(U) of the pre-tRNA whilst the C-terminal end of the protein interacts with the 5' tail. When La is phosphorylated at the C-terminus, binding to the 5' end of the pre-tRNA is inhibited, allowing processing by the ribonuclease RNase P, whilst the 3' end remains protected (Fan *et al.*, 1998). Indeed, La has been shown to have a phosphorylation-dephosphorylation cycle. During apoptosis, the La protein is completely dephosphorylated and around 25% of La proteins are cleaved (Rutjes *et al.*, 1999b). The proposed transcriptional activity of La is inhibited by phosphorylation (Maraia *et al.*, 1994a). Dephosphorylated La (which is active in pol III transcription) inhibits 5'-end processing of pre-tRNAs (Fan *et al.*, 1998). This phosphorylation cycle is disrupted during apoptosis (Rutjes *et al.*, 1999b), and therefore deactivation of La may serve as a key step facilitating cell death.

La has controversially been implicated in regulating translation of both viral and cellular mRNAs. Various studies have used crosslinking analysis and found that La binds to the 5' untranslated region of the poliovirus RNA (Meerovitch *et al.*, 1993) and also to mRNA encoding the endogenous X-linked inhibitor of apoptosis protein (XIAP) (Holcik & Korneluk, 2000). In the latter study, La was shown to bind to the internal ribosome entry site (IRES) and facilitate translation, which was inhibited when the availability of functional La was reduced. However the role of La in translation has been disputed, owing to the extremely high concentrations of La required to re-initiate translation in several La

immunodepletion experiments, and the lack of functional evidence of its involvement in this process (Wolin & Cedervall, 2002).

Finally, the La protein appears to play a role in the gene silencing pathway of animal cells. The RNA interference (RNAi) pathway is governed by the RNA-induced silencing complex (RISC). RISC is composed of a small RNA complementary to an mRNA target, as well as an effector protein called Argonaute which causes target cleavage or translational silencing (Bartel, 2004). RISC turnover, or the rate at which mRNA targets are cleaved and released, has been shown to be dependent on the La protein (Liu *et al.*, 2011). Specifically, La promoted the multiple turnover of RISC via association with the Argonaute protein, AGO2, and by facilitating the release of cleaved mRNA targets. La associates with AGO2 in an RNA-dependant manner (Liu *et al.*, 2011).

1.3.4   Ro52

As well as Ro60 and La, another protein which showed reactivity with autoimmune sera is Ro52. Early investigations claimed that Ro52 was part of the Y RNP complex, binding to the Y RNA via Ro60 (Ben-Chetrit *et al.*, 1988; Slobbe *et al.*, 1991; Slobbe *et al.*, 1992). However, additional fractionation and co-immunoprecipitation experiments demonstrated that Ro52 does not bind to any of the Y RNAs, Ro60 or La (Kelekar *et al.*, 1994; Boire *et al.*, 1995). Ro60 and Ro52 also show differential patterns of cellular localisation; Ro60 was shown to be predominantly cytoplasmic whereas Ro52 showed mainly nuclear staining (Kelekar *et al.*, 1994). Unlike Ro60 and La, Ro52 does not have an RNA

recognition motif (RRM) (Kelekar *et al.*, 1994). Furthermore, it has been demonstrated that owing to the zinc finger and leucine zipper motifs of Ro52, this protein binds DNA rather than RNA targets (Frank *et al.*, 1995).

1.3.5   hnRNP I & K

Most of the additional Y RNA binding proteins discovered since Ro60 and La have been shown to interact via the pyrimidine-rich central loops of these molecules (Figure 1.2). Using *in vitro* transcribed biotinylated Y RNAs in reconstitution experiments with various HeLa cell S100 extracts, several novel Y RNA binding proteins between 53 kDa and 80 kDa were identified which bound preferentially to Y1 and Y3 (Fabini *et al.*, 2000). The identity of two of these proteins was later identified via mass spectrometry to be the heterogeneous nuclear ribonucleoproteins, hnRNP I & K, which bind to a subset of the human Y RNAs via the large central loop, as demonstrated by deletion mutant experiments (Fabini *et al.*, 2001). The hnRNP I protein was shown to be a stable component of Y RNPs 1 and 3. However, hnRNP K appeared to bind only transiently to hY1 and hY3 which were devoid of Ro60 and La (Fabini *et al.*, 2001). The hnRNP I protein was shown to bind a minor fraction of the hY1 and hY3 population (10-20%), and required the presence of La, which presumably acts as a molecular chaperone to assist RNA folding and subsequent recognition. It was noted that both hnRNP I and La have been implicated in similar mRNA processing events, such as translation, and suggested that the Y RNAs may serve as carriers for these molecules (Fabini *et al.*, 2001).

1.3.6   RoBPI

The Ro RNP binding protein (RoBPI) was discovered as a Y RNP component via a ribonucleoprotein interaction assay based on the yeast-three hybrid system (Bouffard *et al.*, 2000). Using recombinant Ro60 and Y RNAs, it was shown that RoBPI only interacted with Ro60 and hY5 together, and not with any one of these components alone. The 559 amino acid RoBPI is approximately 65 kDa, has three RNA recognition motifs (RRM) and displays nuclear colocalisation with Ro60, La and hY5 (Bouffard *et al.*, 2000). The RoBPI protein is structurally similar to mRNA splicing factors, and therefore like La and hnRNP I & K, appears to be involved in controlling expression of coding and non-coding genes. As the hY5 RNP has been shown to be predominantly nuclear in humans and mouse compared with the cytoplasmic hY RNPs 1-4 (Gendron *et al.*, 2001), is targeted by distinct autoantibodies in autoimmune patients (Boulanger *et al.*, 1995), as well as eluting in a separate fraction to hY RNPs 1-4 during anion exchange chromatography (Boire *et al.*, 1995), RoBPI may contribute to this molecular uniqueness of hY5 which separates it from the other Y RNPs.

1.3.7   Nucleolin

Nucleolin is a protein which is localised to the nucleolus and has been shown to be involved in multiple steps of the ribosome biogenesis pathway (Ginisty *et al.*, 1999). Immunopreciptiation of Ro60 and La autoantigens followed by mass spectrometry of associated proteins revealed that nucleolin also appears to

be complexed to hY1 and hY3 RNPs (Fouraux *et al.*, 2002). Both matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and quadruple time-of-flight (Q-TOF) mass spectrometry were used to confirm that the novel binding protein was indeed nucleolin. Again, like several other novel Y RNA related proteins, nucleolin was shown to bind the pyrimidine-rich internal loop of these Y RNAs, and it can do so in concert with hnRNP I. The structure, and not the sequence of the central loop on the Y RNAs, was shown to be the key determinant of nucleolin binding (Fouraux *et al.*, 2002). As nucleolin has been shown to shuttle between the nucleus and cytoplasm of cells (Borer *et al.*, 1989), it was proposed that nucleolin acts as a carrier molecule of hY RNA1 & 3 (Fouraux *et al.*, 2002).

1.3.8   L5

In depth analysis about Ro60's involvement in RNA quality control revealed an RNP-RNP interaction between Y5 RNA-Ro60 and 5S rRNA bound to the ribosomal protein L5 (Hogg & Collins, 2007). Using a novel RNP purification method it was shown that Y5 RNA specifically targets Ro60 to misfolded 5S rRNAs by interacting with the L5 protein via its internal pyrimidine rich loop, which in turn is bound to ribosomal RNA. Y5 RNA then acts as a 'gatekeeper' for the surface of Ro60, helping to discriminate between correctly and incorrectly folded substrates (Hogg & Collins, 2007).

The most recently identified, novel Y RNA binding protein is the zipcode-binding protein ZBP1 (Sim *et al.*, 2012). The 68 kDa ZBP1 protein has been shown to be involved in mRNA localisation and translation regulation (Ross *et al.*, 1997; Huttelmaier *et al.*, 2005). ZBP1 was shown to copurify with Ro60 and Y3 RNA in mouse cells, and this association is increased in cells exposed to UV light. Nuclear accumulation of the mY3/Ro60/ZBP1 complex was also more pronounced following UV light exposure (Sim *et al.*, 2012). Under non-stressed conditions, Ro60 and mY3 were shown to be localised to the cytoplasm by ZBP1, and the Y3 RNP complex accumulated in the nucleus when ZBP1 was depleted (Sim *et al.*, 2012).

1.4   The role of Y RNAs in chromosomal DNA replication

The propagation of life relies on the efficient and timely replication of the cell's genetic material. During cell division in all organisms, parental DNA gives rise to two identical daughter strands, each inheriting one original strand from the parent in a semi-conservative fashion. In eukaryotes, chromosomal DNA replication is tightly regulated and is closely linked to the cell cycle, occurring in S phase.

The current model of DNA replication is already well reviewed in the literature (Takeda & Dutta, 2005; Pospiech *et al.*, 2010), but I will briefly summarise key molecular events here in order to facilitate the understanding of the

role of Y RNAs in this process. Eukaryotic cells start DNA replication via the assembly of the multiprotein pre-replicative complex (pre-RC) in G_1 phase of the cell cycle. This is initiated with the assembly of the six subunit origin recognition complex (ORC), which in turn recognises specific sites where DNA replication initiation occurs - called autonomously replicating sequences (ARS). The Cdc6 and Cdt1 proteins then interact with the ORC and facilitate the recruitment of the minichromosome maintenance proteins Mcm2-7. Together these proteins form the pre-RC in G_1 phase nuclei. Activation of the pre-RC and the formation of DNA replication forks occurs during the transition to S phase and involves a second group of factors. These proteins include the two kinases Cdk2 and Cdc7. Other initiation factors are then recruited, including Cdc45, Mcm10, the GINS complex and replication protein A (RPA). Origin DNA is then unwound and the three DNA polymerases (DNA pol α, DNA pol δ and DNA pol ε) are recruited and eventually commence replication.

Early studies which investigated the biochemistry of DNA replication employed cell fusion experiments, such as those famously conducted by Rao and Johnson in 1970 (Rao & Johnson, 1970). Using this technique, it was shown that S phase cells contain factors which could induce G1, but not G2 phase nuclei, to prematurely trigger DNA replication. More recently, the development and optimisation of a human cell-free system to study DNA replication was established which has resulted in further replication factors being uncovered (Krude *et al.*, 1997; Krude, 2006). In this cell-free system, isolated cell cycle-arrested template nuclei synchronised in late G1 phase are prepared which are licensed to replicate, but lack active DNA replication forks. Then, by adding cytosolic extracts from

proliferating cells containing initiation factors or combinations of specific cellular fractions, novel DNA replication factors can be identified by seeing if arrested nuclei can then initiate DNA replication (Krude, 2006).

Using this cell-free system Krude and colleagues identified the noncoding Y RNAs as essential factors for chromosomal DNA replication (Christov *et al.*, 2006). When the cytosolic extract from actively proliferating cells was fractionated by anion exchange chromatography, this gave rise to two essential fractions called QA and QB. QB was further separated into fractions ArFT and ArE. Fractions QA, ArFT and ArE together were sufficient to initiate DNA replication when added to template nuclei. Surprisingly, fraction ArE was found to be composed mainly of nucleic acids rather than proteins, and it was from this fraction that Y RNAs were purified. The QA and ArFT fractions alone were not sufficient to significantly drive DNA replication in isolated nuclei from human somatic cells, with only 15 % of nuclei replicating in the presence of these two factors (Christov *et al.*, 2006). However when the purified Y RNAs were added along with these two protein fractions, the proportion of replicating nuclei increased in a dose-dependent manner to 50 % (Christov *et al.*, 2006). Furthermore, targeted depletion of Y RNAs from the cytosolic extract inhibited DNA replication. A degree of functional redundancy with regards to the Y RNAs was found in this system as addition of any of the four Y RNAs to the depleted extract was sufficient to reinstate DNA replication. This requirement for Y RNAs for DNA replication was also seen in mouse cell nuclei, where interestingly, human Y RNAs could also initiate mouse DNA replication indicating that these factors are evolutionarily conserved. Non-

vertebrate Y RNAs could not reconstitute DNA replication in vertebrate systems (Gardiner *et al.*, 2009).

Importantly, the sequence and structure of the Y RNAs were shown to be the contributing factors which initiated DNA replication (Christov *et al.*, 2006). A conserved double stranded 9 base pair motif sequence in the upper stem of the Y RNAs identified in screens of hY1 mutants was sufficient to drive DNA replication in the presence of protein fractions QA and ArFT (Gardiner *et al.*, 2009). This motif was found to have a consensus sequence of 5'-GUAGUGGG-3' on the 5' strand and 5'-CCCACUGCU-3' on the 3' strand. This motif was not found in non-vertebrate Y RNAs. The addition of this double stranded sequence alone was sufficient to drive DNA replication without the requirement of the other protein binding domains of the Y RNA (again, in the presence of QA and ArFT). Interestingly, the same results were not observed with the addition of this sequence in a DNA from.

Although Y RNAs can be complexed with cellular proteins such as Ro60, La and nucleolin, none of the associated Y RNA binding proteins are required for DNA replication, and addition of any of these proteins does not affect the percentage of actively replicating nuclei in the cell-free system (Langley *et al.*, 2010). It was found via real-time PCR that around 50 % of cellular Y RNAs were not bound by Ro60/La/Nucleolin, and it was argued that it is the non-protein bound portion of Y RNAs which have a role in DNA replication. This would leave the Ro60/La/Nucleolin bound Y RNAs to carry out distinct functions such as scavenging for misfolded RNAs.

In order to investigate exactly how Y RNAs facilitate chromosomal DNA replication, Krude and colleagues degraded Y1 and Y3 RNA in a cell-free system and monitored replication track extension rate (or the rate at which nascent DNA is polymerised) (Krude *et al.*, 2009). They found that Y RNAs are required for the establishment and initiation of DNA replication forks, but are not required for the elongation of actively replicating DNA. Additionally, the stability of the DNA replication fork was not affected by depleting the Y RNA population. The reduction in the amount of single stranded nascent DNA by degrading Y3 RNA could be negated by the addition of Y1 RNA – further supporting the functional redundancy of Y RNAs in this process. In another study, using fluorescently labelled Y RNAs, it was shown that Y RNAs act in a 'catch and release' mechanism whereby Y RNAs associate with the unreplicated euchromatin in the late G1 phase of the cell cycle, and are displaced once initiation has taken place (Zhang *et al.*, 2011). Around 20-70% of hY RNAs associate with nuclei in G_1 phase cells whereas only 4-10 % associate in G_2. It was further shown that hY1, hY3 and hY4 co-localise with each other on euchromatin sites, but not with hY5 - which was enriched in nucleoli. Although it was previously shown that the upper stem region alone was sufficient to reconstitute DNA replication in template nuclei (Gardiner *et al.*, 2009), it was demonstrated that the loop domain of the Y RNAs was required for targeting Y RNAs specifically to euchromatic sites. Y RNA mutants where only the upper stem maintained a wild-type sequence showed non-specific site binding across the nucleus including heterochromatic regions (Zhang *et al.*, 2011).

Human Y RNAs appear to associate with DNA replication factors. This was demonstrated by conducting RNA pull-down assays where agarose beads are coupled to the 3' termini of the Y RNA (Zhang *et al.*, 2011). All 4 Y RNAs interact with members of the origin recognition complex (ORC 2 & 3 interacted with hY1, 3, 4 & 5 and ORC 4 & 6 interacted with hY1, 3 & 5 only). Furthermore, Cdt1, Cdc6, and the DNA unwinding protein DUE-B all interacted with the Y RNAs. None of the proteins associated with DNA replication elongation such as Mcm2-7, CMG and the DNA polymerases bound to the Y RNAs supporting earlier work that these non-coding RNAs are specifically involved in the initiation stage of replication (Krude *et al.*, 2009; Zhang *et al.*, 2011).

Finally, the most recent work with regards to Y RNAs and chromosomal DNA replication showed that the Y RNAs only act as licensing factors after the midblastula transition (MBT) stage of development (Collart *et al.*, 2011) and not before. Using inhibitory morpholino nucleotides, it was demonstrated that Y RNAs are not needed for DNA replication in *Xenopus* and zebrafish embryos and egg extracts prior to MBT. After this stage, Y RNA specific morpholino nucleotide treated embryos died shortly before gastrulation. The lack of requirement for Y RNAs to facilitate in DNA replication prior to MBT may facilitate the need to replicate the cell's DNA rapidly during early developmental stages. After this developmental stage, DNA replication takes significantly longer - such as in post-MBT cells.

1.5 Y RNAs in cancer

Cancer is a complex genetic disease which results in uncontrolled cell proliferation and tissue invasion. It can have multiple genetic origins, but essentially it arises as a result of the overexpression or mutation of proto-oncogenes (normal genes which have the ability to cause cancer) and the underexpression of tumour suppressors (Dalmay, 2008). Cancer development is a multistep process that involves six hallmark events: uncontrolled proliferation, apoptosis evasion, tissue invasion, the ability to grow without external biochemical signals, the ability to ignore anti-growth signals and sustained angiogenesis (Hanahan & Weinberg, 2000). Owing to the widespread and intricate role of small RNAs (sRNAs) in cellular processes, aberrant sRNA expression may promote cancer progression. The profiles of microRNAs (miRNAs), which are one type of sRNA, are significantly altered in breast cancer cells compared with normal breast tissue (Iorio *et al.*, 2005). Understanding how sRNA profiles influence disease processes will aid the diagnosis and treatment of a variety of disorders. For example, as some cancer cells have distinct miRNA profiles, rapid diagnosis of disease type can be achieved via microarray or quantitative PCR (qPCR) (Dalmay, 2008). Additionally, miRNA-mimics or miRNA inhibitors can be used to target specific microRNAs in order to prevent certain disorders. For example, the use of locked nucleic acid (LNA) in therapeutic studies has already been demonstrated (Elmen *et al.*, 2008).

As discussed, Y RNAs function in chromosomal DNA replication, implicating them in the control of cell proliferation, a characteristic which is

aberrant and uncontrolled in cancer. Indeed, it has been shown that if HeLa cells are transfected with siRNAs against hY1, there is a significant reduction in cell proliferation and a subsequent decrease in the number of cells in the S phase of the cell cycle (Christov *et al.*, 2006). Furthermore, transfection of siRNAs against hY1 and hY3 into a range of different carcinoma cells (including bladder, prostate, cervical and lung) results in a similar reduction in cell proliferation (Christov *et al.*, 2008). This makes Y RNAs promising targets for future pharmalogical treatments in cancer patients.

A study by Christov *et al.* also demonstrated using real-time PCR that all Y RNAs in humans are significantly upregulated in bladder, cervical, colon, kidney, lung and prostate cancers compared to the equivalent non-diseased tissues (Christov *et al.*, 2008). The greatest level of hY1, hY3 and hY4 overexpression was seen in kidney tumours, and the highest hY5 levels were observed in lung tumours. The authors propose that Y RNAs are novel cancer biomarkers, and could be utilized for diagnosis in the future.

A more recent study has shown that small RNAs derived from full length Y RNAs are present in solid tumours (Meiri *et al.*, 2010). Notably, 25 nt long RNAs derived from hY1, hY3 and hY5 were present in solid tumour samples at levels comparable with the cancer associated miRNA, miRNA-21, however, no microRNA-like function of these Y RNA products was shown (Meiri *et al.*, 2010).

1.6 Y RNAs in apoptosis

Apoptosis is a highly regulated and well-orchestrated mode of cell death which is quite different from uncontrolled cell death which is disorganised and can be damaging to neighbouring cells (Taylor *et al.*, 2008). Apoptosis is carried out by a family of proteases called caspases which initiate a molecular cascade resulting in important structural components of the cell being degraded. These include genomic DNA, the nuclear lamina and mitochondrial components. The entire process is self-contained and occurs without allowing the release of immunostimulatory molecules that may harm other cells. The resulting products of degradation are removed by phagocytic white blood cells. Mutations that impair apoptotic pathways can lead to cancer.

Apoptosis appears to strongly involve Ro RNPs. Firstly, it is well documented that the Ro60, Ro52 and La autoantigens cluster at the surface of apoptotic cells (Casciola-Rosen *et al.*, 1994; Ohlsson *et al.*, 2002). Specifically, Ro52 is found surface-exposed in small apoptotic blebs whereas the larger apoptotic bodies contain cell surface La and also Ro60. Intriguingly, La is dephosphorylated and cleaved at the C-terminus by caspase-3 during apoptosis (Rutjes *et al.*, 1999b). La is known to play a role in preventing apoptosis by activating the translation of the X-linked inhibitor of apoptosis protein (Holcik & Korneluk, 2000).

The RNA component of Ro RNPs also appears to be altered during apoptosis. Human Y RNAs are rapidly and specifically cleaved during apoptosis in a caspase-dependant manner, and this occurs under the duress of a range of

apoptotic stimuli (Rutjes *et al.*, 1999a). The Y RNA fragments produced during apoptosis are approximately 22-25 nts, with a larger fragment of 31 nts also being generated. The 31 nt fragment was still found bound to Ro60 and La, whereas the shorter fragments could only be immunoprecipitated by Ro60 antibodies, suggesting that the La binding site on these fragments was lost (Rutjes *et al.*, 1999a). Interestingly, the degradation of these Y RNAs is one of the first phenotypes of apoptotic cells, with cleavage products being detectable only 1.5 hours after apoptosis induction. This makes Y RNA cleavage detection via qPCR an effective tool in assaying programmed cell death progression in cell samples (Asselbergs & Widmer, 2003). Other RNAs such as tRNAs and snRNAs were found to remain intact in the study by Rutjes *et al.*, implying a specificity for Y RNAs (Rutjes *et al.*, 1999a). Furthermore, this cleavage can be inhibited by apoptosis inhibitor proteins such as Bcl-2. The most conserved region of the Y RNAs, the main stem at the 3'/5' end, remains bound to the autoantigens after cleavage, with Ro60 having the strongest affinity for the degradation products (Rutjes *et al.*, 1999a).

1.7   Small RNA biology

1.7.1   The advent of 'deep sequencing' and the discovery of the 'small RNA world'

The birth of small non-coding RNA biology (sRNAs) (RNAs < 200nt) came with the discovery that small double stranded RNA molecules can bind to complementary mRNA targets and inhibit their expression into proteins (RNA interference – RNAi) (Lee *et al.*, 1993; Fire *et al.*, 1998; Hamilton & Baulcombe, 1999). Since then, the advent of small RNA sequencing techniques based on 3' and 5' adapter ligation and subsequent reverse transcription PCR (RT-PCR) (Elbashir *et al.*, 2001) have allowed 'deep sequencing' of RNA samples. Various 'Next-generation sequencing' (NGS) platforms such as Illumina (Solexa) and 454 pyrosequencing have been developed which have increased the number of sequence 'reads' which can be obtained, to hundreds of millions per sample (Metzker, 2010). The improvement of this technology, the reduction of running costs and increase in availability has resulted in the discovery of millions of novel small RNAs (Wittmann & Jaeck, 2010). This in turn has completely revolutionised our view of how the cell controls the flow of genetic information and the hierarchical structure by which this is established.

1.7.2 MicroRNAs (miRNAs): A small RNA case study

MicroRNAs are some of the most abundant, well studied small RNAs, and have long been the focus of thousands of deep sequencing studies. MicroRNAs (miRNA) are endogenously produced post-transcriptional regulators which are imperfectly complementary to the 3'UTR (untranslated region) of their targets. They are ~22nt in length and form distinct, stem loop structures as a result of imperfect self-complementarity, (Lau *et al.*, 2001). MiRNAs are thought to regulate at least 30%, and maybe as much as half of mammalian protein coding genes (Filipowicz *et al.*, 2008), with each miRNA being able to regulate up to 100 targets (Brennecke *et al.*, 2005). Unlike siRNAs which direct cleavage of their mRNA targets, animal miRNAs are thought to primarily inhibit translation of their semi-complementary transcripts by firstly inhibiting 5' methylguanosine cap-dependant translation initiation and then by deadenylation of the mRNA (Fabian *et al.*, 2009), resulting in degradation by the exosome (Doma & Parker, 2007).

Most microRNAs are transcribed by RNA polymerase II (Lee *et al.*, 2004), and following transcription, the miRNA folds into a dsRNA conformation containing a stem-loop structure at one end. This primary microRNA (pri-miRNA), which can be hundreds or thousands of nucleotides in length, can contain a variety of different miRNA species which will eventually be separated from each other (Filipowicz *et al.*, 2008). In the nucleus, the pri-miRNA is enzymatically cleaved by the 650 kDa Microprocessor complex (Denli *et al.*, 2004) which is composed of the RNase III-like enzyme Drosha, and the RNA binding protein, DGCR8 (or *pasha* in the case of *Drosophila*). The resultant precursor-miRNA (pre-miRNA) is

around 60-80 nt, and is characterised by a 2 nt overhang at its 3' end. The role of the Microprocessor complex is redundant in some miRNA species, where the pre-miRNA forms a debranched intron and is generated during splicing. These 'mirtrons' such as miR-1004, can be transported directly to the cytoplasm (Ruby *et al.*, 2007).

Following nuclear processing, the pre-miRNA is exported into the cytoplasm by the RanGTP dependant exportin-5 located in the nuclear envelope (Bohnsack *et al.*, 2004). In the cytoplasm, another RNase III enzyme Dicer, along with the HIV-1 transactivating response RNA binding protein (TRBP), process the pre-miRNA into a 20-22 nt miRNA/miRNA* duplex (Chendrimada *et al.*, 2005). Although Dicer is involved in the canonical microRNA maturation pathway, studies have demonstrated that a number of miRNAs circumvent the use of Dicer altogether (Cheloufi *et al.*, 2010; Cifuentes *et al.*, 2010). Pre-miR-451 in mouse and zebrafish for example does not require Dicer as part of its biogenesis, but instead requires the slicer activity of Ago2 for its maturation.

The miRNA duplex contains a guide miRNA and a passenger miRNA*. The miRNA is semi-complementary to its mRNA target, and contains a 2-8 nt 'seed' region at its 5' end. The seed region of the miRNA has a near perfect degree of complementarity to the miRNA recognition element (MRE) at the 3' UTR of the mRNA (Brennecke *et al.*, 2005). The passenger miRNA* strand is usually discarded, however recent studies have shown that several species of miRNA*s can be actively selected for RISC incorporation (Ghildiyal *et al.*, 2010). TRBP also functions as a recruiting factor by bringing the DICER-miRNA complex and the

Argonaute protein (AGO) based silencing complex (RISC) together, thereby forming the RISC loading complex (RLC) (Gregory *et al.*, 2005).

Once loaded into RISC, the process of gene silencing can take place. There is still an uncertainty about how exactly miRNA gene silencing is executed in animals, and the prevalence of translational repression and storage in cytoplasmic foci *versus* mRNA transcript degradation (Guo *et al.*, 2010; Djuranovic *et al.*, 2012).

1.8 Small RNAs derived from longer RNAs

1.8.1 Filtering of deep sequencing data reveals extensive RNA fragmentation

Since 2008, there has been an abundance of deep sequencing studies which have identified small RNA fragments derived from longer RNAs, and this has been the subject of several excellent reviews (Wittmann & Jaeck, 2010; Rother & Meister, 2011; Tuck & Tollervey, 2011). Previous deep sequencing studies overlooked these fragments, as they were filtered out of data sets as assumed degradation products. For example, small nucleolar RNAs (snoRNAs), which are involved in ribosome biogenesis, have been shown to give rise to RNA fragments which can act as microRNAs (Ender *et al.*, 2008; Brameier *et al.*, 2011). The first snoRNA-derived RNA (sdRNA) was discovered in HEK293 cells by deep sequencing of immunoprecipitated AGO1 & AGO2 associated RNAs (Ender *et al.*, 2008). This sdRNA was found to be processed by dicer (but independent of Drosha) and its RNAi function demonstrated by luciferase assay. However other sdRNAs

show potential dicer-independent maturation (Brameier *et al.*, 2011). Further sdRNA discovery confirmed that these 18-35nt RNAs primarily mapped to the 5' and 3' stem of the full length snoRNA parent (Brameier *et al.*, 2011).

Indeed, sRNA fragments from rRNAs, snoRNAs, snRNAs and tRNAs seem to give fragments predominantly from the 5' and 3' ends of mature parental RNA transcripts (Li *et al.*, 2012). This seems to occur asymmetrically in human and mouse cells, with one end appearing more prominently in sequencing studies (similar to microRNA/ microRNA* level bias) (Li *et al.*, 2012).

Additional AGO immunoprecipitation and deep sequencing studies uncovered small RNA fragments derived from ribosomal RNAs (rRNA), small nuclear RNAs (snRNAs), vault RNAs (vRNAs), transfer RNAs (tRNAs) and even messenger RNAs (mRNAs), all of which associated with the silencing protein family (Burroughs *et al.*, 2011).

Vault particles are cellular organelles thought to be involved in nuclear export and play a role in multidrug resistance. They contain the 88-98 nt non-coding vault RNAs (vRNAs). It was shown that these vRNAs can be cleaved to give rise to small 23nt vault-derived sRNAs (svRNAs), independent of dicer processing (Persson *et al.*, 2009). Intriguingly, one of these svRNAs associates with an Argonaute protein (part of the RNAi silencing machinery) and can downregulate the mRNA encoding enzyme CYP3A4 which has a role in drug metabolism (Persson *et al.*, 2009).

An analysis of a large collection of publically available, independent deep sequencing datasets has revealed that long non-coding RNAs (lncRNAs) can be processed into sRNA fragments (Jalali *et al.*, 2012). The sRNA clusters

predominantly mapped to the 3' end of lncRNAs. Interestingly, lncRNA-derived sRNA fragments were shown to be differentially processed in a variety of cell and tissue types, implying functionality rather than being simple products of degradation (Jalali *et al.*, 2012).

1.8.2   tRNA fragmentation

To date, the most well studied sRNA fragments studied are those derived from the tRNAs. The first reports of tRNA cleavage came in a study which showed that tRNA cleavage in *Escherichia coli* occurs following infection by bacteriophage T4 (Levitz *et al.*, 1990). A more recent deep sequencing study in an animal system has found that sRNA fragments derived from tRNAs are more abundant than microRNAs (or certainly a close second), and are not simply the product of degradation pathways (Cole *et al.*, 2009; Lee *et al.*, 2009).

There are two distinct classes of tRNA fragments; the tRNA regulatory fragments (tRFs) and the stress induced tRNA-derived RNAs (tiRNAs or 'tRNA halves'). The tRFs are composed of three types of fragments (17-26nt) which were first discovered during deep sequencing of prostate cancer cell lines (Lee *et al.*, 2009). The tRF-1 species is released by RNAse Z (ELAC2) from the 3' trailer of pre-tRNAs as part of tRNA maturation (Lee *et al.*, 2009) and tRF-3 and tRF-5 fragments are generated from mature tRNAs by dicer from the 3' end and 5' end of the mature tRNA respectively (Cole *et al.*, 2009; Lee *et al.*, 2009). RNAi mediated knockdown of one of the tRFs, tRF-1001, resulted in reduced cell proliferation (Lee *et al.*, 2009). Although tRFs have been shown to be dicer dependant and some

tRFs can associate with Argonaute proteins (Cole *et al.*, 2009; Haussecker *et al.*, 2010), the gene silencing capability of tRFs still requires further validation. Although modest gene silencing by tRFs has been observed despite widespread Argonaute association, the mechanism of tRF function is still under debate. It has been argued that rather than acting as *bona fide* microRNAs, tRFs globally modulate gene silencing by competing for AGO binding, thereby regulating AGO's availability to bind with microRNAs (Haussecker *et al.*, 2010). However, the latest research has shown that AGO2 associated 3' tRFs are complementary to retroviral sequences in the human genome and can target these transcripts via RNAi (Li *et al.*, 2012).

The second class of tRNA fragments is the stress induced tRNA-derived RNAs (tiRNAs). These tiRNAs, or 'tRNA halves' are longer than tRFs (31-40 nts) and are generated via cleavage in the anticodon loop of mature tRNAs following various stress stimuli (Thompson *et al.*, 2008; Fu *et al.*, 2009; Yamasaki *et al.*, 2009). Both 5' and 3' end derived tRNA halves can be detected. A variety of stress conditions, including nutrient depletion, hypoxia, hypothermia, heat shock and ultraviolet light irradiation have been shown to yield tiRNAs extremely rapidly (within 20 minutes) (Fu *et al.*, 2009; Yamasaki *et al.*, 2009). A low level of tiRNAs was also detected in unstressed cells (Thompson *et al.*, 2008). Furthermore, this cleavage is not a mechanism to simply reduce the number of available tRNAs as only a small proportion of the full-length tRNA population are targeted (Fu *et al.*, 2009; Yamasaki *et al.*, 2009).

The RNase A family member, angiogenin, was found to be the nuclease responsible for stress induced tRNA cleavage (Fu *et al.*, 2009). Angiogenin is

bound to the ribonuclease inhibitor RNH1 under normal conditions. When angiogenin was overexpressed in cells there was an increase in tiRNA production (Fu *et al.*, 2009) and when RNH1 was targeted via siRNA, a similar result was observed (Yamasaki *et al.*, 2009). It was demonstrated that the function of tiRNAs was to inhibit translation. Specifically, transfection of 5' tiRNAs (but not 3' tiRNAs) inhibited protein synthesis and RNH1 depletion promoted translational arrest (Yamasaki *et al.*, 2009).

Angiogenin-induced tiRNA fragments were shown to specifically target the initiation stage of translation. Protein synthesis is inhibited by tiRNAs association with the translational silencer YB-1, and subsequent displacement of the eukaryotic initiation factor eIF4G/eIF4A (Ivanov *et al.*, 2011). A subset of these tiRNAs (5′-tiRNA^{Ala} and 5′-tiRNA^{Cys}) were found to be the most potent translation inhibitors, and uncapped mRNAs were found to be preferentially targeted. This specific targeting of certain mRNAs suggests that tiRNAs reprogramme gene expression in stressed cells (Yamasaki *et al.*, 2009).

To complicate the picture of tRNA cleavage fragment biogenesis and function further, a recent study has shown that several RNases, including angiogenin can also give rise to ~20nt fragments derived from the 3' end of mature tRNAs (Li *et al.*, 2012). These 3'tRFs can associate with AGO2 and direct gene silencing. It is truly surprising that multiple sRNAs can be derived from tRNAs with such diverse modes of function.

1.8.3   Y RNA-derived small RNAs

Over the last two years, there has been a flurry of publications where Y RNA-derived small RNAs have been detected in deep sequencing studies. Next-generation sequencing of different human tumour samples identified 25nt Y RNA fragments from hY1, hY3 and hY5 which were more abundant than many microRNAs (Meiri *et al.*, 2010). These fragments, which were first described during studies in apoptotic cells (Rutjes *et al.*, 1999a), were suggested to have a function due to their high abundance and sequence stability, and not simply due to the protective properties of Ro60 which remains bound to the fragments after cleavage (Meiri *et al.*, 2010). These Y RNA fragments were also detected in the serum of healthy people in this study.

The study by Meiri *et al.* investigated the potential gene silencing capability of Y RNA fragments using the luciferase assay, but no silencing activity was demonstrated. However, the possibility that these fragments could still be acting as microRNAs has not been ruled out, and has been the subject of a recent review (Verhagen & Pruijn, 2011). Verhagen and Pruijn hypothesised that Y RNAs could be acting as pre-miRNAs in disguise and suggest that Dicer might potentially process the full length Y RNAs into microRNA molecules. This hypothesis has yet to be tested in the laboratory.

Plant siRNA and miRNA targets can be discovered by detecting degraded or cleaved mRNAs, as cleavage is the primary RNAi response in plants. This so called 'degradome' analysis (Addo-Quaye *et al.*, 2008) is a technique that is seldom applied to animal systems owing to the assumed prevalence of translational

repression over cleavage in animal gene silencing. However, Bracken *et al.* applied the technique to mouse embryo and adult tissues (Bracken *et al.*, 2011). As well as detecting microRNA-mediated cleavage products, fragments from endonucleases other than AGO2 were also identified. Multiple instances of cleavage events occurring in non-coding RNAs (ncRNAs) were found, including reads corresponding to both the 3' and 5' stems of the Y RNAs (Bracken *et al.*, 2011). Intriguingly, as the methods used in this study only selected for RNAs with polyadenylated tails, the results show that these Y RNA cleavage products must have been polyadenylated, implying either selective biochemical stabilisation, or indeed, targeting for degradation *in vivo*. Polyadenylation is essential for the function and stabilisation of mRNA transcripts in the cytoplasm, but paradoxically is also a marker for degradation of some sRNAs by the exosome in the nucleus (LaCava *et al.*, 2005).

Finally, a variety of sRNAs have been sequenced in a population of 'shuttle RNAs' present in the vesicles released by immune cells (Nolte-'t Hoen *et al.*, 2012). RNA cleavage products derived from vault RNAs, tRNAs and Y RNAs were detected, whereas microRNAs were significantly underrepresented in the shuttle RNA population. Y RNAs were some of the most abundant RNAs present in the immune vesicles, with both 28 nt fragments and full-length Y RNAs being detected. The authors were astonished by the similarity by which Y RNAs were assimilated into immune cell-derived vesicles (Nolte-'t Hoen *et al.*, 2012), and the nature by which they are selectively packaged into viruses which was reported previously (Garcia *et al.*, 2009). They speculate that the Y RNAs could play a role

in sorting of regulatory RNAs into these vesicles, in stabilising the RNA during export or in guiding shuttle RNAs to specific locations in vesicle-targeted cells.

1.9   Aims and objectives of the thesis

Like many microRNA research laboratories, our laboratory (under Professor Tamas Dalmay) also came across Y RNA fragments in one of our deep sequencing data sets. Originally looking for stress-induced microRNAs, we stumbled across Y RNA-derived sRNAs which were upregulated under certain stress stimuli. Previous reports have only described these fragments, and have never studied them in detail. Here I aimed to characterise the nature, occurrence, biogenesis and potential function of these Y RNA fragments, which I will refer to here on in as 'YsRNAs'. In reading the thesis, it should be appreciated that understanding the biology of YsRNAs would not only shed light on these specific molecules, but would hopefully serve as a model for the emerging class of elusive 30-40nt sRNA fragments already observed, with potential universal biological function.

Chapter 2

Materials and Methods

2.1 Cell culture

2.1.1 Cell lines and growth media

MCF7, HeLa, HepG2 and HEK293 cells were cultured in Dulbecco's modified Eagle's medium (Gibco, 10938) supplemented with 10% fetal bovine serum (FBS) (PAA, A15-101), 1% L-Glutamine (Gibco, 25030) and 1% Penicillin-Streptomycin (Gibco, 15140).

MCF10A cells were purchased from the American Type Culture Collection (ATCC) and were cultured in MEGM Bulletkit medium (Lonza, cc-3150) supplemented with 100ng/mL cholera toxin (Sigma, C8052).

DLD-1 wild-type and DLD-1 dicer$^{-/-}$ exon 5 deletion cell lines were purchased from Horizon Discovery (Cambridge, UK). Both the DLD-1 cell lines, and HCT116 were cultured in DMEM/F-12 + Glutamax (Gibco, 31331), supplemented with 10% FBS and 1% Penicillin-Streptomycin. The dicer$^{-/-}$ genotype was confirmed by Northern blot analysis using a probe for microRNA-21 which was done for each experiment.

Wild-type and RNase L$^{-/-}$ mouse embryonic fibroblast cells (MEF) were provided by Professor Robert H. Silverman's laboratory (Zhou *et al.*, 1997). The MEF cells were grown in RPMI 1640 + GlutaMAX media (Gibco, 61870), supplemented with 10% FBS and 1% Penicillin-Streptomycin. The genotype was checked for each experiment via polymerase chain reaction (PCR) (see following section).

The wild-type and Ro60$^{-/-}$ mouse embryonic stem (mES) cell lines were provided by Professor Sandra L. Wolin's laboratory (Chen *et al.*, 2003). The mES cells were cultured in Knockout DMEM (Gibco, 10829), supplemented with 15% ES cell FBS (Gibco, 10439), 1% L-Glutamine and 1% Penicillin-Streptomycin, 1% mem Non-essential amino acids (Gibco, 11140), 100 u/mL ESGRO leukaemia inhibitory factor (Millipore) and 0.1 mM 2-Mercaptoethanol (Sigma, M7522). Stem cells were grown in flasks prepared with CELLstart CTS substrate (Gibco), or on gelatinised plates prepared with 0.1% gelatin solution (Millipore, ES-006-B) according to the manufacturer's instructions. The Ro60$^{-/-}$ genotype was confirmed for each experiment via Western blot (see section 2.11 on antibodies).

Frozen stocks of cells were prepared in 1 ml cryovials by adding 2.5 - 5 x 10^6 cells in a solution containing 80 % media, 10 % additional FBS and 10 % dimethyl sulfoxide (DMSO) (Sigma). After addition of DMSO, cryovials were placed in an isopropanol bath at -80°C and left overnight. The following morning, cryovials were transferred to a liquid nitrogen filled Dewar for storage.

2.1.2   PCR genotyping of RNase L$^{-/-}$ cell lines

The RNase L$^{-/-}$ cell lines were created by using a targeting vector to insert the neomycin resistance gene (*neo*) in the reverse orientation into the RNase L gene (Figure 2.1) (Zhou *et al.*, 1997). In order to confirm the genotype of these cells, PCR analysis of extracted DNA from RNase L$^{-/-}$ and wild-type cells was done using primers which map a stretch of DNA flanking the *neo* gene (Figure 2.1). PCR products for insertional mutants should be 2.1 kb and 900 bp for wild-type

cells. A 'no DNA' control was also performed. DNA was extracted from mutant and wild-type cells (4 x 10^6 cells each) using the DNeasy blood and tissue kit (Qiagen). PCR primers used were RNase L (Hind2) 5'-GGAGGAGAAGCTTTACAAGGTG-3' and RNase L (F1) 5'- GCATTGAGGACCATGGAGAC-3'. PCRs were performed using Phusion Hot Start high-fidelity polymerase (NEB). PCR cycling parameters were: 98°C for 2 min, followed by 44 cycles of: 98 °C (30 s), 54 °C (30 s), 68 °C (5 min). This was then followed by 72 °C (10 min), and finally a 4 °C end temperature. PCR products were analysed on a 1 % agarose gel made with 0.5 x Tris/Borate/EDTA buffer (TBE) and stained with ethidium bromide (EtBr). Samples were electrophoresed in a glycerol/bromophenol blue loading buffer. A 100 bp DNA molecular weight ladder was used for fragment size analysis (Roche, 11721933001).
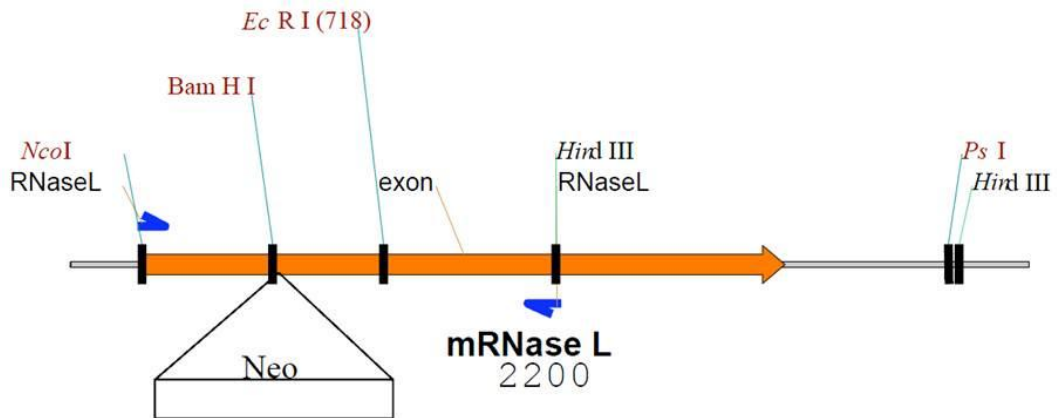
Figure 2.1 Map for the 5' end of the RNase L gene and primer binding sites in mouse embryonic fibroblast cell lines.

RNase L$^{-/-}$ cells were made by insertional inactivation using the neomycin resistance gene (*neo*) into the 5' end of the RNase L gene (Zhou *et al*., 1997). Genotype of both wild-type and mutant cells was achieved via PCR using the primers shown in blue. PCR products for insertional mutants should be 2.1 kb and 900 bp for wild-type cells. Fragment sizes were assessed on agarose gels. Image courtesy of the Silverman laboratory.

2.1.3   Passaging and cell growth conditions

All cells were grown in a 37 °C, 5 % CO_2 humified incubator and were only allowed to grow to ~75 % confluency prior to passaging.

All cell lines, except mES and MCF10A cells, were passaged using 0.25% Trypsin - ethylenediaminetetraacetic acid (EDTA) (Gibco). This was achieved by removing cell media from a flask of growing cells, and replacing with a volume of trypsin sufficient to cover the entire surface of the flask. Cells were incubated with trypsin for 3 minutes at 37 °C at which point all cells had dissociated from the flask

surface. After agitation of the flask by tapping and shaking to dislodge any remaining adhered cells, the trypsin digestion reaction was stopped by adding an equal volume of appropriate growth media to the flask, followed by gentle mixing. Cells were counted using a haemocytometer.

Following cell dissociation from the flask and cell counting, cells were centrifuged to a pellet at 1,500 rpm (363 x g) for 2 min. The supernatant media was subsequently removed, and cells were resuspended in fresh growth media, and re-seeded at a 1/10 dilution in order to achieve an appropriate confluency in the newly passaged flask.

For mES and MCF10A cells, 0.05 % Trypsin-EDTA was used. In the case of MCF10A cells, 0.1 % soybean trypsin inhibitor (Gibco, 17075-029) in phosphate buffered saline solution (Gibco, 14190) was used to neutralise the trypsin, as opposed to culture media.

2.2   siRNAs and transfection

All transfections were carried out on cells (~ 60 % confluency) in 6-well cell culture plates (Nunc) using Lipofectamine 2000 reagent (Invitrogen, 11668) and Opti-MEM + GlutaMAX (Invitrogen, 51985) according to the manufacturer's instructions. Cells were seeded at 150,000 cells/well in a 6-well plate the day before transfection reactions in media without antibiotics.

All siRNAs were used at a final concentration of 50 nM, and were transfected twice over a 48 hour period. Transfection reactions were allowed to

progress for 5 hours, after which the siRNA/Optimem media was replaced with the appropriate growth media.

All siRNAs used were purchased from Qiagen: The negative control siRNA used in all experiments was the ALLStars negative control (Qiagen, 1027281), sequence not disclosed. Ro60 siRNA: FlexiTube siRNA Ro60 (Hs TROVE2) (Qiagen, SI03239628), (+)5'-GGUAGUAUACUCAACGCUATT-3', (-)5'-UAGC GUUGAGUAUACUACCCA-3'. La siRNA (custom designed): (+)5'-GAAACAG ACCUGCUAAUACTT-3', (-)5'-GUAUUAGCAGGUCUGUUUCTT-3', already used in a previous study (Sommer *et al.*, 2011). Two siRNAs were used together to target nucleolin: siRNAs used were FlexiTube siRNA nucleolin (Qiagen, SI00300923), (+)5'-GCUAUGGAGACUACACCAGTT-3', (-)5'-CUGGUGUA GUCUCCAUAGCTT-3', and (Qiagen, SI02654925), (+)5'-GGAAAUGGCCAA ACAGAAATT-3', (-)5'-UUUCUGUUUGGCCAUUUCCTT-3'. Two siRNAs were used together to target RNase L: siRNAs used were FlexiTube siRNA RNase L (Qiagen, SI00129493), (+)5'-AGACUACACUAGUCCAUAATT-3', (-)5'-UUAUGGACUAGUGUAGUCUGG-3' and (Qiagen, SI03079279), (+)5'-CUA UGAUUGGCAAACUCAATT-3', (-)5'-UUGAGUUUGCCAAUCAUAGGG-3'. hAgo2 siRNA (custom designed): (+)5'-GCACGGAAGUCCAUCUGAAUU-3', (-)5'-UUCAGAUGGACUUCCGUGCAC-3'. Angiogenin siRNA: Flexitube siRNA (Qiagen, SI02780197) (+)5'-CGUUGUUGUUGCUUGUGAATT-3', (-)5' UUCACAAGCAACAACAACGTT-3'. RNH1 siRNA: Flexitube siRNA (Qiagen, SI03164805) (+)5'CGGUUAGCAACAACGACAUTT-3', (-)5'-AUGUCGUU GUUGCUAACCGTG'3'.

2.3   Poly(I:C) treatment

Polyinosinic:polycytidylic acid potassium salt (PolyI:C) (Sigma, P9582) was prepared in distilled H_2O, and was used at a final concentration of 10 μg/ml. Poly(I:C) was transfected into cells using Lipofectamine 2000 reagent and Opti-MEM + GlutaMAX as already described. Poly(I:C) transfection media was removed from cells after 5 hours of incubation and replaced with fresh media. Cells were harvested for RNA/Protein extraction 3 hours after this media change (8 hours after exposure to poly(I:C).

For experiments where gene knockdown was required prior to poly(I:C) stress treatment, the following experimental plan was implemented: day 1: cells were seeded, day 2: first siRNA transfection, day 3: second (repeat) siRNA transfection, day 4: poly(I:C) transfection followed by RNA and protein extraction after 8 hours.

2.4   RNA extraction

Total cellular RNA was extracted from cells using the phenol based Trizol reagent (Invitrogen, 15596) according to the manufacturer's instructions, with amendments to the protocol mentioned below. Specifically, 1 ml of Trizol was used per 3 wells in a Nunc 6-well plate. Growth media was removed via pipette, and Trizol added, followed by a brief 1 min incubation period at room temperature. This was followed via vigorously pipetting the Trizol up and down in order to dissolve cells. For cells treated with poly(I:C) where most of the cells were floating

in the growth media, the media was first centrifuged in order to obtain a pellet of cells. The cell pellet was then dissolved in Trizol along with any remaining cells which were still adhered to the bottom of the well.

All stages of RNA extraction following treatment of cells with Trizol were performed under 'RNase-free' conditions. All pipettes, bench surfaces, gloves and other consumables were treated with RNaseZAP (Sigma) prior to RNA extraction in order to avoid enzymatic degradation of the extracted RNA.

Trizol used to dissolve cells (1 ml) was centrifuged at 13,000 rpm for 5 min at 4°C to remove insoluble debris. After transfer of the supernatant to fresh centrifuge tubes, 200 µl chloroform was added and vortexed vigorously for 15 s, followed by 2 min incubation at room temperature. The solution was then centrifuged at 13,000 rpm for 15 min. The aqueous phase from the Trizol–chloroform step was added to excess Isopropanol (1 ml) as this has been found to be more effective at precipitating small RNAs (unpublished data). Additionally, RNA precipitation was carried out for at least 2 h at -20°C, which also improves sRNA precipitation. After precipitation, 1 µl GlycoBlue (Ambion, AM9515) was added and mixed via vortexing prior to centrifugation in order to improve formation and visualisation of the RNA pellet. RNA pellets were washed in 1 ml 75 % ethanol. Air-dried RNA pellets were resuspended in 30 µl of analytical reagent grade H_2O (Fisher Scientific) and stored at -80°C.

2.5   Assessment of RNA quality and concentration

RNA quantification was achieved using the Nanodrop 2000 spectrophotometer (Thermo Scientific). Sample purity was assessed using the 260/280 ratio, which should be ~2.0 for RNA. RNA quality and integrity was evaluated by analysing 1µl of RNA on a 1.5 % agarose gel made with 0.5 x Tris/Borate/EDTA buffer (TBE) and stained with ethidium bromide (EtBr). RNA was added to a 95 % formamide based gel loading buffer; Gel loading buffer II (Ambion, 8546G), and denatured for 2 min at 70°C. Following 20-30 min electrophoresis at 120 V, the 28S and 18S ribosomal bands were observed by eye in order to assess RNA degradation.

2.6   Protein extraction and quantification

Proteins were extracted from cells using RIPA buffer (Sigma, R0278) supplemented with protease inhibitor (Sigma, P8340) according to the manufacturer's instructions. Specifically, 300 µl of RIPA buffer was used per 3 wells in a Nunc 6-well plate. Growth media was first removed, followed by washing the adhered cells with cold Phosphate Buffered Saline (PBS) (Gibco, 14190). After removal of PBS, RIPA buffer was added to cells, followed by vigorous 'scratching' of cells with an upturned 1 ml pipette tip. The RIPA buffer was then incubated on ice for ~ 5 min. Protein lysate was then clarified by centrifugation at 8,000 x g for 10 min at 4°C. This was followed by flash freezing the lysate using liquid nitrogen. For cells treated with poly(I:C), the growth media

was first centrifuged to obtained a cell pellet from the floating cells, as already described, followed by RIPA buffer treatment.

Protein quantification was done using the Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Scientific, 23227) following the manufacturer's protocol. Detection of colour change due to Cu^+ production by the BCA reagent was achieved using the Wallac 1420 VICTOR2 multilabel, multitask plate reader. A standard curve was generated for each quantification by determining the absorption of Bovine Serum Albumin (BSA) solutions of known concentrations in triplicate. Standard curves and experimental sample concentrations were generated and calculated in Microsoft Excel.

2.7   Small RNA Northern blotting

2.7.1   Polyacrylamide gels for small RNA Northern blotting

In order to detect small RNAs, total cellular RNA (1-10 μg) was separated on a 15 % denaturing polyacrylamide gel using the mini PROTEAN III system (Bio-Rad) using 1.0 mm gel plates. To make a single gel, 2.1 g urea was dissolved in 1.25 ml distilled water and 0.5 ml 5 x TBE in a 50 ml centrifuge tube (with ~20 s microwaving to facilitate dissolving of the urea). Following this, 1.85 ml 40 % acrylamide/bis solution 19:1 (Bio-Rad) was added to the centrifuge tube, and the final volume was increased to 5ml using distilled water. After gentle mixing of the acrylamide-urea solution via several tube inversions, acrylamide polymerisation was achieved by adding 2.5 μl tetramethylethylenediamine (TEMED) and 50 μl of

a 10 % ammonium persulfate (APS) solution. This solution was pipetted into the gel cast rapidly, followed by adding a 10 well comb with 1.0 mm thickness. Additionally, it was found that the quality of the wells could be improved by applying constant pressure to the top of the gel plates during polymerisation by using small bulldog clips. This technique should only be used with urea based gels as application of the clips to standard non-urea polyacrylamide gels can cause a formation of a gap between the gel and the cover plate during polymerisation.

For experiments involving poly(I:C), 1 μg total RNA per lane was typically sufficient to detect Y RNA cleavage fragments. RNA samples were added to 95 % formamide based Gel loading buffer II (Ambion, 8546G) and denatured for 2 min at 70°C prior to gel loading. Electrophoresis was done in 0.5 x TBE buffer at 110 V for ~ 2 h.

2.7.2   Northern blotting of RNA

Following gel electrophoresis, gels were stained with ethidium bromide (EtBr) and imaged on a UV transilluminator (UVP) to check successful RNA separation. Transfer of RNA to nylon membrane was performed as previously described (Pall *et al.*, 2007). Briefly, semi-dry transfer was achieved by placing the gel on a piece of 9 cm x 7 cm nylon hybridisation membrane (Amersham Hybond-NX, GE Healthcare) pre-soaked in 0.5 x TBE. The gel and membrane were sandwiched between layers of filter paper (BLT2, Munktell), with three pieces of 9 cm x 7 cm on either side, also pre-soaked in 0.5 x TBE. The layers of filter paper, gel and membrane were arranged in a semi-dry electroblotter (Scie-Plas, V20-

SDB), removing excess buffer and air bubbles by rolling the surface with a 50 ml centrifuge tube. Transfer was done at 3 mA/cm^2 for 45 min at 4°C. For example, 250 mA for a single 9 cm x 7 cm membrane at constant Amps. This equates to approximately 15-20V for 45-60 minutes at 4°C.

2.7.3   Carbodiimide-mediated cross-linking of RNA

RNA was chemically cross-linked to Hybond-NX as previously described (Pall *et al.*, 2007). It was proposed by Pall *et al.* that 3-(N-morpholino) propanesulfonic acid (MOPS) should be used instead of tris-based buffers during the gel making and running process, as it was thought that tris might interfere with 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-mediated cross-linking. This is not the case, as discovered by researchers in the laboratories of Professor Tamas Dalmay, and Professor Sir David Baulcombe. Indeed, use of TBE buffers instead of MOPS actually results in superior gels and 'crisper' band detection following oligonucleotide hybridisation, as demonstrated by side-by-side comparisons in the Dalmay laboratory (data not shown).

Cross-linking solution was prepared by adding 122.5 μl 12.5 M 1-methylimidazole to 10 ml distilled water, with the pH adjusted to pH 8 using 10 μl HCL. This solution was added to 0.373 g EDC (Sigma), giving a 0.13 M solution. The final volume was adjusted to 12 ml final using distilled water.

A piece of filter paper slightly larger than the nylon membrane was then soaked in the EDC solution, with excess solution subsequently removed. The nylon membrane was then placed onto the EDC soaked filter paper, with the RNA-side

facing up. This was then wrapped in Saran wrap and incubated at 60°C for 1-2 h. The membrane was then washed twice in distilled water in order to remove excess cross-linking solution.

2.7.4   Pre-hybridisation of Northern blot

Membranes were placed into hybridisation tubes with the RNA side facing the inside of the tube. A volume of UltraHyb-Oligo hybridisation buffer (Ambion) was added which was sufficient to cover the membrane. The tubes were then placed in a HB-1000 hybridisation oven (UVP), rotating at 37°C for 1 h.

2.7.5   Northern blot probe synthesis

Antisense DNA oligonucleotide probes were labelled with $[\gamma\text{-}^{32}\text{P}]$-ATP (Perkin Elmer, NEG-502A-500UC) using 1 µl of T4 polynucleotide kinase (PNK) (Invitrogen). Specifically, 3 µl of γ-ATP (~ 1.4 MBq) was used to label 2µl of 10 µM oligonucleotide (without a 5' phosphate) using the 'forward' reaction buffer. The solution, with a final volume of 20 µl, was incubated at 37°C for 1 h. Following incubation, unincorporated nucleotides were removed by first, diluting the reaction solution with an equal volume of distilled water, and applying the γ-^{32}P oligonucleotide labelled solution to a home-made Sephadex G-25 (GE Healthcare) spin column, and centrifuging at 2,000 rpm for 2 min. The eluate was then added to the hybridisation buffer in the tube containing the Northern blot

following pre-hybridisation. Probe hybridisation was conducted at 37°C in a rotating oven overnight.

Sephadex G-25 columns were made by piercing the base of a 0.5 ml centrifuge tube with a 0.8 mm x 40 mm (21 G x $1\frac{1}{2}$ in) needle (Becton Dickinson) (only inserting half of the 'face' of the needle through the tube), and placing this tube into a larger 1.5 ml centrifuge tube. A few drops of glass bead solution was added to the 0.5ml centrifuge tube, followed by 300-400 µl Sephadex G-25. Column formation was achieved by centrifugation at 2,000 rpm for 2 min.

2.7.6   Post-hybridisation washing of Northern blots and signal detection

After overnight incubation, hybridisation buffer containing the γ-^{32}P labelled probe of interest which had not hybridised to the membrane was discarded. The nylon membrane was then subjected to 2 x 30 min washing steps using a 0.2 x Saline-Sodium Citrate (SSC), 0.1 % Sodium Dodecyl Sulphate (SDS) solution at 37°C. Membranes were then wrapped in Saran wrap, and signal was detected using phosphorimager screens (Fujifilm). The Bio-Rad molecular imager, FX proplus, was used for signal visualisation. Membranes could be re-used by stripping hybridised radioactive oligonucleotides by placing membranes in a solution of boiling 0.1 % SDS, with shaking for 30 min - 1 h.

2.8 Quantitative Northern blotting and statistical analysis of band intensities

In order to calculate the concentration of an endogenous sRNA, oligonucleotides with the same sequence as the endogenous target were designed and purchased either in DNA form (Sigma) or RNA form (Dharmacon). Serial dilutions of synthetic sequences for a given target with concentrations ranging from 0.1 fmoles – 75 fmoles (depending on the experiment) were loaded on a 15 % denaturing polyacrylamide gel and subjected to Northern blotting. Between 0.1-10 µg of total cellular RNA from an experimental sample was also run on a Northern blot. Both the serial dilution and experimental blots were hybridised to a probe with a complementary sequence to the target of interest in the same hybridisation tube, and exposed to phosphorimager screens together for the same time period. Signal strength of band intensities was quantified using the Quantity One software (Bio-Rad). Background signal strength was first removed from each of the raw values of the bands. A calibration curve was then created using Microsoft Excel for the serial dilution of synthetic targets. This calibration curve was then used to calculate the concentration of sRNA in the experimental target.

For analysis of Y RNA fragmentation from full length Y RNAs, signal strength intensities for a given Y RNA/YsRNA were normalised by dividing by the band intensity values of U6 snRNA (loading control) for each respective lane.

For statistical analysis comparing the ratios of particular band intensities, t-tests and Z-tests were performed by Dr Irina Mohorianu from the School of Computing Sciences, University of East Anglia.

2.9   Oligonucleotide probe sequences for Northern blot hybridisations

2.9.1   Antisense DNA oligonucleotide probe sequences (Sigma):

Y5 3' end: 5'-AGCTAGTCAAGCGCGGTTGTGGGGG-3',

Y4 3' end: 5'-AGCCAGTCAAATTTAGCAGTGGGGG-3,

Y3 3' end: 5'-TAGTCAAGTGAAGCAGTGGGAG-3',

Y1 3' end: 5'-AGACTAGTCAAGTGCAGTAGTGAGAA-3',

Y5 5' end: 5'-TAACCCACAACACTCGGACCAACT-3',

Y4 5' end: 5'-TAACCCACTACCATCGGACCAGCC-3',

Y3 5' end: 5'-AACACCACTGCACTCGGACCAGCC-3',

Y1 5' end: 5'-TAACTCACTACCTTCGGACCAGCC-3',

miR-21: 5'-TCAACATCAGTCTGATAAGCTA-3',

miR-25: 5'-TCAGACCGAGACAAGTGCAATG-3',

miR-93: 5'-CTACCTGCACGAACAGCACTTTG-3',

tRNA Met 5' end: 5'-AGCACGCTTCCGCTGCGCCACTCTGCT-3',

U6 snRNA: 5'-GCTAATCTTCTCTGTATCGTTCC-3'.

2.9.2   Locked nucleic acid (LNA) modified oligonucleotide probes (Exiqon):

LNA is a modified RNA nucleotide which has a high binding affinity for RNA and results in an increased sensitivity when detecting target molecules. The LNA oligonucleotides used here contained DNA nucleotides with an LNA

molecule substituted every third base (denoted by a '+' prefixed to the base which was substituted with LNA).

Y5 (LNA) 'loop' probe: 5'- G+AGA+CAA+TGT+TAA+ATC-3'

2.10   Western blotting

2.10.1 Polyacrylamide gel electrophoresis of proteins

For protein electrophoresis, 10 µg of total protein was boiled in 1.5 ml centrifuge tubes for 5 min at 100°C in an equal volume of Laemmli buffer (Sigma), and then returned to ice. Proteins were separated on a precast NuPAGE 4-12 % Bis-Tris polyacrylamide gel (Invitrogen) using the XCell Surelock mini-cell electrophoresis system (Invitrogen). Electrophoresis was performed at 120 V in NuPAGE Mops SDS running buffer (NP0001). The protein size markers used were Kaleidoscope pre-stained standards (Bio-Rad, 161-0324) and Precision Plus protein standards (Bio-Rad, 161-0374).

2.10.2 Blotting of proteins

Proteins were transferred to Immun-blot Polyvinylidene fluoride membrane (PVDF), 0.2 µm (Bio-Rad, 162-0177). PVDF was first wetted in methanol. The gel and six pieces of filter paper (BLT2, Munktell), were wetted in NuPAGE transfer buffer (Invitrogen, NP0006-1) supplemented with 20 % methanol. The PVDF was

also wetted in transfer buffer following methanol 'activation'. The wetted gel was put on the PVDF and placed between pieces of filter paper (3 pieces each side). The layers of gel, PVDF and filter paper were then secured between layers of transfer buffer soaked sponge pads. The entire complex was than arranged in the XCell II blot module (Invitrogen) and filled with transfer buffer. Transfer was done at 300 mA for 45 min at 4 °C.

Following protein transfer to PVDF membrane, the membrane was then washed three times in phosphate buffered saline (made using Dulbecco `A' Tablets, Oxoid). To avoid non-specific antibody hybridisation, the PVDF membrane was blocked for 1 h in a 5 % milk solution (Marvel) in PBS, supplemented with 0.1 % Tween20 (PBS-Tween).

2.10.3 Western blot antibody hybridisation

Western blot membranes were hybridised to the appropriate primary antibody at dilutions recommended by the manufacturers. Hybridisations were performed in 10 ml of 5 % milk, 0.1 % PBS-Tween in 50 ml centrifuge tubes rotating at 4 °C overnight in a hybridisation oven. The following day, membranes were washed three times for 15 min in PBS-Tween (0.1 %), without milk. Hybridisation of horseradish peroxidise (HRP) conjugated secondary antibodies, at dilutions specified by the manufacturers, was achieved in 5 % milk PBS-Tween in 50 ml centrifuge tubes rotating at room temperature for 1 h. After hybridisation with the secondary antibody, membranes were washed three times as before in PBS-Tween.

Antibody signal was detected using enhanced chemiluminescence (ECL) SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, 34080). The membrane was then incubated in activated substrate for 30 s - 3 min and then quickly removed and wrapped in Saran wrap. The Glow Writer Pen (Diversified Biotech) was used to mark on the Saran the positions of the size markers. Signal was detected using the Fujifilm darkbox (LAS-3000). After signal visualisation, PVDF membrane was stripped of antibodies using ReBlot Plus Strong antibody stripping solution (Millipore) according to the manufacturer's instructions.

2.11   Antibodies for Western blotting

All antibodies were used at dilutions recommended by the manufacturer. Antibodies used were: human Ro60 (Santa Cruz, sc-20961), La (Abnova, H00006741-M01), Nucleolin (abcam, ab13541), Ago2 (abcam, ab57113), β-actin (abcam, ab6276), Angiogenin (abcam, ab10600 and Santa Cruz, sc-1408) and RNase L (abcam, ab13825). For detection of Ro60 in mouse, a previously described home-made antibody from the Wolin laboratory was used (Chen *et al.*, 2003). Secondary antibodies used were: Rabbit polyclonal Secondary Antibody to Mouse IgG - H&L (HRP) (Abcam, ab6728) used for detecting mouse Ro60, La, Nucleolin, RNase L, angiogenin (abcam antibody) and β-actin primary antibodies. Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat (Sigma, A0545) for detecting human Ro60 primary antibody. Finally, donkey anti-goat IgG-HRP (Santa Cruz, sc-2020) was used for detecting angiogenin (Santa Cruz antibody).

2.12   Y5 expression construct

The Y5 expression construct (pY5) was created by another PhD student in the lab, Carly Turnbull. Using the forward primer: 5'-AATACTAGTGAAGATCCATGGAGGTACATC-3' and the reverse primer 5'-GTAAACGTTGTCTACTACTGTTATTAGTGC-3' the hY5 gene, including endogenous promoter and terminator sequences, was amplified from HeLa genomic DNA and cloned into the pGEM-T Easy vector (Promega). The sequence was checked by Sanger sequencing on both the sense and antisense strands using the pUC/M13 forward (5'-CGCCAGGGTTTTCCCAGTCACGAC-3') and pUC/M13 reverse (5'- TCACACAGGAAACAGCTATGAC-3') primers.

The pY5 construct was transfected (1 µg per well in a 6-well plate with a 1 ml volume) using Opti-MEM + GlutaMAX and Lipofectamine 2000 reagent as already described.

2.13   Angiogenin expression construct

The angiogenin expression construct (pcDNA-ANG) was a gift from Xiaofei Zheng's laboratory (Fu *et al.*, 2009). The sequence was checked by Sanger sequencing on both the sense and antisense strands using the T7 (5'-TAATACGACTCACTATAGGG-3') and SP6 (5'ATTTAGGTGACACTATAG-3') primers.

The pcDNA-ANG construct was transfected (0.5 μg per well in a 6-well plate with a 1 ml volume) using Opti-MEM + GlutaMAX and Lipofectamine 2000 reagent as already described.

2.14   Plasmid purification

Plasmids were transformed into super competent DH5α bacterial cells via heat shock. This was done using a 25s heat shock at 42 °C, followed by addition of 1 ml of super optimal broth medium supplemented with 20 mM glucose (SOC) with a one hour recovery period in a 37°C shaker incubator. Bacteria were then plated and colonies containing the desired plasmid were selected for on Luria broth (LB) plates with 100 μg/ml ampicillin. Plasmid recovery was achieved via a purification kit (Qiagen) following inoculation of a single bacterial colony in 5 ml LB with 100 μg/ml ampicillin. All bacterial handling and plating was done using aseptic techniques.

Following sequence confirmation, desired plasmids were purified to high concentration using large scale 'midiprep' plasmid purification (Qiagen). Concentration and quality of DNA was assessed using the Nanodrop 2000 spectrophotometer (Thermo Scientific). A sample absorbance 260/260 nm ratio was checked, which should be ~1.8 as an indication that the analysed sample is 'pure' DNA.

2.15 Dideoxynucleotide chain termination (Sanger) sequencing

Following plasmid purification, sequences were checked via dideoxynucleotide chain termination (Sanger) sequencing using the Big Dye terminator kit (Applied Biosystems). The Big Dye sequencing reaction contained 1 μL of recovered plasmid, 1 μL of BigDye, 1 μL of Half Big Dye, 1 x sequencing buffer and 1 μL of sequencing primer in a 10 μL reaction. The polymerase chain reaction (PCR) used to generate dideoxynucleotide termination products was performed using the following cycling parameters: 96 °C (4 min) followed by 25 cycles of: 96 °C (30 s), 50 °C (15 s), 60 °C (4 min). This was followed by a 4 °C holding temperature. The PCR reactions were sent to the Genome Analysis Centre, Norwich where the sequencing reactions were run. Sequence analysis was performed using the FinchTV software, version 1.4.0 (Geospiza).

2.16 Anion exchange chromatography of non-stressed and poly(I:C) treated cells

MCF7 cells were plated in 175 cm^2 flasks (4 flasks per experiment) and allowed to grow to 70 % confluency. Flasks were either treated or not-treated with poly(I:C). For poly(I:C) treated flasks, cells were harvested 24 h after poly(I:C) induction, using 0.25 % Trypsin-EDTA as already described. Following trypsin neutralisation and pellet formation, cells were washed in cold Phosphate Buffered Saline (PBS) (Gibco, 14190). Following PBS washing and cell pellet formation via centrifugation (from 4 flasks combined), cells were lysed in 1 ml of column buffer

(20 mM Tris/HCl pH 7.5, 300 mM NaCl, 5 mM MgCl2, 5 mM DTT) supplemented with RNase inhibitor at 1 µl/ml (Promega) and protease inhibitor at 5 µl/ml (Sigma, P8849). This was achieved via pressure/suction induced lysis using a 2ml syringe and a 0.5 mm x 16 mm (25 G x $\frac{5}{8}$ in) needle (Becton Dickinson). A lysate was created by drawing the cell-lysis buffer mixture through the needle into the syringe and expelling the fluid back into the collection tube ten times. Insoluble material was removed by centrifugation at 16,000 g for 5 min at 4 °C.

Anion exchange was performed by a postdoctoral researcher in the laboratory, Dr Tibor Csorba. The extracts were separated using a diethylaminoethanol (DEAE) -anion exchange column (Pharmacia). Fractions from the anion exchange column were eluted in increasing salt concentrations between 100 mM and 800 mM. RNA was extracted from every second fraction and separated on a 15 % denaturing polyacrylamide gel and subjected to Northern blotting as already described.

2.17   Ago2 pull-down assay

Ago2 immunoprecipitation (pull-down) was performed by a postdoctoral researcher in the laboratory, Dr Francisco Esteban Nicolas. Four flasks of 175 cm^2 confluent MCF7 cells were treated with trypsin and washed twice with PBS. Cells were lysed using 2 ml of Lysis Buffer (Promega, E3971) plus 2.5 mg/ml of heparin (Sigma). Lysate was spun down for 10 min at 10,000 rpm and the supernatant was moved to a clean tube. Then 10 µl of primary antibody Ago2 (Abcam, ab57113)

was added followed by incubation and rocking for 4 hours at 4ºC. Following this, 50µl of protein A beads were blocked in 2 ml of lysis buffer plus 2.5 mg/ml of heparin for 1 h. Beads were spun down (2,000 rpm, 3 min) and mixed with the lysate, followed by incubation and rocking for 1 h at 4ºC. This was then washed three times followed by spinning the beads down and resuspension in 500 µl of lysis buffer with heparin 5 mg/ml. Then there was a final wash only with lysis buffer. This was eluted three times with 100 µl of glycin/HCl pH=2.3 100 mM. These three elutions were pooled and neutralized with 75 µl of Tris/HCl pH 9.0 (1M). This was then purified with 1.2 ml of trizol.

2.18   Angiogenin *in vitro* cleavage experiments

Experiments using recombinant angiogenin protein to cleave total cellular RNA or cell lysate were based on previous experimental designs (Fu *et al.*, 2009; Li *et al.*, 2012). Recombinant human angiogenin powder (R&D Systems, 265-AN-050) was made up to a 10 µM (140 µg/ml) stock solution in a filter sterilised buffer (PBS containing 0.1 % bovine albumin), and used at a working concentration of 1 µM. Recombinant RNase A (Ambion, AM2269) (73 µM stock, 1 mg/ml) was used at a working concentration of either 1 µM or 1.5 nM. The experimental 'angiogenin buffer' (30 mM HEPES pH 6.8, 30 mM NaCl, 0.0001 % BSA) was adjusted to pH 7 and then filter sterilised. The working buffer in the absence of any RNase was used as a control.

Reactions were performed at 100 µl final volumes using 20 µg total RNA per reaction in 1.5 ml centrifuge tubes. For experiments using cell lysates, cells

were first dissociated from tissue culture flasks using trypsin. The cells were then washed in PBS, and then 2×10^6 cells were lysed in 90 µl of angiogenin buffer using pressure/suction induced lysis with a 2ml syringe and a 0.5 mm x 16 mm (25 G x $\frac{5}{8}$ in) needle. The number of cells used was based on the assumption that one representative mammalian cell contains ~ 10 pg RNA (Tang *et al.*, 2011). This would equate to 20 µg RNA per 2×10^6 cells, which would make the *in vitro* cleavage RNA and cell lysate experiments comparable. After addition of angiogenin or RNase A, reactions were incubated on a heating block at 37 °C for between 2 min – 1h. Reactions were stopped by addition of 1 ml Trizol. RNA was then extracted as already described.

2.19 Adjustment of pH and filter sterilisation of buffers

The pH of solutions and buffers was measured using the Suntex microprocessor pH meter SP-2200, applying dropwise addition of concentrated HCl or NaOH to adjust the pH value. Buffers and solutions which could not be autoclaved were filter sterilised by passing liquid through 0.2 µm filters (Sartorius Stedim Biotech) using a syringe.

2.20 Mouse organ dissection and subsequent RNA isolation

All mice were provided by the Disease Modelling Unit, School of Biological Sciences, University of East Anglia. Adult male and female mice (~ 2 months old) from BALB/c or C57BL/6 strains were sacrificed by a technician.

This was done humanely by cervical dislocation at the same time of the morning each time a mouse was used. Dead mice were quickly dissected (within 30 min of sacrifice) following brief sterilisation of the fur with 100 % ethanol. All dissection instruments used were autoclaved and sterile. Organs dissected were the brain, heart, both lungs, liver, both kidneys and femoral muscle from the left leg. Once extracted, organs were quickly cut into slices where at least one dimension was no more than 0.5 cm across. These tissue samples were placed in 5 volumes of RNA*later* solution (Sigma) and kept on ice. Prior to RNA extraction, tissue fragments for each organ were removed from RNA*later* solution and weighed on a digital scale. The mass of tissue fragments for each organ was used to calculate the volume of Trizol (Invitrogen) required to extract RNA as recommended by the manufacturer. Tissues were ground in Trizol using sterile pestle and mortars, applying liquid nitrogen intermittently. RNA pellets were resuspended in 200 µl of distilled H_2O for each organ.

2.21   Purification of sRNA (<200 nt) from total RNA

In order to isolate the sRNA fraction (<200 nt) from tissues, organs and cell culture extracts, 20 µg of total RNA was processed using the *mir*Vana miRNA Isolation kit (Ambion). The small RNA fraction was eluted in a 50 -100 µl volume of distilled H_2O. This RNA was then concentrated via ethanol precipitation and resuspended in 10 µl distilled H_2O, giving a total recovery of ~ 2 µg sRNA from 20 µg total RNA.

2.22 Concentrating RNA via ethanol precipitation or vacuum drying

In order to concentrate RNA, two techniques were employed. Where only a small concentration was required (~ 20 µl reduction in volume), vacuum concentration was used. RNA solutions were centrifuged in the Eppendorf Concentrator 5301 in open-capped 1.5 ml centrifuge tubes. The rate of reduction in volume during centrifugation is approximately 1 µl/min.

Where a greater volume concentration was required, ethanol precipitation was implemented. Taking a 100 µl volume of RNA, 12.5 µl of 3 M sodium acetate, 375 µl 100 % ethanol and 2 µl GlycoBlue (Ambion, AM9515) was added and briefly vortexed. This solution was incubated at -80°C for 2 h, followed by centrifugation at 13,000 g for 20 min. The supernatant was then removed and followed by washing with 95 % ethanol. The air-dried RNA pellet was then resuspended in the desired volume of distilled H_20.

2.23 Small RNA library preparation

Total RNA was isolated via Trizol extraction (Invitrogen) and small RNA was enriched using the *mir*Vana miRNA Isolation kit (Ambion) as already described. Small RNA libraries were created using the TruSeq Small RNA preparation kit (Illumina, RS-200-0012 & RS-200-0024) using variable index primers (barcodes) to allow for multiplexing of samples. The protocol provided by the manufacturer was used to create libraries with amendments specified below. For the mouse organ sequencing experiments, 500 ng input sRNA was used per

reaction, and for the mouse embryonic stem cells/Ro60$^{-/-}$ experiments, 2 µg input sRNA was used. Two biological repeats were done for both the mouse organ experiment and for the mES/Ro60$^{-/-}$ experiment.

Reactions were performed using half the volumes recommended by the manufacturer up to the reverse transcription (RT) stage, at which point full recommended reaction volumes were used. Truncated T4 RNA ligase 2 (NEB) was used for 3' adapter ligation, and SuperScript II reverse transcriptase (Invitrogen) for creation of cDNA. PCR amplification of cDNA was achieved using Phusion Hot Start high-fidelity polymerase (NEB) using 4 µl of the RT reaction in two PCR reactions with different cycle numbers, with 20 µl final volume. The PCR conditions provided in the manufacturer's protocol were used except 15 & 17 cycles were used for the mouse organ library preparation, and 17 & 19 cycles for the mouse embryonic stem cells/Ro60$^{-/-}$ experiment. These cycling parameters were optimised for library preparation of the 30-40mer sRNA fraction. The two separate PCR reactions for each sample were pooled for gel extraction.

DNA libraries were separated on 8 % polyacrylamide gels and stained using SYBR gold. The 30-40mer class of sRNAs were isolated by cutting the appropriate portion of the gel with a sterile razor blade using the size markers provided by the kit as a guide.

2.24   Illumina (Solexa) sequencing and bioinformatics

Libraries were sent to Baseclear (Netherlands) where they were checked on a BioAnalyser (Agilent DNA 1000 chip) and quantified. Libraries were then

pooled and sequenced using the HiSeq 2000 system (Illumina) with a 50 cycle read length. The sequencing was done using two different lanes for each experiment. For the mouse organ experiment, the two biological repeats (6 organs from 2 different mice) were sequenced on different lanes. For the mES/Ro60$^{-/-}$ experiment, sequencing was performed in duplicate such that each sample was sequenced on each of 2 lanes. Data analysis was done by Dr Helio Pais, Matthew Beckers and Dr Irina Mohorianu from the School of Computing Sciences, University of East Anglia.

Chapter 3

Y RNAs are fragmented into Y RNA-derived sRNAs in a range of cell types under stressed and non-stressed conditions

3.1   Introduction

Cells need to respond to adverse conditions brought about by environmental change or internal fluctuations in biochemical states. This 'stress' response needs to be implemented quickly and efficiently in order to restore cellular homeostasis. One of the ways cells respond to stress is by altering the expression of microRNAs, which in turn tailors gene expression in order to adapt to new environmental conditions (Leung & Sharp, 2010). In the laboratory of Professor Tamas Dalmay, we study microRNAs in both animal and plant systems. Previous work in the Dalmay laboratory has demonstrated that plants use microRNAs to respond to stress. Specifically, it was shown that miR-395 in *Arabidopsis* is upregulated in response to sulphur starvation in different cell types (Kawashima *et al.*, 2009), a nutrient which is essential for the synthesis of some amino acids. Increased miR-395 levels results in the differential regulation of sulphate transporter proteins which in turn facilitates the favoured translocation of sulphate to plant shoots (Kawashima *et al.*, 2009; Kawashima *et al.*, 2011).

A different study in the Dalmay laboratory aimed to uncover novel stress related microRNAs in animal cells. Here, postdoctoral researcher Dr Francisco Esteban Nicolas implemented a range of stress stimuli on the common breast cancer cell line, MCF7. These included heat shock, double heat shock, nutrient starvation, viral infection and a control. In order to mimic viral infection, polyinosinic:polycytidylic acid (Poly I:C) – an immunostimulant structurally similar to double-stranded RNA, was transfected into cells. RNA was then extracted from all stressed and control cells and subjected to Solexa (Illumina)

sequencing. The RNA size class corresponding to ~22 nt miRNAs was selected for library preparation. Strikingly, only three microRNAs were identified as being differentially regulated during stress. MiR-1246, miR-1979 and miR-1975 showed a significant increase in abundance in cells subjected to poly(I:C) treatment only. Further investigation via a literature search and using the Basic Local Alignment Search Tool – BLAST (Altschul *et al.*, 1990) revealed that the 25 nt miRs-1979 and 1975 were in fact fragments derived from non-coding RNAs called Y RNAs.

Y RNAs were first discovered during investigations into systemic lupus erythematosus (SLE) as they were found to associate with the autoantigen Ro60 (Lerner *et al.*, 1981). Y RNAs have since been implicated in biological processes such as chromosomal DNA replication (Christov *et al.*, 2006) and RNA quality control (O'Brien & Wolin, 1994).

The Dalmay laboratory found that miR-1979 was derived from the 3' end of human Y3 RNA (hY3) and miR-1975 was derived from the 3' end of human Y5 RNA (hY5). Indeed, a study published shortly after we discovered this also noted that these putative miRNAs were actually Y RNA fragments (Meiri *et al.*, 2010), and they were subsequently removed from the microRNA online database, miRBase. Interestingly, these Y RNA fragments were first described 11 years prior to our discovery in cells undergoing apoptosis, where Y RNAs appeared to be specifically targeted for degradation (Rutjes *et al.*, 1999a). No study has yet looked at these Y RNA-derived sRNAs (YsRNAs) in detail.

In this chapter, the aim was to characterise the nature and occurrence of Y RNA fragmentation in a range of animal cell lines. YsRNAs mapping to both the 3' and 5' ends of all four Y RNAs were identified in Solexa (Illumina) sequencing

data. The YsRNAs were detected in stressed cells via small RNA Northern blot and were also detected in non-stressed cells at levels comparable with microRNAs.

3.2   Results

3.2.1   Y RNAs give rise to fragments originating from their 5' and 3' ends

Preliminary work by Dr Francisco Esteban Nicolas comparing poly(I:C) treated and untreated MCF7 cells utilised Solexa sequencing libraries which selected for 19-24 nt sRNAs. It was in these experiments that YsRNAs were shown to be upregulated during poly(I:C) treatment. Following Northern blot validation, it was noted that several of these candidates also showed larger bands, ~30 nt. It was then decided that new libraries should be made which would include this longer size class of small RNAs.

MCF7 cells were subjected to a stress treatment by transfecting cells with poly(I:C). The effect of poly(I:C) on cell morphology can be seen in figure 3.1. Untreated MCF7 cells displayed a polygonal shape typical of other epithelial cells and were adhered to the cell culture flask. After poly(I:C) treatment (8h post-transfection – when RNA was harvested), many cells had disassociated from the flask and displayed a shrunken, spherical morphology with blebbing of the plasma membrane, typical of cells undergoing apoptosis (Figure 3.1). RNA was then extracted from poly(I:C) treated and untreated cells and cDNA libraries were generated for the 19-33 nt size range of small RNAs. A size distribution of sequences revealed an expected peak around 22 nt representing the microRNA population, and an unexpected peak of around 30-33 nt (Figure 3.2). Following Solexa sequencing, 21,837,679 redundant reads were obtained for the poly(I:C) treated dataset, of which 79 % mapped to the genome. Out of the total number of

reads, this included 490,826 unique (non-redundant) reads of which 33 % mapped to the genome. The discrepancy between the percentage of mapped reads for redundant and unique sequences was owing to the presence of a small proportion of sequences with very high read counts which could be mapped. The 50 sequences with the highest read counts made up 83 % of the mapped reads (34 % of which were microRNAs, and 56 % of which were tRNA, snoRNA and rRNA fragments). Sequence reads were normalised using the total redundant read numbers which mapped to the genome.
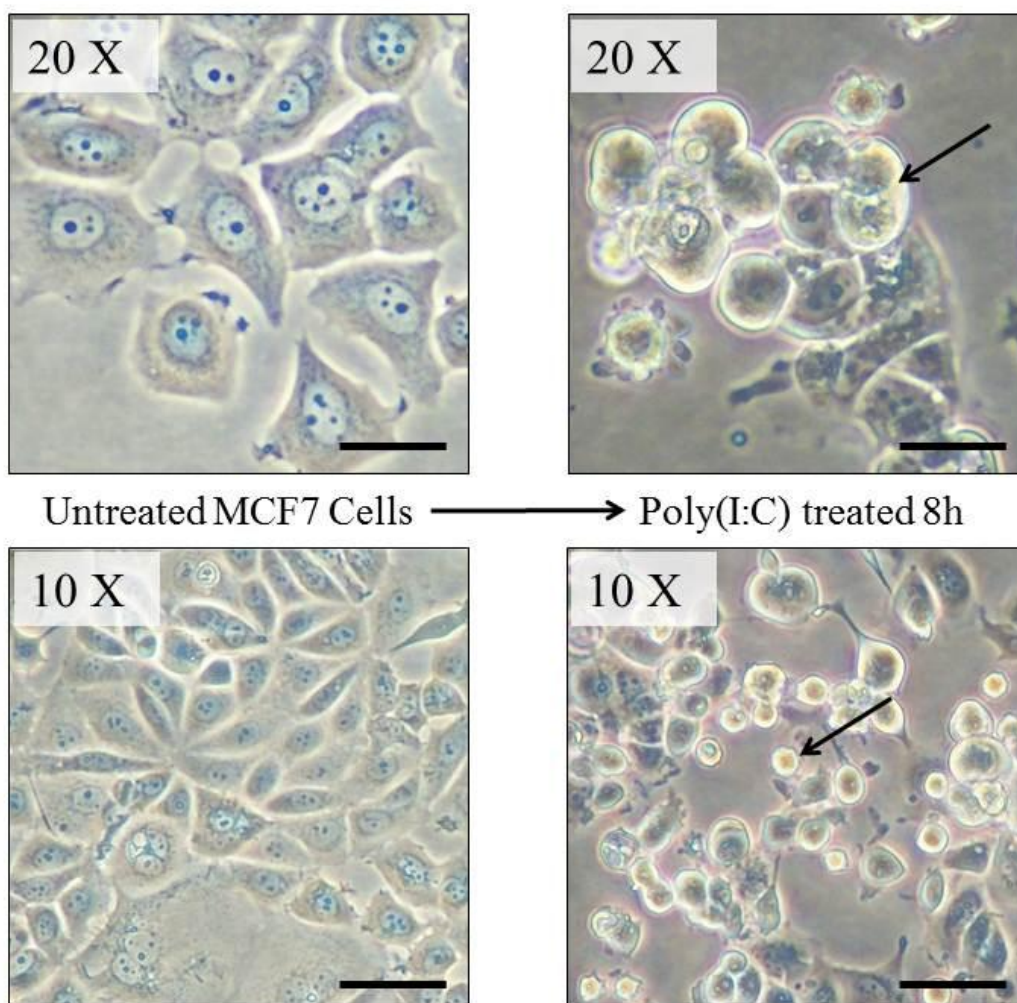
Figure 3.1 Morphological effect of poly(I:C) on MCF7 cells.

Photographs of normal MCF7 cells prior to treatment, and 8 h after poly(I:C) transfection (when RNA/protein samples were harvested). Photographs taken down a microscope lens at 20 X magnification (top photographs) and 10 X magnification (bottom photographs). Untreated MCF7 cells (left panels) looked normal for healthy epithelial cells, with a typical polygonal shape and were firmly attached to the cell culture flask in discrete patches. After 8 h following transfection with poly(I:C) (right panels), many cells had disassociated from the flask, with adhered cells looking more spherical and shrunken in size, with blebbing of the plasma membrane (indicated by arrows) characteristic of apoptosis. Scale bar represents 20 μm (approximate size).
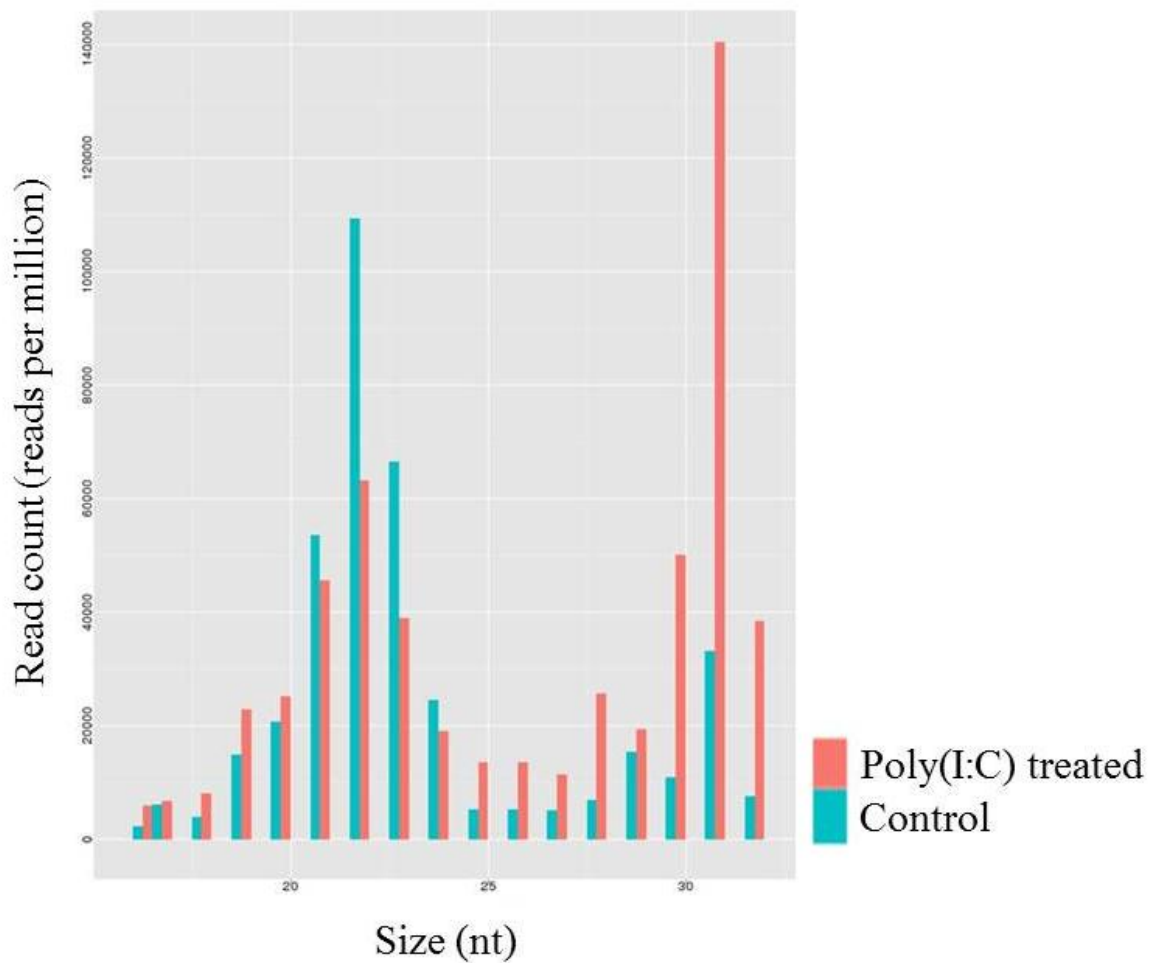
Figure 3.2 Size distribution of sequences from MCF7 cells treated with poly(I:C) or not treated (control).

RNA was extracted from poly(I:C) and untreated MCF7 cells. Libraries of 19-33 nt were then generated (work by Dr Francisco Esteban Nicolas) and sequenced using Solexa sequencing technology. Two peaks are clearly visible: a peak of ~22 nt representing the microRNA population, and a peak of around ~30 nt which includes Y RNA fragment sequences.

Reads corresponding to YsRNAs were detected which were derived from all four human Y RNAs. When mapped to the Y RNA structures, these sequences corresponded exactly to the 5'/3' stems of the Y RNAs, and included some of the single stranded region of the large internal pyrimidine rich loop (Figure 3.3A). No sequences corresponding to the remaining loop region were detected.

Figure 3.3 YsRNA sequences detected in poly(I:C) treated MCF7 cells by Solexa sequencing.

(A) Y RNA structures (adapted from Christov *et al*., 2006) with the principle YsRNAs generated following poly(I:C) treatment highlighted in yellow. The Y RNA fragment size is also shown. (B) YsRNA read counts (normalised as reads per million using redundant reads mapping to the genome). The number of reads for the principle cleavage fragment generated is shown first. Read counts for the summed abundance of the consensus sequence and detected Y RNA fragment variants, (sequences with a +/- 2 nt deviation and up to 2 mismatches from the consensus sequence), is shown in brackets. The lack of detected variants for some YsRNAs (for example, both Y5sRNAs), may be due to bioinformatic or sequencing bias rather than lack of biological occurrence. Y RNA reference sequence obtained from NCBI database. Y RNA length differences between (A) and (B) is owing to heterogeneous poly-U tails as result of transcription by RNA Pol III.

The YsRNAs ranged in size from 27-32 nt, with the exception of the Y4-5'
fragment which had a shorter length of 18 nt (Figure 3.3A). The fragments from
the 5' and 3' ends of each Y RNA with the highest read counts were selected for as
the 'YsRNA consensus sequence' for each fragment (Figure 3.3B). Fragments with
a +/- 2 nt deviation and up to 2 mismatches from the consensus sequence were also
selected for analysis, and will be referred to as 'YsRNA variants'.

The Y5-3' fragment had the highest read count, showing 1,553 reads per
million reads (RPM) for the consensus sequence (Figure 3.3B). This was followed
by the Y4-3' fragment with 746 RPM and Y5-5' fragments with 357 RPM. All
other YsRNAs had read counts which were around 5-20 fold less than these three,
with the exception of the Y4-5' fragment which had a very low read count of 1
RPM. Half of the eight different YsRNA species had a number of YsRNA variants
with significant read counts, (summed abundance of read counts for the consensus
sequence and YsRNA variants is shown in brackets in figure 3.3B).

The Y1-3' variants showed the highest proportion of sequences contributing
to the summed abundance of variants and consensus read counts, with 41 % of
total fragment reads coming from YsRNA variants. Neither of the Y5-3' or Y5-5'
YsRNAs gave rise to variants using the criteria used.

As a comparison, the normalised read counts of two microRNAs expressed
in poly(I:C) treated MCF7 cells were calculated, taking into account reads with a
+/- 1 nt deviation from the microRNA sequence in miRBase. The normalised read
counts for miR-25 was 215 RPM and for miR-93 was 621 RPM. The read numbers
for the two microRNAs were similar to the read numbers for the YsRNAs.

3.2.2　Detection of YsRNAs via Northern blot

In order to see whether these YsRNAs were *bona fide* small RNAs and not merely unspecific degradation products, the fragments were validated via Northern blot. Initially, the Y5-3' YsRNA was focused on as it had significantly higher read counts compared to the other YsRNAs. Total RNA was extracted from both untreated and poly(I:C) treated MCF7 cells and analysed on a Northern blot, probing with a sequence complementary to the 3' end of Y5 RNA (Figure 3.4A). In the lane for untreated MCF7 cells, a single band corresponding to ~80 nt was detected. In the lane with RNA from poly(I:C) treated cells, the same ~80 nt band was present, along with two additional smaller bands. A band of 32 nt with a very strong signal was present, and a weaker band of 24 nt was also present. The size of these fragments was confirmed using size markers (Figure 3.4B). Comparing the Northern blot with the sequencing data, the larger 32 nt YsRNA fragment corresponded with the principle Y5 3' fragment and the 24 nt fragment corresponded with the Y5 3' stem minus the single stranded tail sequence (Figure 3.4C). This 24 nt fragment was also present in the sequencing data at 289 RPM.
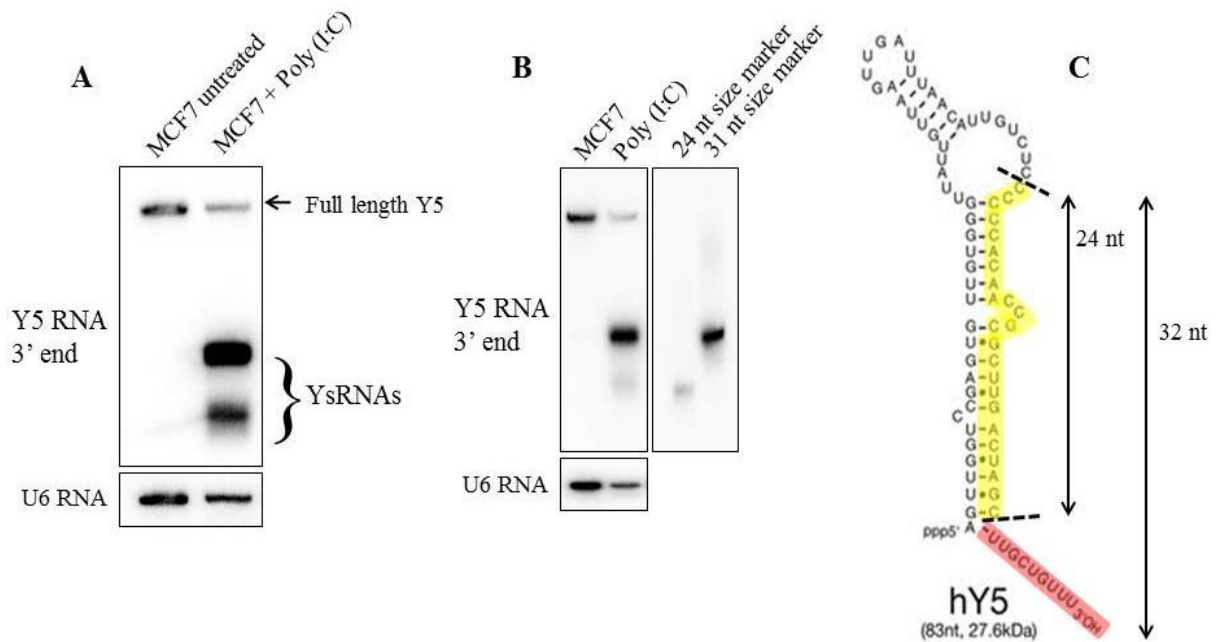
Figure 3.4 Y5 RNA fragment validation (3' end).

(A) Northern blot analysis of Y5 RNA fragmentation following poly(I:C) stress treatment. Full length Y5 RNA gives rise to two smaller fragments. (B) Validating size of Y5 YsRNAs using size markers. Solexa sequencing data showed a 32 nt Y5 3' fragment was the principle fragment and a 24 nt fragment with an overlapping sequence was produced at a lower abundance. 31 nt size marker used instead of 32 nt as preliminary analysis hinted that the longer Y5 3' fragment was 31 nt. (C) The 32 nt and 24 nt fragments mapped to the full length Y5 RNA highlighted in yellow and red. The two fragments share the same sequence and differ by the single stranded 3' tail shown in red.

3.2.3 Y5 RNA 3' fragments are generated under stress in a range of cell types

In order to see whether YsRNAs could be generated under poly(I:C) stress in cells other than MCF7, a range of human cell lines were cultured and either treated with poly(I:C) or not treated. These included the cancerous cell lines HeLa

(cervical cancer cells), DLD-1 (colorectal adenocarcinoma cells), HCT116 (colorectal carcinoma cells) and HepG2 (hepatocellular carcinoma). Y RNAs have been shown to be upregulated in a range of cancerous tissues (Christov *et al.*, 2008), and YsRNAs also have a higher abundance in tumour samples (Meiri *et al.*, 2010). Therefore, two non-cancerous cell lines: HEK293 (embryonic kidney cells) and MCF10A (mammary epithelial cells), were investigated to see whether YsRNA generation was cancer specific. Total RNA was extracted from poly(I:C) treated and untreated cells and analysed via Northern blot, probing for the Y5 RNA 3' fragment (Figure 3.5). Northern blots for all cancerous and non-cancerous cell lines showed a single band representing full length Y5 RNA for untreated cells. All poly(I:C) treated cells also shared this ~80nt band, along with the two shorter bands representing YsRNAs. The band of 32 nt had a stronger signal than the band of 24 nt which was usually weaker and less defined. Upon closer inspection, it was also noted that YsRNA were also detectable at very low levels in some of the untreated cells.

3.2.4   Y5 RNA 3' fragments are produced in non-stressed cells at microRNA-like levels

Only a short exposure time was sufficient to detect Y5 RNA and its cleavage products on a phosphorimager screen after poly(I:C) treatment, and only 1 μg of total RNA was required to detect the YsRNAs. This could have resulted in fragments which were produced at low levels being overlooked. Indeed, weakly produced YsRNAs appeared to be detectable in non-stressed cells (Figure 3.5).

Therefore, 10 μg total RNA from untreated MCF7 cells was analysed on a Northern blot with a Y5 RNA 3' specific antisense probe, using 0.1 μg RNA from poly(I:C) treated cells as a size marker for the YsRNAs (Figure 3.6). The hybridised Northern blot was also left to expose for a longer time period (7 days instead of 7 h) in order to detect possible YsRNAs in cells grown under normal conditions. Intriguingly, the longer Y5 3' fragment (32 nt) was clearly visible in non-stressed cells after Northern blot signal was left to develop for a week (Figure 3.6). An siRNA with no target in the human genome was transfected as a control to test the effect of Lipofectamine 2000 on YsRNA generation. YsRNA generation mildly increased under these conditions (Figure 3.6).
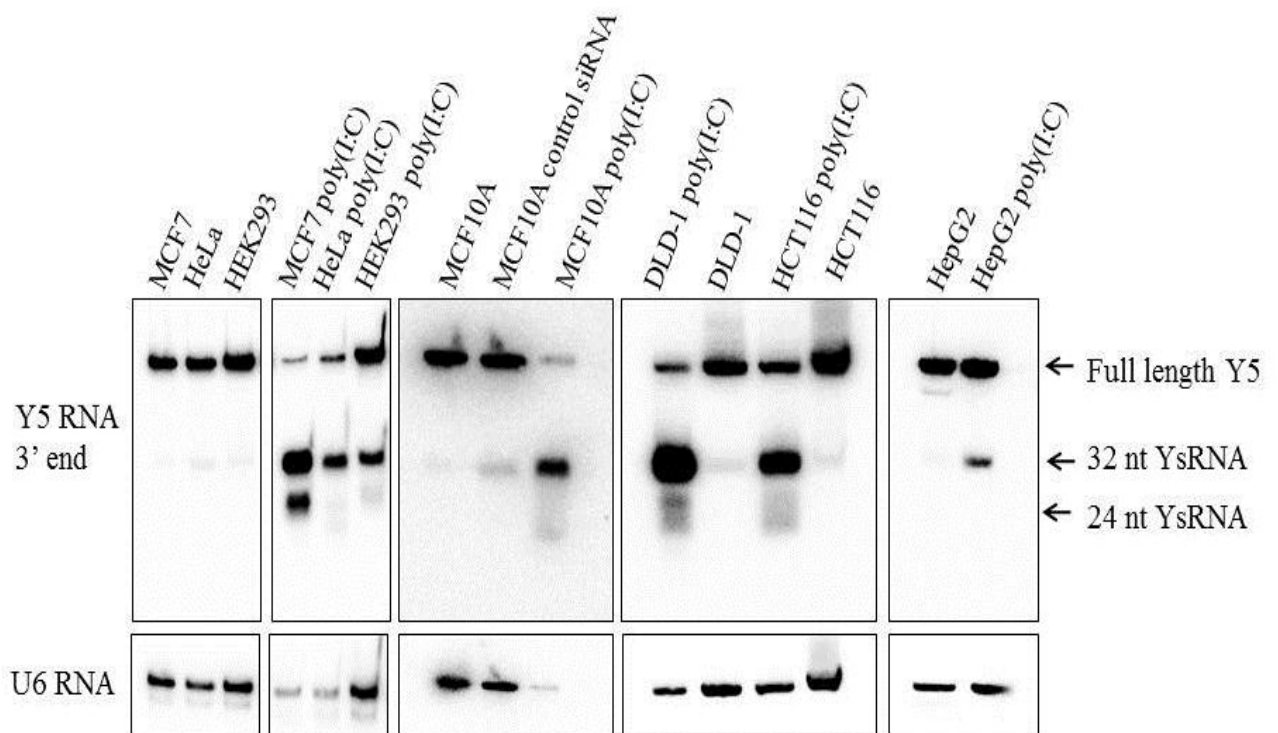
Figure 3.5 Multiple cell lines show Y5 RNA fragmentation during poly(I:C) treatment.

Total RNA extracts from various cell lines, either treated or untreated with the immunostimulant poly(I:C), were analysed by Northern blotting using a Y5-3' antisense probe. These included the cancerous cell lines MCF7 (breast adenocarcinoma), HeLa (cervical cancer cells), DLD-1 (colorectal adenocarcinoma cells), HCT116 (colorectal carcinoma cells) HepG2 (hepatocellular carcinoma), as well as two non-cancerous cell lines; HEK293 (embryonic kidney cells) and MCF10A (mammary epithelial cells). The upper most band is the full length Y5 RNA (83 nt) and the two shorter bands are the Y5 RNA 3' cleavage products of 32 nt and 24 nt. A non-specific siRNA was transfected into MCF10A cells as a transfection control. The membranes were stripped and re-probed with a U6 specific probe to show equal loading.
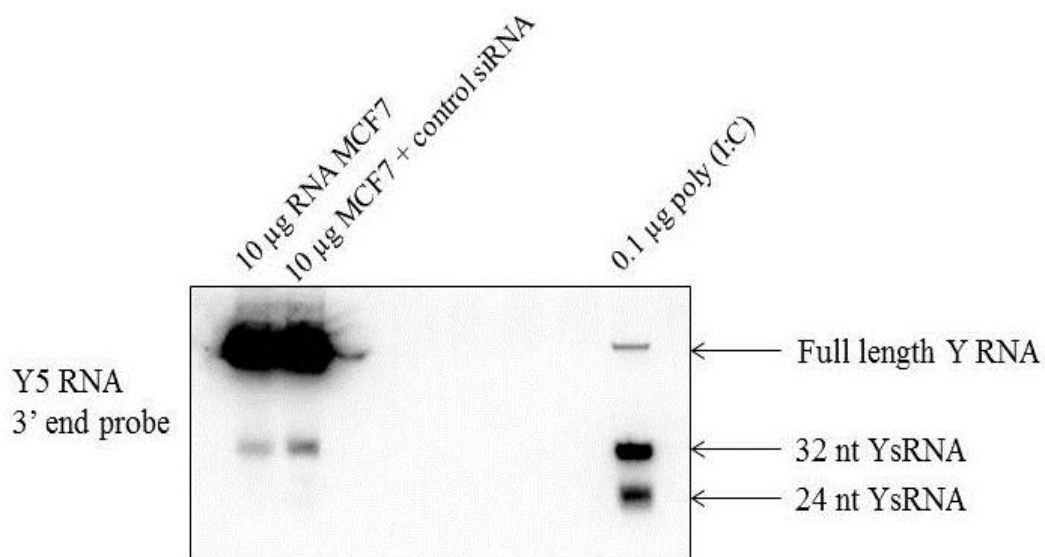
Figure 3.6 Y5 RNA fragments are produced in non-stressed cells.

Total RNA (10 μg) from MCF7 cells grown under normal conditions was analysed on a Northern blot using a probe specific to the Y5 3' RNA fragment. A non-specific siRNA was transfected into MCF7 cells to see how the process of transfection affected YsRNA generation. A small amount of RNA (0.1 μg) from poly(I:C) treated cells was used as a YsRNA size marker. The top band in each of the three lanes represents full length Y5 RNA. The YsRNAs detected in non-stressed cells are indicated with an arrow.

Based on the Northern blot signal (Figure 3.6), the level of Y5sRNA production in untreated cells appeared to be on a par with other sRNAs found in the cell. Therefore, Y5 RNA cleavage products were quantified in non-stressed cells and compared to the levels of two microRNAs. Using a synthetic RNA oligonucleotide corresponding to the longer Y5sRNA, a dilution series was made (Figure 3.7). Dilution series were also made using oligonucleotides corresponding to miR-25 and miR-93. After Northern blot hybridisation and generation of a calibration curve, the normal physiological level of the longer Y5 sRNA was calculated at 0.27 femtomoles/μg total RNA. To determine whether this level was

similar to other short non-coding RNAs, similar experiments were carried out for the two miRNAs. The level of miRNA-25 (0.34 femtomoles/µg of total RNA) and miRNA-93 (1.4 femtomoles/µg of total RNA) was similar to that of the Y5sRNA.

	Cellular concentration (femtomoles/µg total RNA)
Y5 sRNA	0.27
miR 25	0.34
miR 93	1.4
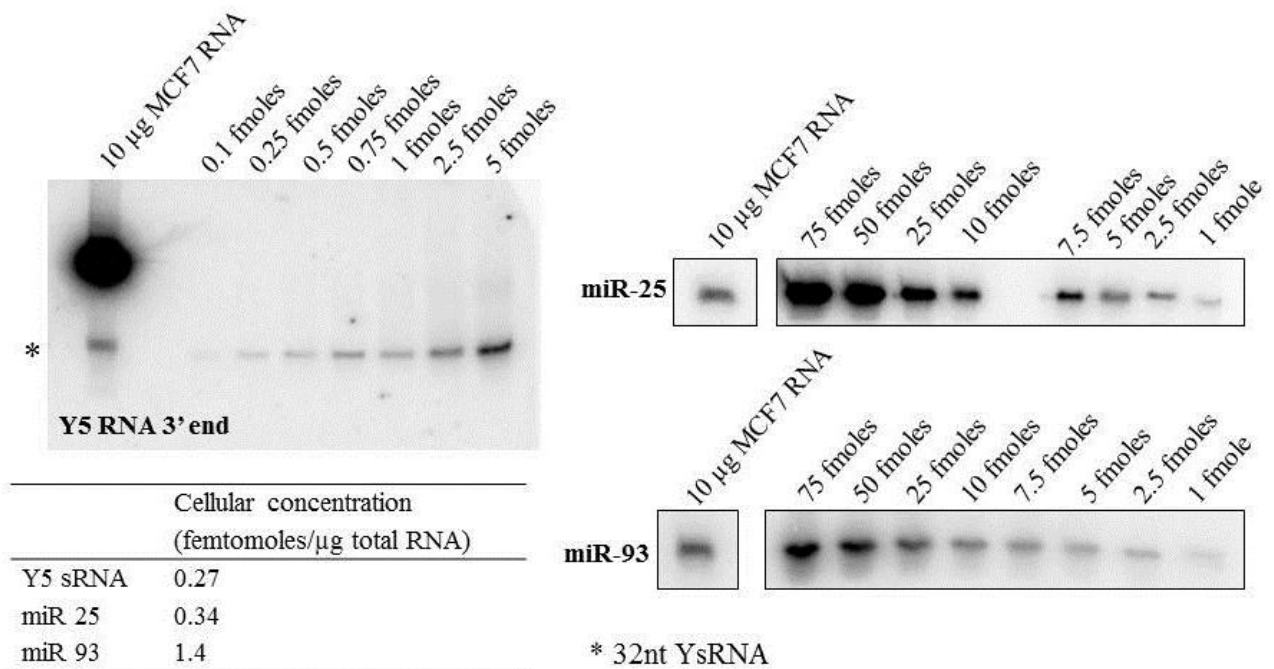
Figure 3.7 Quantification of endogenous Y5sRNA and microRNAs in untreated MCF7 cells.

Serial dilutions were made using RNA oligonucleotides with the same sequence as the 32 nt Y5 3' sRNA fragment. After probing with a Y5 3' specific antisense probe, a calibration curve was created and then used to quantify the endogenous level of the Y5 3' fragment in untreated MCF7 cells. Similar experiments were conducted for two microRNAs as a comparison (miR-25 and miR-93). The Y5sRNA was found to have a cellular concentration of 0.27 femtomoles/µg total RNA. This is comparable to microRNA levels in the cell such as miRNA-25 (0.34 femtomoles/µg total RNA) and miRNA-93 (1.4 femtomoles/µg total RNA). For miRNA quantification, both experimental and serial dilution membranes were hybridised together and signal was developed and exposed under the same conditions.

3.2.5   Validation of other YsRNAs

In order to experimentally validate fragments generated from the other Y RNAs, oligonucleotide probes complementary to the 3' ends of all the Y RNAs were used to analyse RNA from poly(I:C) treated RNA (Figure 3.8). The same membrane was re-probed, and a size marker was used to confirm the size of the YsRNAs generated.

As there is considerable sequence complementarity between the 3'/5' stems of the Y RNAs where the YsRNAs are derived (48 % sequence similarity based on a sequence alignment of the 25 nt from the last dsRNA stretch at the 3' end of all the Y RNAs), the four YsRNA 3' specific probes were tested using complementary sequences to check there was no cross-hybridisation of the different probes to the different Y RNAs (Figure 3.9). The four probes appeared specific and no cross-hybridisation was detected for each probe used (Figure 3.9).

The sizes of the YsRNAs (3' end) from all four Y RNAs largely agrees with the sizes derived from the Solexa sequencing data in Figure 3.3. The Solexa data indicates that Y1, Y3 and Y4 RNAs give 27 nt 3' derived YsRNAs. The Northern blot analysis appears to agree with this (Figure 3.8). Similarly, the Y5 YsRNA appears to be the longest, around 32 nt, which is also confirmed by Northern blot data (Figure 3.4B and 3.8).
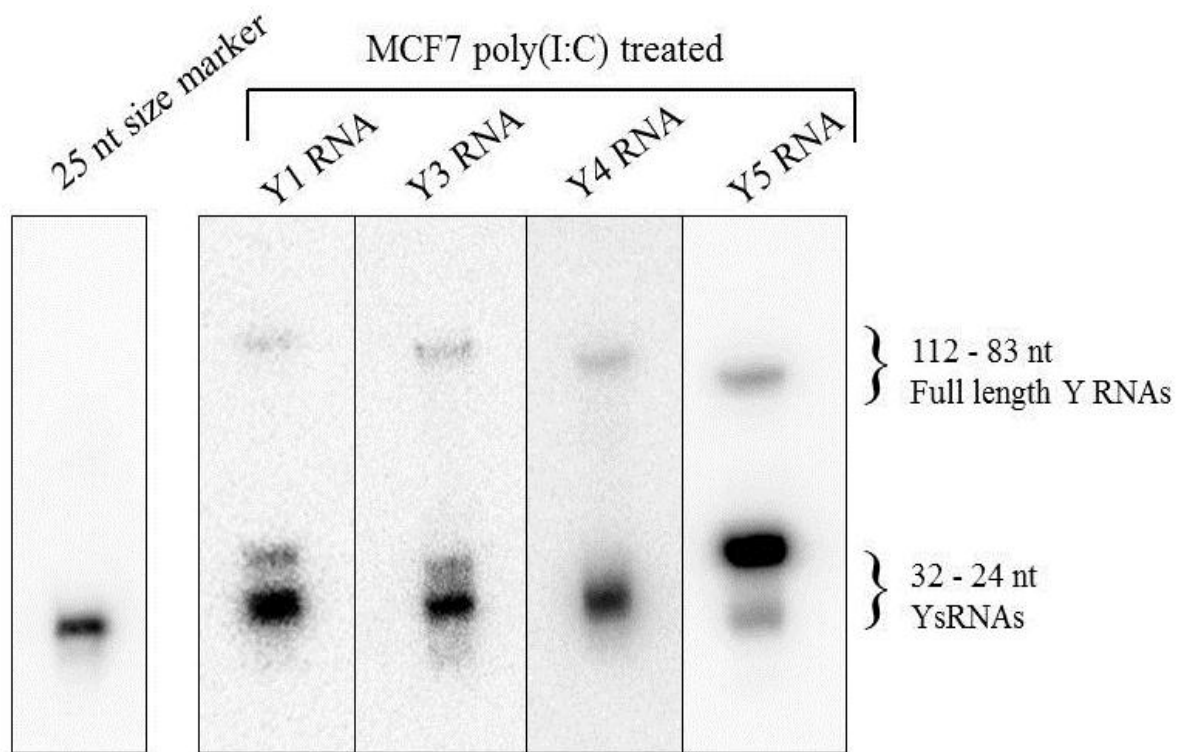
Figure 3.8 Validation and comparison of YsRNAs from all four Y RNAs.

MCF7 RNA from poly(I:C) treated cells was analysed via Northern blot. The same membrane was re-probed with oligonucleotides complementary to the 3' ends of all the Y RNAs. Owing to the sequence similarity of the stem regions of the Y RNAs, the probes used to detect YsRNAs 1-5 were all tested to make sure there was no cross-hybridisation of the different probes to the different Y RNAs.
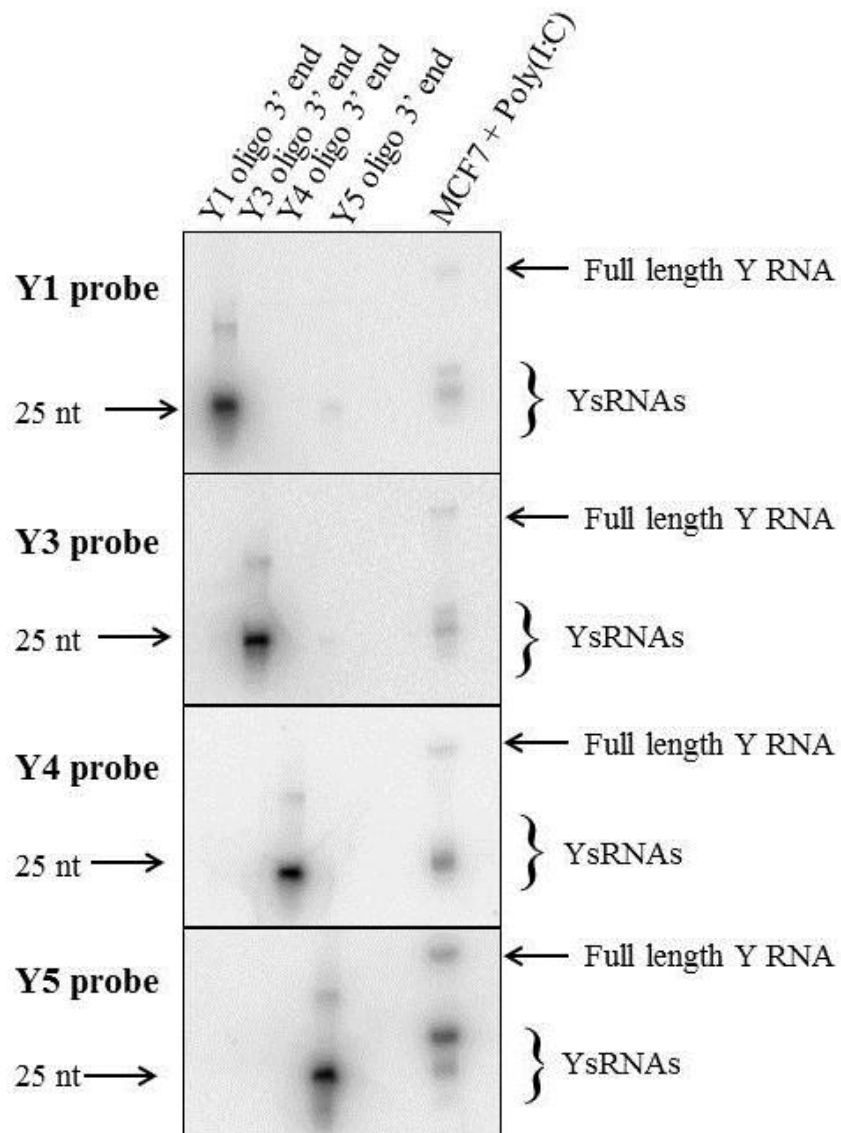
Figure 3.9 Specificity of the probes used to detect YsRNAs (3' end derived).

Due to the sequence complementarity between the stems of the Y RNAs where the YsRNAs are derived (48 % sequence similarity based on a sequence alignment of the 25 nt from the last dsRNA stretch at the 3' end of all the Y RNAs), the four YsRNA 3' specific probes were tested using complementary sequences to check there was no cross-hybridisation of the different probes to the different Y RNAs. The same membrane was used for all hybridisations.

3.3   Discussion

Previously, YsRNAs have been detected in cells undergoing apoptosis (Rutjes *et al.*, 1999a) and were also found to be upregulated in cancerous cells (Meiri *et al.*, 2010). These two findings seem to be contradictory because if YsRNAs were upregulated to assist the uncontrolled cell proliferation associated with cancer cells, why would they also be upregulated in cells undergoing cell death? As Y RNAs have been shown to be essential for DNA replication (Christov *et al.*, 2006), one explanation could be that the fragmentation of Y RNAs somehow disrupts the controlled regulation of chromosomal DNA replication - which would certainly fit with the apoptosis scenario, as replication of genetic material is not a priority in cells undergoing cell death. In the case of increased fragmentation in cancerous cells, perhaps some oncogenic factor would replace the role of Y RNAs as initiators of DNA replication which would facilitate uncontrolled cellular proliferation? However, the fragmentation of Y RNAs appears to maintain the integrity of the 9 base pair motif sequence which has previously been shown to be essential for chromosomal DNA replication when in a double stranded form (Gardiner *et al.*, 2009). As we do not know whether the 5' and 3' YsRNAs remain bound to each other following fragmentation, it is not clear what affect Y RNA fragmentation would have on the function of this 9 bp motif sequence.

Alternatively, YsRNA generation and upregulation during cell stress could facilitate the reported quality control function of Ro60 (O'Brien & Wolin, 1994) resulting in a moderated chaperone activity so that Ro60 targets a different subset of misfolded RNAs generated under adverse conditions.

Here it was shown that all four Y RNAs give rise to fragments deriving from the 3' and 5' ends which has not previously been reported. Strikingly, this pattern of 3'/5' fragmentation is identical to the fragmentation pattern of other sRNAs fragments derived from longer RNAs such as tRNA, snoRNA and rRNA fragments (Brameier *et al.*, 2011; Li *et al.*, 2012). Indeed, Li *et al.* found that the processing of sRNAs from these longer precursor molecules occurs asymmetrically, similar to microRNA/microRNA* strand level bias. Comparing the normalised read counts for YsRNAs, there also appears to be an asymmetric bias for either the 5' or 3' YsRNA for each of the Y RNAs. With the exception of Y3 RNA, YsRNAs derived from the 3' end have higher read counts than those from the 5' end.

The asymmetric processing of all the YsRNAs needs to be confirmed via Northern blot analysis. During library preparation for Solexa sequencing of stress induced microRNAs, sRNA populations were not dephosphorylated to remove 5'-polyphosphates. The T4 RNA ligase used during library preparation can only ligate the 5' adapter to 5'-monophosphates, which does not present a problem during microRNA library preparation as these molecules have a native 5'-monophosphate. However, with novel sRNAs, 5'-tri or diphosphates (or indeed other modifications) may be present, such as with the Y RNAs which all have a 5'-PPP. This would not present a problem with the 3' YsRNAs, as following cleavage, a 5' phosphate and 3' OH group would be present which is required for Solexa library preparation. The differential biochemical modifications at the termini of 5' and 3' YsRNAs may explain the high read counts from 3' YsRNAs compared with 5' YsRNAs. In retrospect, if it was known that the investigation would have developed into a non-microRNA based project, the library preparation protocol would have been revised.

This would have involved firstly, dephosphoroylating all sRNAs using an enzyme such as alkaline phosphatase, followed by re-phosphorylation of a 5'-monophosphate using polynucleotide kinase. Experimental strategies used for sequencing novel sRNAs with a variety of 5' and 3' end modifications has been well reviewed elsewhere (Zhuang *et al.*, 2012). The reason why 5' YsRNAs still had significant read counts in the experiment reported here despite not dephosphorylating/re-phosphorylating during library preparation may be due to the presence of a small population of Y RNAs with 5' monophosphates, perhaps due to degradation or processing errors, or due to a varying dynamic range of 5'-monophosphate containing Y RNAs within a given cellular population.

In the case of some of the Y RNAs, two fragments with an overlapping sequence appear to be generated from the same strand. This is clearly demonstrated by weak, smeared bands shown in figure 3.8 which run near to the principally generated fragment. Specifically for the 3' Y5sRNA, a 32 nt and a 24 nt fragment are both clearly present in the sequencing data and also appear on a Northern blot, with the 32 nt being much more abundant. As the 3' tail region is the only difference between the longer and shorter fragments (Figure 3.4C), this differential fragment formation may be due to differential binding of the La protein which has been shown to bind to this region (Pruijn *et al.*, 1991). This will be investigated in chapter 5. It remains to be seen if the longer and shorter Y5 3' sRNAs both have biological functions. As the 32 nt Y5sRNA is produced at a much higher level, it is possible that the 24 nt Y5sRNA fragment is functionally redundant.

The level of the longer Y5sRNA in normal, non-stressed cells was found to be comparable to some microRNAs (Figure 3.7). As microRNAs are thought to

regulate 30-50 % of mammalian protein coding genes (Filipowicz *et al.*, 2008), this indicates that YsRNA molecules may have specific functions even in the absence of apoptosis. However, it cannot be ruled out that in a population of cultured cells, there may be a small number of cells that are undergoing apoptosis even if this is not evident under the microscope.

Chapter 4

Biogenesis of Y RNA-derived small RNAs is

independent of the microRNA pathway

4.1   Introduction

MicroRNAs are endogenously produced, small ~22 nt molecules which regulate gene expression by targeting protein coding transcripts and preventing their translation or inducing mRNA cleavage (Bartel, 2004). A detailed description of microRNA biogenesis and function can be found in chapter 1 (section 1.7.2).

Y RNA-derived sRNAs (YsRNAs) are generated in cells following poly(I:C) treatment, and are also produced in non-stressed cells, as demonstrated in chapter 3. These YsRNAs are similar in size to microRNAs. Indeed, other non-coding RNAs which have biological functions which are unrelated to gene silencing, such as snoRNAs and tRNAs, have been shown to produce RNA fragments which can enter the microRNA pathway (Ender *et al.*, 2008; Cole *et al.*, 2009). One group has tried to address the possibility that YsRNAs have gene silencing potential by using the Luciferase assay experiment (Meiri *et al.*, 2010). Meiri *et al*. tried to see whether the activity of a Luciferase gene containing a complementary sequence to the YsRNAs could be reduced in cancerous cells which produce these fragments. No gene silencing activity was detected, although the authors did not rule out the possibility that these YsRNAs could act as microRNAs. Meiri *et al.* relied on endogenous YsRNA levels in Hep3B cells to mediate silencing and did not transfect artificial siRNA-like molecules containing YsRNA sequences. Therefore, the endogenous levels of possible miRNAs derived from Y RNAs might not have been high enough to detect silencing capability using the Luciferase assay system. Interestingly, two YsRNAs derived from the 3' ends of Y3 and Y5 RNA were found in the microRNA database miRBase annotated as

miR-1979 and miR-1975 respectively. Both miR-1975 and miR-1979 have subsequently been removed from miRBase because they are derived from Y RNAs.

A recent review has also discussed the hypothesis that Y RNAs might be processed into microRNAs (Verhagen & Pruijn, 2011). It was argued that Y RNAs and pre-miRNAs are both very similar in length and share a similar stem-loop secondary structure. It was noted that microRNA genes are found clustered in the genome, as are Y RNA genes which are within a single cluster on chromosome 7q36. Verhagen and Pruijn (2011) also suggested that under stress conditions, Ro60 might be released from the Y RNAs and enter the nucleus, allowing Dicer to access the exposed Y RNAs and generate the YsRNAs. These fragments would then be available to bind to AGO proteins and mediate gene silencing.

As the question of whether YsRNAs can act as microRNAs remains disputed and unresolved in the literature, this hypothesis was addressed experimentally here. Using human dicer$^{-/-}$ cells it was shown that YsRNAs derived from all four Y RNAs are still produced following poly(I:C) treatment. Collaborative work also revealed that YsRNAs are in different complexes than microRNAs following analysis of cellular fractions from anion exchange chromatography, and that YsRNAs are not immunoprecipitated by Ago2. Therefore, it was concluded that YsRNAs do not enter the microRNA pathway.

4.2   Results

4.2.1   Production of YsRNAs is Dicer independent

As the vast majority of microRNAs are produced from pre-miRNAs by the RNAse III enzyme Dicer, the generation of YsRNAs was investigated in dicer deficient cells (dicer$^{-/-}$). Both wild-type DLD-1 and dicer$^{-/-}$ DLD-1 cells were treated with the immunostimulant poly(I:C) in order to induce a stress response which has been shown to upregulate YsRNA generation (Chapter 3). Total RNA from these cells was then analysed by Northern blot, using antisense oligonucleotide probes for all of the YsRNAs (3' end derived) (Figure 4.1A), and also using miR-21 to demonstrate the Dicer$^{-/-}$ genotype (Figure 4.1B).

As expected, a single band representing the full length Y RNA was present in both wild-type and dicer knockout cells under normal conditions for all four Y RNAs. When transfected with poly(I:C), wild-type DLD-1 cells gave rise to the expected YsRNAs described in Chapter 3. A principle YsRNA was present for each Y RNA represented by the band with the strongest signal. Additionally, YsRNAs were also represented by the presence of weaker bands or a smeared signal.

For dicer knockout cells treated with poly(I:C), the YsRNA fragments from each of the four Y RNAs were still produced, even in the case of Y3sRNA which is only weakly produced in wild-type cells. Indeed, the production of YsRNA looked stronger in Dicer$^{-/-}$ cells in the case of Y5sRNA and Y4sRNA. Interestingly,

in the case of Y1 RNA, the band intensities for the different YsRNAs slightly changed in dicer knockout poly(I:C) treated cells compared to the wild-type cells.
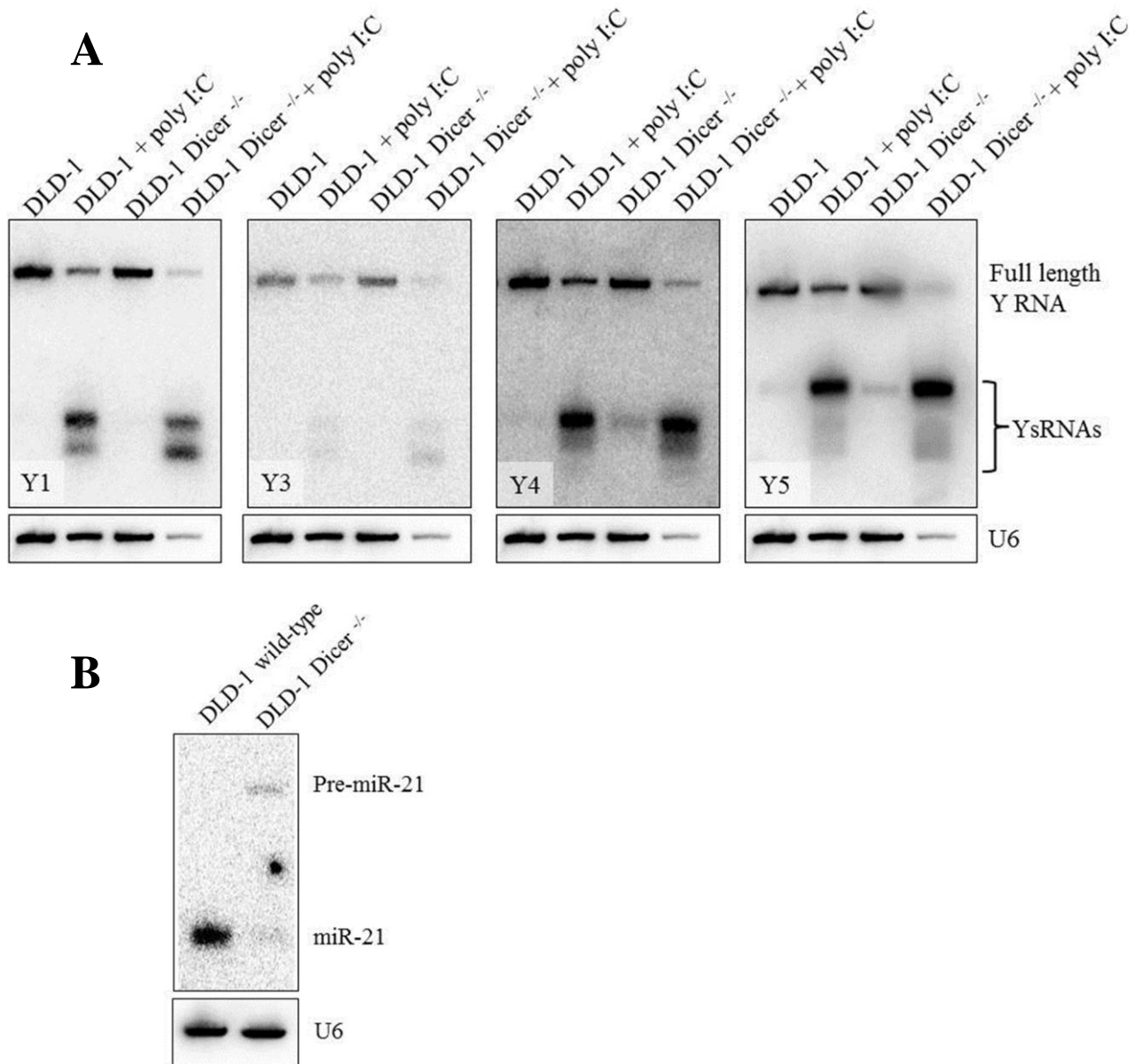
Figure 4.1 YsRNAs are produced in Dicer knockout cells.

(A) Total RNA from wild-type or Dicer$^{-/-}$ DLD-1 (human colorectal adenocarcinoma cells) either treated or not treated with the immunostimulant poly(I:C) was analysed by Northern blot. Oligonucleotide probes antisense to each of the four YsRNAs (3' end) were hybridised to the same Northern blot membrane in separate hybridisation reactions. Bands representing YsRNAs were generated following poly(I:C) treatment in both wild-type and Dicer $^{-/-}$ cells. (B) Genotyping of DLD-1 wild-type and Dicer$^{-/-}$ cell lines via Northern blot. An oligonucleotide probe for microRNA-21 was used. MiR-21 was present in the wild-type cell line, but only the precursor miRNA was present in Dicer knockout cells. The membrane was stripped and re-probed for U6 RNA to show equal loading.

4.2.2 YsRNAs and microRNAs are in different protein complexes

MicroRNAs bind to Argonaute proteins within the RISC complex. Therefore, it was investigated whether YsRNAs are present in the same complex as miRNAs. Protein lysates from poly(I:C) treated cells were prepared, and then separated according to their net charge by anion exchange chromatography (chromatography done by Dr Tibor Csorba). Fractions were eluted at increasing salt concentrations and RNA was extracted from every second fraction. RNA derived from these fractions was then analysed by Northern blot (Figure 4.2). First, antisense oligonucleotide probes specific to the 3' end of all 4 Y RNAs were used to analyse the RNA content of the different fractions. The full length Y RNAs were mainly detected in fractions 63-69 for Y1, Y3 and Y4 RNAs. In the case of Y5 RNA, the full length molecule was eluted earlier, mainly in fractions 41-43. The YsRNAs from all Y RNAs appeared to reside mainly in fractions 39-49. Interestingly, the 32 nt Y5sRNA and the 24 nt Y5sRNA appeared in different fractions: the 32 nt Y5sRNA mainly in 39-41 and the 24 nt Y5sRNA in fractions 45-49 (Figure 4.2).

The RNA contained within the different protein fractions was also hybridised with a probe against miR-21. MiR-21 appeared to be eluted at a very low salt concentration (100mM), mainly in fractions 11-15 (Figure 4.2). Therefore YsRNAs and microRNAs are in different complexes. However, it was noted that a small amount of miR-21 and a small amount of both full length and fragmeted Y RNAs appeared together in fractions 67-69.
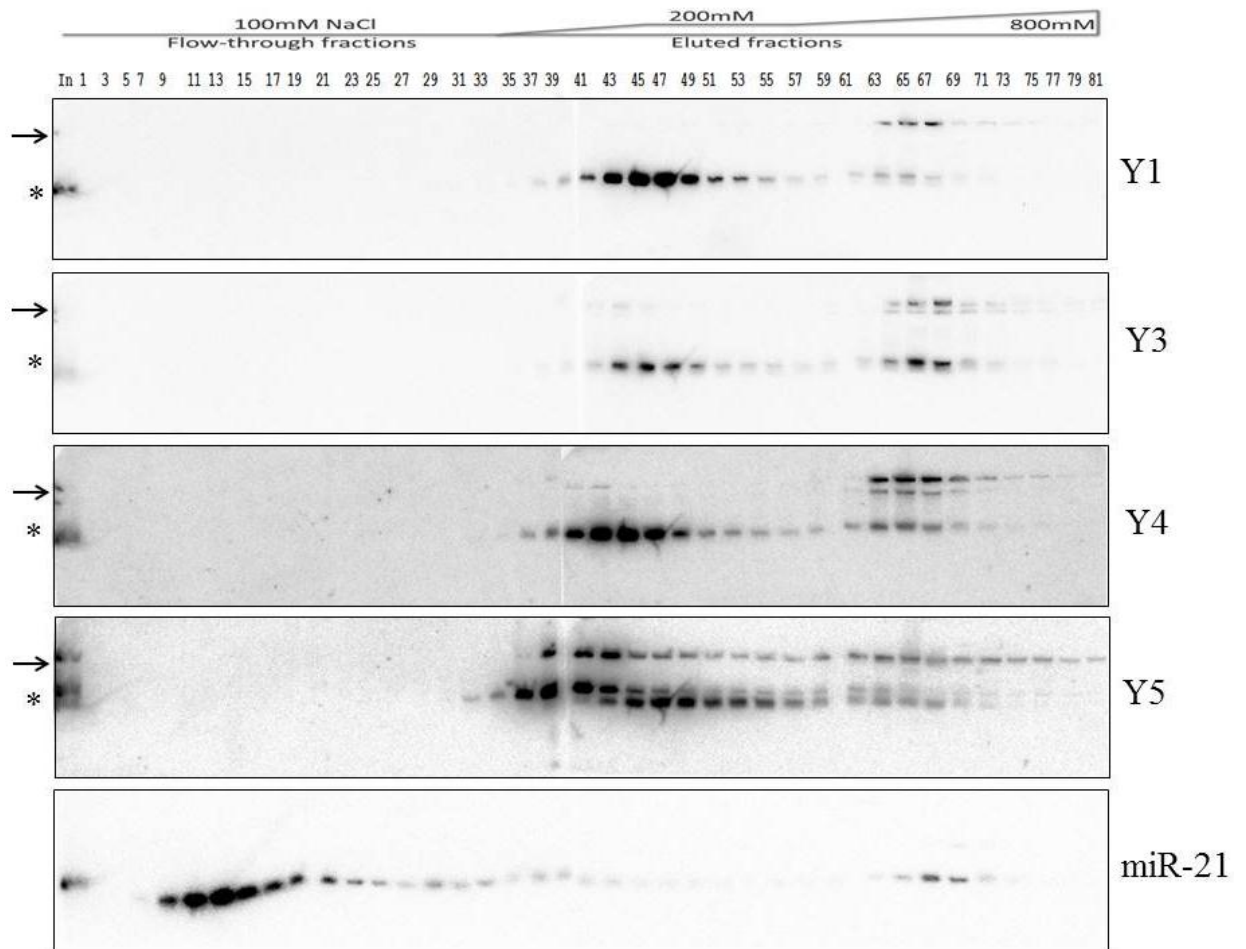
Figure 4.2 YsRNAs and microRNAs are in different complexes.

Protein lysates of poly(I:C) treated MCF7 cells were separated by anion exchange chromatography. Fractions were eluted with increasing salt concentration between 100 mM and 800 mM. RNA was extracted from every second fraction and analysed by Northern blot. Membranes were probed in different hybridisation reactions using antisense oligonucleotide probes to Y1, Y3, Y4 and Y5 RNAs (3' end). The membrane was also probed for miR-21. MiR-21 is present mainly in fractions 11-15. The YsRNAs are all mainly present in fractions 39-49. Interestingly for Y5sRNA, the longer 32 nt and shorter 24 nt fragments are in different fractions. Ro60 protein levels in the respective fractions were not analysed in this particular experiment for technical reasons. Arrows indicate full length Y RNAs, asterisks indicate YsRNAs.

4.2.3 YsRNAs are not generated by or associated with the Argonaute 2 silencing protein

In addition to the Dicer knockout and anion exchange chromatography experiments, the Argonaute 2 (Ago2) protein was also investigated in relation to YsRNA biogenesis and function. Ago2 is a key component of the microRNA guided RNA-induced silencing complex (miRISC) and some pre-miRNAs can also be processed into mature microRNAs by Ago2 instead of Dicer (Cheloufi *et al.*, 2010).

Firstly, Ago2 was depleted via siRNA in MCF7 cells, and the cells were then subjected to poly(I:C) treatment (Figure 4.3). The level of miR-21 in poly(I:C) treated cells which had been transfected with an siRNA against Ago2 was significantly lower compared to cells treated with poly(I:C) only, as expected. However, the level of Y5sRNA remained at a high level in both poly(I:C) treated cells with normal or depleted Ago2 levels.

Secondly, another member of the Dalmay laboratory, Dr Francisco Esteban Nicolas conducted an Argonaute 2 (Ago2) immunoprecipitation experiment (Figure 4.4). Total cellular lysate from non-treated and poly(I:C) treated MCF7 cells was analysed by Northern blot and hybridised to oligonucleotide probes complementary to Y3sRNA, Y5sRNA and miR-21. For the poly(I:C) treated sample, both the 32 nt and 24 nt Y5sRNA fragments were detected by the Y5 RNA probe, with a single band being detected by the Y3sRNA and miR-21 probes. Non-treated MCF7 cells only showed a signal for miR-21 and not for either of the YsRNAs.

Following Ago2 immunoprecipitation, miR-21 was pulled-down in both treated and non-treated cells as expected (Figure 4.4). However, neither Y5sRNA nor Y3sRNA were immunoprecipitated by Ago2.
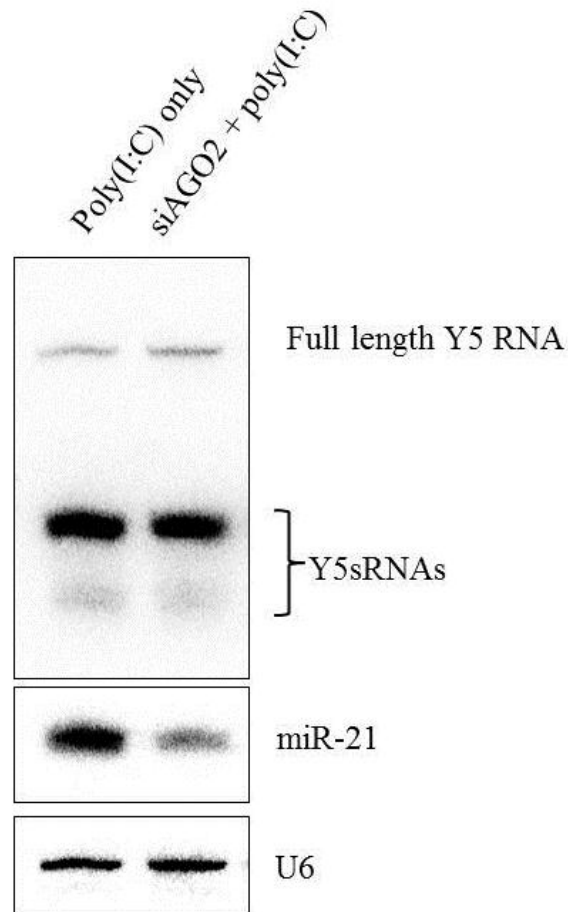
Figure 4.3 YsRNA generation is not affected by depletion of Ago2.

The Ago2 protein was depleted in MCF7 cells via siRNA, followed by treatment of cells with poly(I:C). The abundance of miR-21 was assessed by Northern blot analysis which indicated that the level of the microRNA in cells transfected with the Ago2 siRNA was greatly reduced compared to cells treated with poly(I:C) only, as expected. However, when the same membrane was analysed with a Y5 RNA specific probe, the abundance of Y5sRNA generation appeared unaffected by depletion of Ago2.
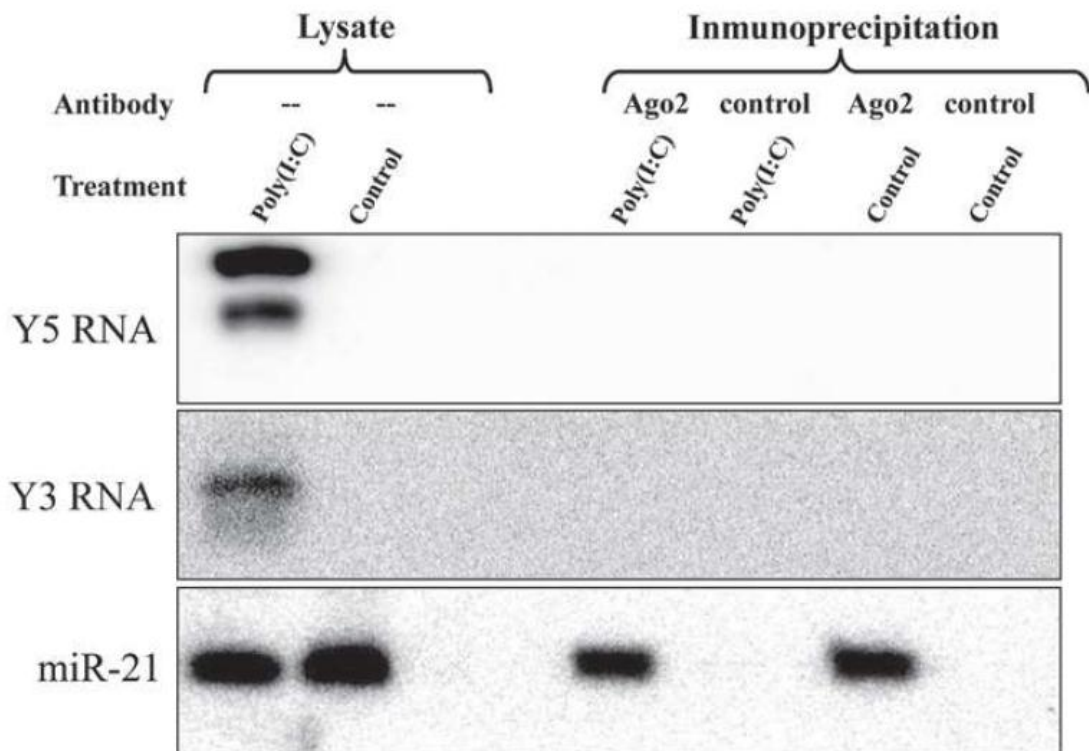
Figure 4.4 YsRNAs are not bound by the Ago2 silencing protein.

Northern blot analysis of lysates from poly(I:C) treated and non-treated cells before and after immunoprecipitation by Ago2. Lysate samples shows that miR-21 was present in both samples, whereas the YsRNAs were only present in poly(I:C) treated samples. Following Ago2 immunoprecipitation, miR-21 was pulled down in both treated and non-treated cells. However, YsRNAs were not immunoprecipitated under either condition. This shows that YsRNAs are not present in the miRISC complex. Work by Dr Francisco Esteban Nicolas.

4.3 Discussion

Canonical microRNAs are generated from precursor stem-loop containing molecules (pre-miRs) and are processed into mature ~ 22nt RNAs by Dicer. These mature miRNAs are then loaded into the gene silencing complex (RISC) which contains one of four Argonaute effector proteins in mammals (Ago1-4) (Filipowicz *et al.*, 2008). Although Agos 1-4 can all associate with microRNAs, only Ago2 is catalytically active (Meister *et al.*, 2004). The precise roles of the other Ago proteins are not as clear. For example, Ago1 has been shown to associate with Dicer (Okamura *et al.*, 2004), indicating a role in miRNA biogenesis, but Ago1 has also been implicated in heterochromatin silencing (Janowski *et al.*, 2006).

Long non-coding RNAs unrelated to pre-miRs have been shown to be processed into fragments which then enter the microRNA pathway (Ender *et al.*, 2008; Cole *et al.*, 2009). Additionally, it has been proposed that fragments derived from the hairpin-containing Y RNAs could potentially enter the microRNA pathway (Meiri *et al.*, 2010; Verhagen & Pruijn, 2011). In this chapter, this hypothesis was investigated.

Here, it was shown that YsRNAs are not generated by Dicer (Figure 4.1). Intriguingly, processing of the full length Y RNA 1-4 appears to be more pronounced in poly(I:C) treated cells without Dicer, and leads to greater YsRNA production (for example see Y5sRNA in Figure 4.1, lane 4). This might be explained by the fact that Dicer can process some of the poly(I:C), since it mimics dsRNA, thereby reducing the amount of poly(I:C) available to cause cell stress. In Dicer knockout cells, this would result in a higher poly(I:C) concentration,

resulting in a stronger stress response. However, this hypothesis does not explain the altered abundance between the different YsRNA products for Y1sRNA between wild-type and dicer$^{-/-}$ poly(I:C) treated cells (see Figure 4.1, Y1 RNA lanes 2 and 4).

Some microRNAs can be generated independent of Dicer through direct Ago2 processing of pre-microRNAs (Cheloufi *et al.*, 2010). Therefore other steps in the microRNA pathway needed to be investigated in order to see whether YsRNAs could act as microRNAs, starting with the protein complex in which the YsRNA resides. Anion exchange chromatography revealed that miRNAs and YsRNAs are mostly in very different complexes (Figure 4.2). However, it was noted that small amounts of miR-21, full length Y RNA and YsRNA appeared together in fractions 67-69. Generally, YsRNAs appear to be in a complex which is much more negatively charged than miRNAs. Intriguingly, YsRNAs from all four Y RNAs do not appear in exactly the same fraction, indicating potentially diverse functions. The Y1sRNAs appear to be principally eluted in fraction 47, (Figure 4.2), Y3sRNA in fraction 45, Y4sRNA in fraction 43 and the longer Y5sRNA in fraction 41. The different protein association is quite striking considering that the stem of the Y RNAs from which the YsRNAs are derived is the most highly similar region between the Y RNAs. Additionally, Y1 and Y3 sRNAs are found in closely related fractions and Y4 and Y5 sRNAs are also in closely related fractions, but distinct to those of Y1/Y3 sRNAs. This might reflect the evolutionary origin of these two pairs of molecules. Y RNAs evolved from a common gene, which through gene duplication, gave rise to Y1/Y3 in one lineage and Y4/Y5 in the other (Mosig *et al.*, 2007).

Depletion of Ago2 in poly(I:C) treated cells by RNAi showed that, although the level of a common microRNA was greatly reduced, YsRNA levels were unaffected (Figure 4.3). This indicates that the slicer activity of Ago2 is not required for generation of YsRNAs from full length Y RNAs and secondly; YsRNAs do not bind to the Ago2 complex. If Ago2 levels are depleted, the amount of microRNA within the cell will be reduced (as shown in Figure 4.3). This could be a result of retardation of the microRNA biogenesis pathway due to inhibition of a key step downstream of Dicer, reducing the amount of microRNA which can be produced. Alternatively, increased turnover and degradation of non-Ago2 protected mature microRNAs could also result in miR depletion. Both scenarios could explain the drop in miR-21 abundance shown in figure 4.3.

In order to investigate whether YsRNAs are actually associated with miRISC, immunoprecipitation experiments were performed. Ago2 immunoprecipitation revealed that YsRNAs do not bind Ago2 (Figure 4.4), whereas miR-21 was shown to be complexed with this silencing protein. However, this experiment did not rule out a possible association of YsRNAs with Ago1, 3 or 4. Nevertheless, taken together with the Dicer knockout and anion exchange chromatography experiments, as well as the Luciferase assay results from Meiri *et al*, these data demonstrate that YsRNAs are not microRNAs. The results from this chapter have subsequently been published by our laboratory (Nicolas *et al.*, 2012).

Interestingly, preliminary data by Verhagen and Pruijn showed that neither Ro60 nor La could be co-immunoprecipitated by anti-hAgo antibodies (unpublished). However, work in Gunter Meister's lab has shown that Ro60 preferentially interacts with Ago2 in Dicer knockout mouse embryonic fibroblasts

(MEFs) (Frohn *et al.*, 2012). These results could explain the mildly increased YsRNA generation in dicer knockout cells, as when Ro60 binds to Ago2 under these artificial conditions, perhaps Ago2 inflicts slicer activity on Ro60-bound Y RNAs. Why Ro60 would bind Ago2 under dicer$^{-/-}$ conditions is unclear, but may point towards a primordial RNA quality control mechanism whereby misfolded RNAs are processed by AGO, directed by Ro60.

Chapter 5

YsRNA generation is dependent on the autoimmune

protein Ro60

5.1   Introduction

Most RNAs within the cell can associate with proteins, forming ribonucleoproteins (RNPs). Y RNAs have been shown to bind to a range of cellular proteins with diverse cellular functions (Hendrick *et al.*, 1981; Bouffard *et al.*, 2000; Fabini *et al.*, 2000; Fabini *et al.*, 2001; Fouraux *et al.*, 2002; Hogg & Collins, 2007; Sim *et al.*, 2012). The Ro60 and La autoimmune proteins are the primary constituents in Y RNPs (Hendrick *et al.*, 1981; Pruijn *et al.*, 1991). Furthermore, the ribosome maturation related protein, nucleolin, has been shown to bind to a subset of the Y RNAs (Y1 and Y3), and potentially functions as a Y RNA nuclear export factor (Fouraux *et al.*, 2002). Unlike Ro60 and La which bind to the lower stem and 3' tail regions of Y RNAs respectively, nucleolin binds to the pyrimidine-rich internal loop.

Long RNA molecules can be processed into shorter RNAs via enzymatic cleavage by proteins called RNases. There are eight canonical, catalytically active RNases in humans (numbered 1-8) (Sorrentino, 2010), but there are also many other proteins which have RNase activity, such as Ago2 - which has a domain similar to ribonuclease H (Song *et al.*, 2004). RNA cleavage can occur at a specific region on the RNA molecule owing to recognition of a sequence or structural element by the enzyme (like with Dicer) or RNA cleavage can occur non-specifically (like with RNase A).

The biogenesis pathway of novel RNA fragments has recently started to be investigated. For example, angiogenin has been uncovered as being the nuclease responsible for the generation of stress-induced tRNA halves (Fu *et al.*, 2009;

Yamasaki *et al.*, 2009). These tRNA fragments then go on to inhibit protein synthesis by targeting the initiation stage of translation (Ivanov *et al.*, 2011). However, it is possible that multiple RNases have a role in generating different functional RNA fragments under various stimuli. For instance, RNase L is an enzyme which is activated downstream of the 2',5'-oligoadenylate pathway (OAS) as a part of the interferon induced antiviral response (Silverman, 2007). RNA degradation is one of the main mechanisms through which this antiviral response is mediated. RNase L cleaves RNA both from the host and invading virus and it is possible that some of the resultant RNA fragments are functional and not merely products of degradation.

Although YsRNAs have been shown to be generated by Dicer independent mechanisms and are not part of the microRNA pathway (see Chapter 4), the proteins involved in the biogenesis of these molecules remains unknown. In this chapter, the mechanism of YsRNA production was investigated experimentally. First, the enzyme angiogenin was investigated as a possible candidate owing to its role in generating tRNA halves. RNA *in vitro* cleavage experiments showed that angiogenin non-specifically degraded the Y RNAs, but specifically cleaved tRNAs to generate the previously reported tRNA halves. The stress induced enzyme RNase L also contributed to YsRNA generation, as determined by RNA interference (RNAi) and RNase L knockout cell line experiments. Y RNA binding proteins were also specifically investigated by RNAi which showed that nucleolin had little effect on YsRNA generation whereas the La protein was a contributory factor in the biogenesis of Y5sRNAs. In order to elucidate the role of Ro60 in YsRNA biogenesis, gene knockdown, knockout and *in vitro* cleavage experiments

showed that Ro60 is essential for YsRNA generation. YsRNA production appears to be due to Ro60 protection of the lower stem region of the full length Y RNAs from a cocktail of stress induced cellular RNases. This was further supported by the failure to detect the pyrimidine-rich internal loop region of Y5 RNA using an LNA probe as this region would be detectable if a specific cleavage reaction took place.

5.2 Results

5.2.1 Angiogenin non-specifically degrades Y RNAs but specifically generates tRNA halves

As angiogenin has been shown to be the RNase involved in generating tRNA halves from full length tRNAs, it was investigated as a possible candidate for generating the YsRNAs. First, siRNAs were used to knockdown angiogenin in MCF7 cells followed by poly(I:C) treatment. RNAi mediated knockdown was also applied to the angiogenin inhibitor (RNH1) in a similar experiment. Total RNA was then analysed by Northern blot (Figure 5.1A). As expected Y5sRNAs were produced in poly(I:C) treated cells. Y5sRNAs were not produced in cells treated with a negative control siRNA. RNAi mediated knockdown of angiogenin (ANG) and RNH1 also failed to result in Y5sRNA generation. When poly(I:C) was transfected 48 h after both siRNA treatments, Y5sRNAs were still produced. Additionally, tRNA halves were produced from tRNA Methionine in all three poly(I:C) treated samples (Figure 5.1A). However, angiogenin could not be detected by Western blot using two different antibodies (Figure 5.1B). Despite failure to detect angiogenin by Western blot, tRNA halve generation can be used as an indicator of a functional angiogenin protein. Therefore, it was concluded that the angiogenin targeting siRNA did not knock down the expression level of angiogenin.

Another approach was taken in order to investigate angiogenin in relation to YsRNA biogenesis. An angiogenin expression construct (pcDNA-ANG) used in

a previous study (Fu *et al.*, 2009) was transfected into MCF7 cells with or without subsequent poly(I:C) treatment (Figure 5.2). YsRNA generation appeared mildly increased in cells transfected with pcDNA-ANG compared to cells transfected with a negative control GFP expression construct (pcDNA-GFP). However, tRNA halves were not generated at sufficient levels in pcDNA-ANG transfected cells to indicate that angiogenin was present. As measurement of tRNA halves is an alternative method of measuring angiogenin activity instead of direct Western blot detection, actively expressing pcDNA-ANG could not be validated.
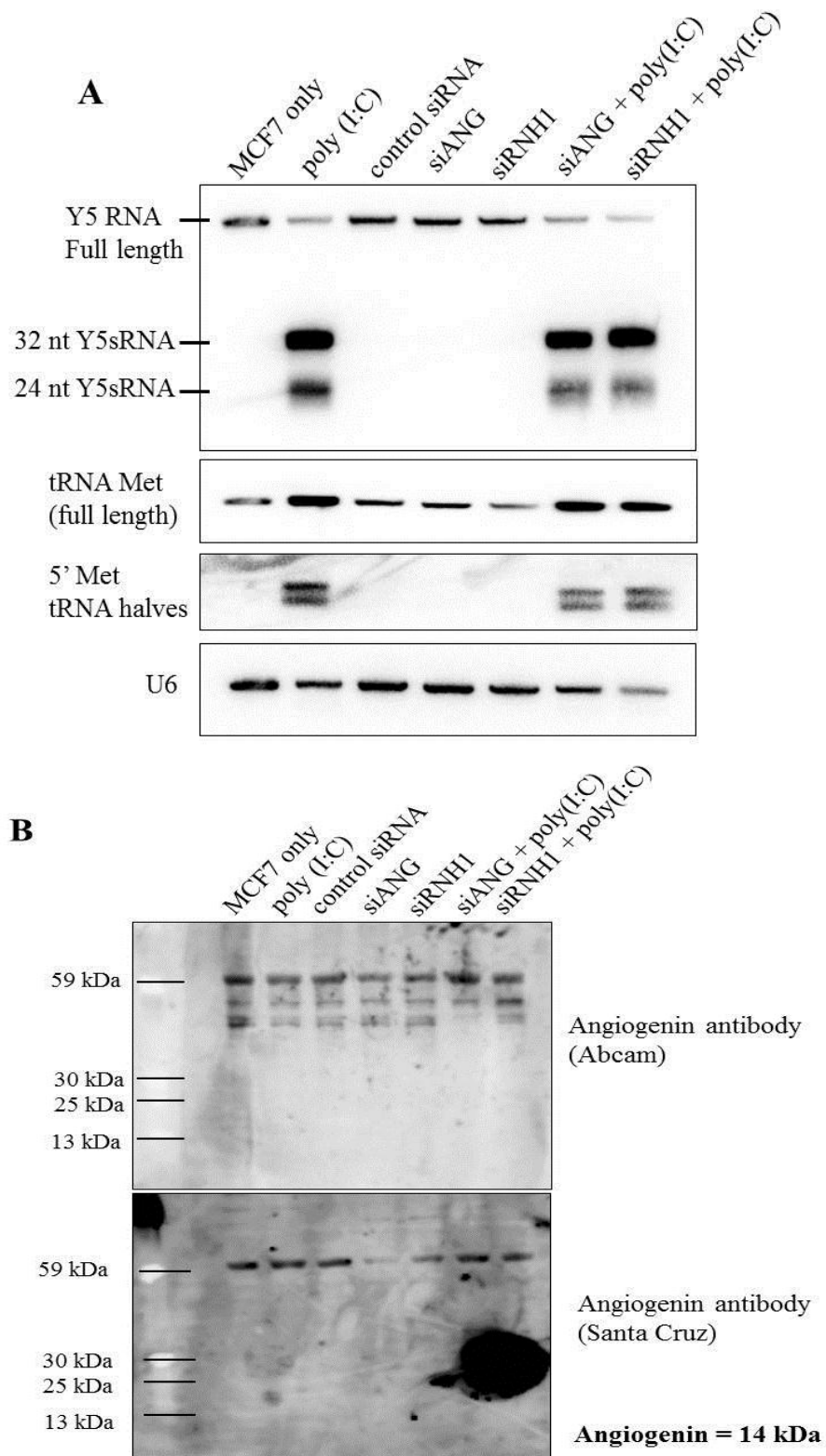
Figure 5.1 Targeting of angiogenin and the angiogenin inhibitor (RNH1) via RNAi.

MCF7 cells were transfected with siRNAs specific to either angiogenin (ANG), RNH1 or with a negative control siRNA. Cells were then either treated or not treated with poly(I:C). (A) Total RNA was analysed by Northern blot. Full length Y5 RNA was detected in all lanes. Y5sRNAs were generated in poly(I:C) cells with or without siRNA knockdown of ANG and RNH1, and the stress induced tRNA halves were also produced in all poly(I:C) treated cells despite attempted knockdown of angiogenin. Membrane was reprobed for U6 RNA to show equal loading. (B) Western blot analysis using two different antibodies raised against angiogenin failed to detect the 14 kDa protein.
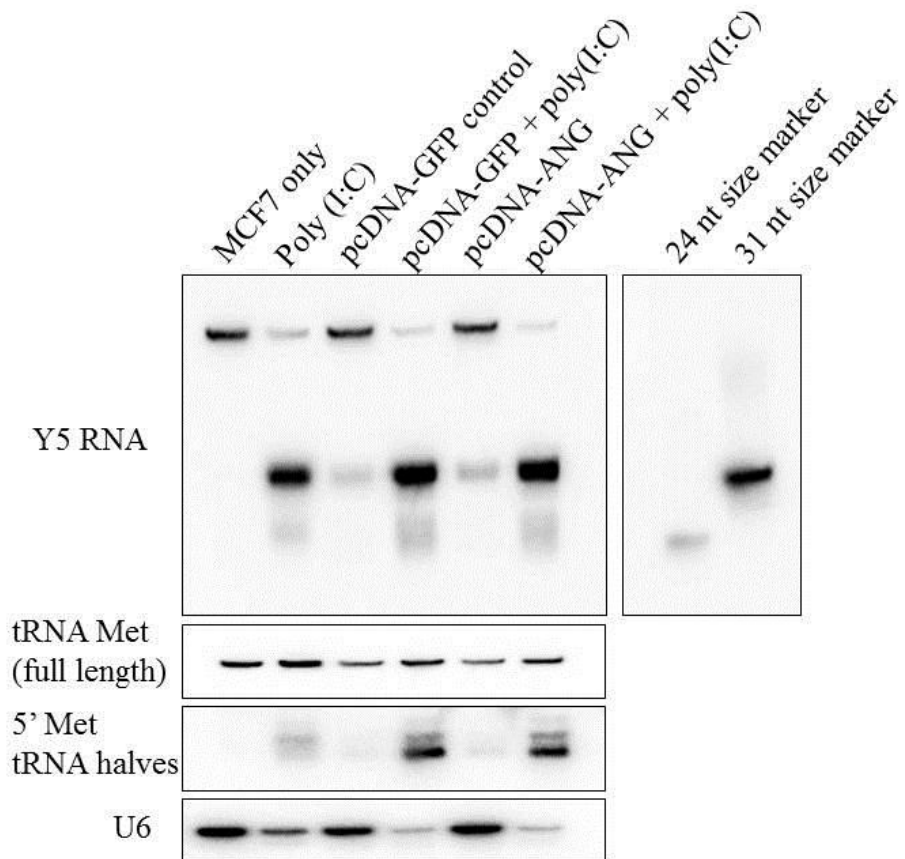
Figure 5.2 Effect of angiogenin overexpression on Y5sRNA generation.

An angiogenin overexpression construct (pcDNA-ANG) was transfected into MCF7 cells which were then either treated or not treated with poly(I:C). A plasmid expressing green fluorescent protein (GFP) was transfected as a negative control. Y5sRNAs were generated following poly(I:C) treatment in MCF7 cells alone, and also in cells transfected with the negative control plasmid. YsRNA generation appeared mildly increased in cells transfected with pcDNA-ANG compared with cells transfected with pcDNA-GFP, however tRNA halves were not generated despite the presence of an angiogenin overexpression plasmid. Cells with pcDNA-ANG which were treated with poly(I:C) did not show a marked increase in YsRNA generation compared to cells treated with poly(I:C) alone or with the negative control plasmid.

As angiogenin could not be detected by Western blot, and owing to the failure to validate expression from transfected pcDNA-ANG, a third and final approach was employed. Recombinant human angiogenin was used directly for RNA *in vitro* cleavage experiments – an approach also used in two recent studies (Fu *et al.*, 2009; Li *et al.*, 2012). Firstly, a pilot experiment was done to see the effect of angiogenin on Y5 RNA in MCF7 cells. Total RNA was incubated either with the reaction buffer only, or with the angiogenin enzyme for 10 min and 30 min at 37 °C (Figure 5.3). RNA was then analysed by Northern blot. The full length Y5 RNA appeared to be targeted during angiogenin incubation, but so was U6 RNA and ribosomal RNA. However, bands representing Y5sRNAs did not show a high production level during angiogenin incubation comparable to those usually generated following poly(I:C) treatment. Although no size-marker was used for this pilot experiment, 2-4 extra faint bands were detected using a Y5 probe for both angiogenin incubations (Figure 5.3). When the same membrane was analysed for tRNA Methionine, specific tRNA halves were generated for both angiogenin incubation time points, but were absent in both the RNA used for the experiment (input) and the buffer only incubation.

In order to investigate the role of angiogenin on Y RNA fragmentation further, a more comprehensive study was performed. Both purified total RNA and total cell lysate (with RNA and bound proteins) were analysed in separate experiments for MCF7, incubating with angiogenin or RNase A (control RNase). Preliminary experiments found that if the same final concentration of RNase A was used as angiogenin (1 μM), all RNA was degraded in the sample rapidly at all time-points (data not shown). Therefore, an RNase A serial dilution experiment

was used to determine an appropriate concentration of RNase A which could be

used to have comparable effects on total RNA as angiogenin (Figure 5.4).
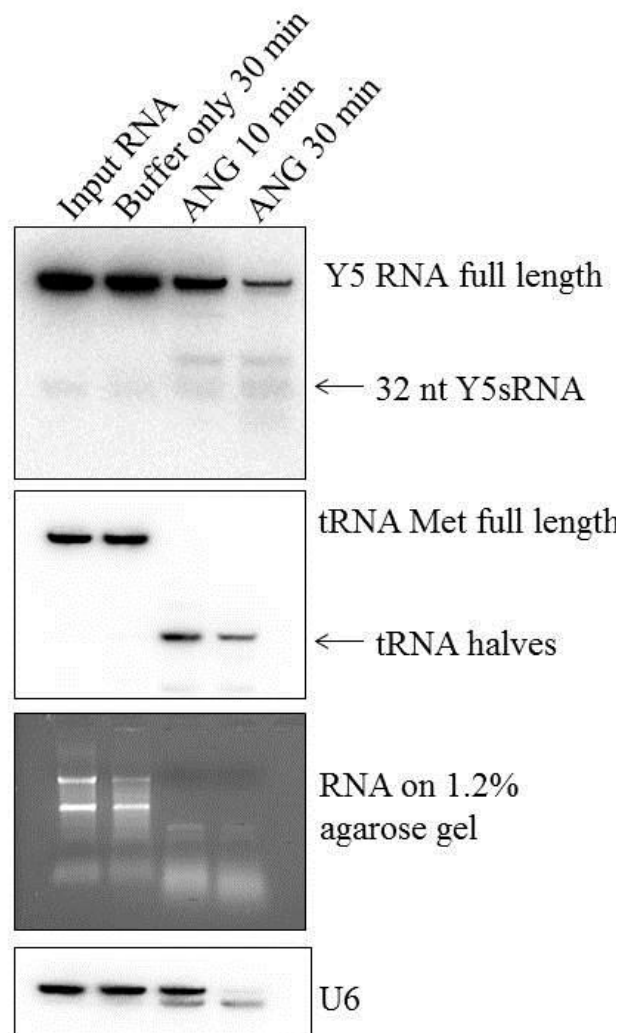
Figure 5.3 Angiogenin *in vitro* cleavage of MCF7 RNA (pilot experiment).

Total RNA was incubated with reaction buffer only, or recombinant human angiogenin in reaction buffer for 10 min or 30 min at 37 °C. RNA was then analysed by Northern blot. Input RNA is the RNA used for the *in vitro* experiment. Y5 RNA showed an overall reduction in abundance following angiogenin (ANG) incubation and the level of full length Y5 RNA decreased with ANG incubation time compared with incubation with reaction buffer only. No significant signal was detected for bands corresponding to YsRNAs. However, when the same membrane was analysed for tRNA-Met, a strong specific signal was detected for tRNA halves in both angiogenin treated samples. RNA was also analysed on an agarose gel to monitor the effect on ribosomal RNA, which showed degradation following ANG incubation. The U6 loading control also showed degradation following extended (30 min) incubation with ANG.
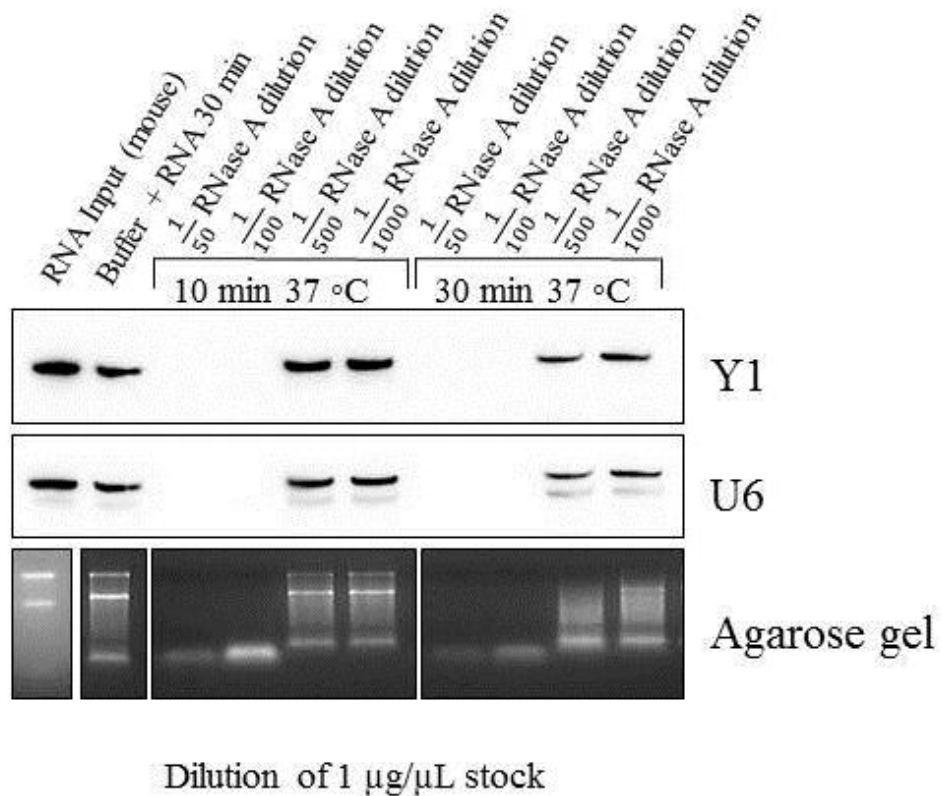
Figure 5.4 RNase A dilution series and effect on RNA degradation.

Mouse total RNA was incubated with varying dilutions of RNase A from a 1 µg/µL stock (73 µM stock). Reactions were incubated in buffer at two different time points. RNA was then analysed by Northern blot and agarose gel. Full length Y1 RNA appeared to be completely degraded in the $\frac{1}{50}$ and $\frac{1}{100}$ RNase A dilutions, but showed only moderate degradation at $\frac{1}{500}$ (comparing $\frac{1}{500}$ RNase A for 30 min with the RNA incubated with buffer only). This degradation pattern was also similar for U6 RNA and also for ribosomal RNA (when analysed on an agarose gel). Therefore $\frac{1}{500}$ dilution RNase A was selected for use in other experiments.

First, MCF7 total RNA was subjected to angiogenin or RNase A *in vitro* cleavage for 2 min, 10 min or 30 min (Figure 5.5). The 1 h time point was omitted as preliminary data showed complete degradation of RNA using both enzymes for this duration (data not shown). Following incubation with both RNases, total RNA was analysed by Northern blot and agarose gel. The full length Y5 RNA showed a mild reduction in abundance in a time-dependent manner with both angiogenin and RNase A (Figure 5.5). The longer Y5sRNA was only detected at a weak level for all reactions, including the input RNA, and did not vary significantly between experimental conditions. However, it was noted that a previously unreported Y5sRNA fragment longer than 32 nt was detectable in 10 min and 30 min angiogenin incubations at a very weak level. When the membrane was analysed for tRNA Methionine, tRNA halves were shown to be specifically generated at high levels in both the 10 min and 30 min angiogenin treated samples, but not in RNase A treated samples. Analysis of ribosomal RNA (rRNA) on an agarose gel showed that rRNA was significantly degraded in all angiogenin incubations and to a lesser extent by RNase A. As no change was seen in Y5sRNA abundance across the time-course, this result suggests that the 32 and 24 nt Y5sRNAs are not generated by angiogenin, but angiogenin may specifically generate a longer than 32 nt fragment from the Y5 RNA with a very low efficiency.

In order to investigate the effect of native RNA binding proteins on YsRNA biogenesis, total cell lysates were used for *in vitro* cleavage reactions. Cell lysate from MCF7 cells were treated with angiogenin and RNase A (Figure 5.6). The full length Y5 RNA was detected in all samples, and showed a reduction in abundance after 1 h treatment with either ANG or RNase A. The Y5sRNA appeared to be

increased slightly in all 1 h incubations using buffer alone, angiogenin or RNase A. The tRNA halves were specifically generated at high levels in both the 30 min and 1 h angiogenin incubation reactions, but not with RNase A. This result suggested that neither angiogenin nor RNase A could specifically generate the YsRNAs but that something in the lysate could produce them instead.
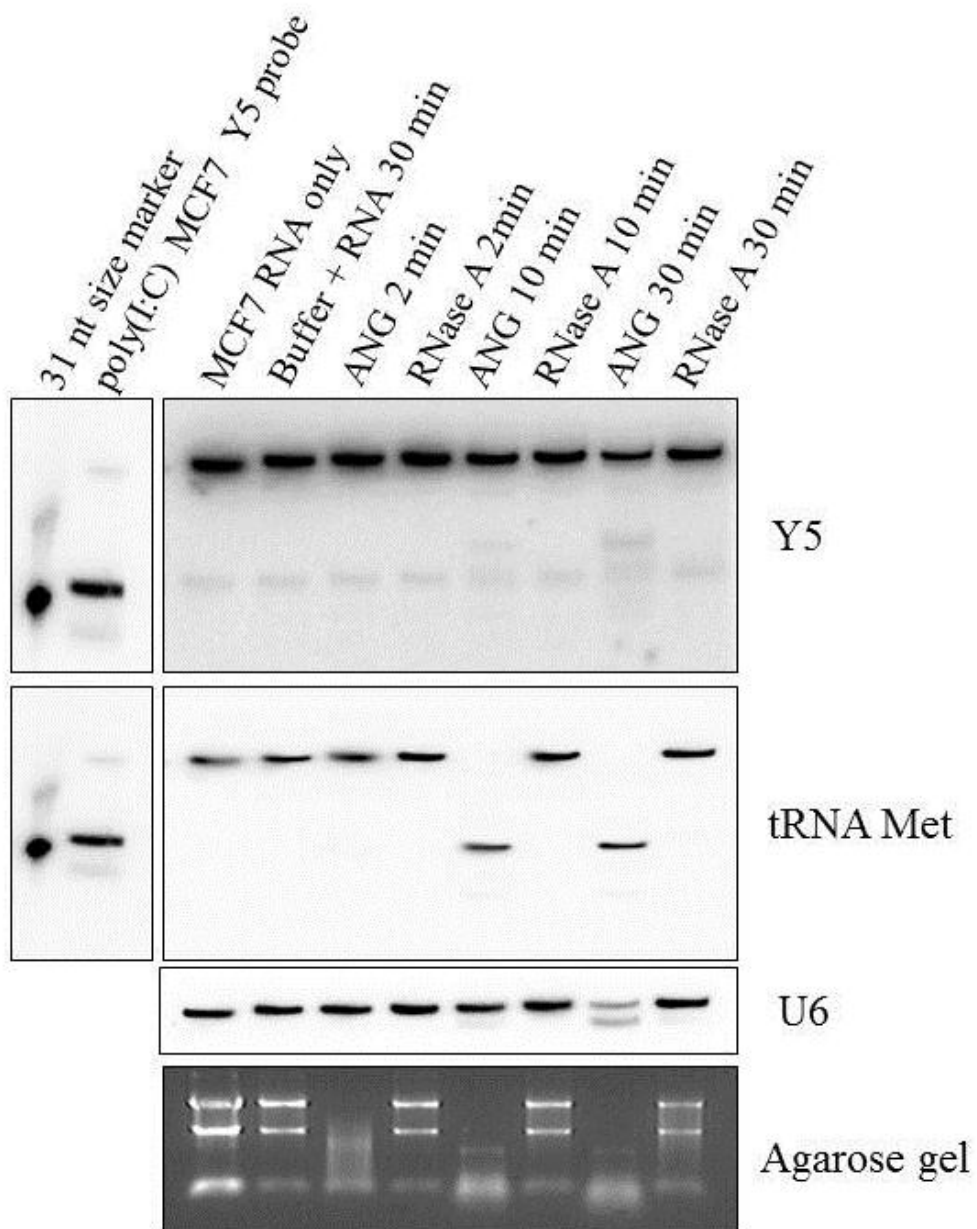
Figure 5.5 MCF7 total RNA subjected to angiogenin and RNase A *in vitro* cleavage.

Following incubation of total RNA with recombinant human angiogenin (ANG) or RNase A, analysis by Northern blot showed that overall abundance of full length Y5 RNA decreased slightly in a time-dependent manner with both enzymes. The natural endogenous 32 nt Y5sRNA was detectable at a weak level in all samples, including the input RNA used for the reactions, and did not increase with ANG incubation. However, a previously unreported Y5sRNA fragment which was longer than 32 nt was detected at weak levels in both 10 min and 30 min ANG incubations. Comparing with tRNA cleavage, tRNA halves from tRNA Methionine were specifically generated at high levels in ANG treated samples at both 10 and 30 min time points.
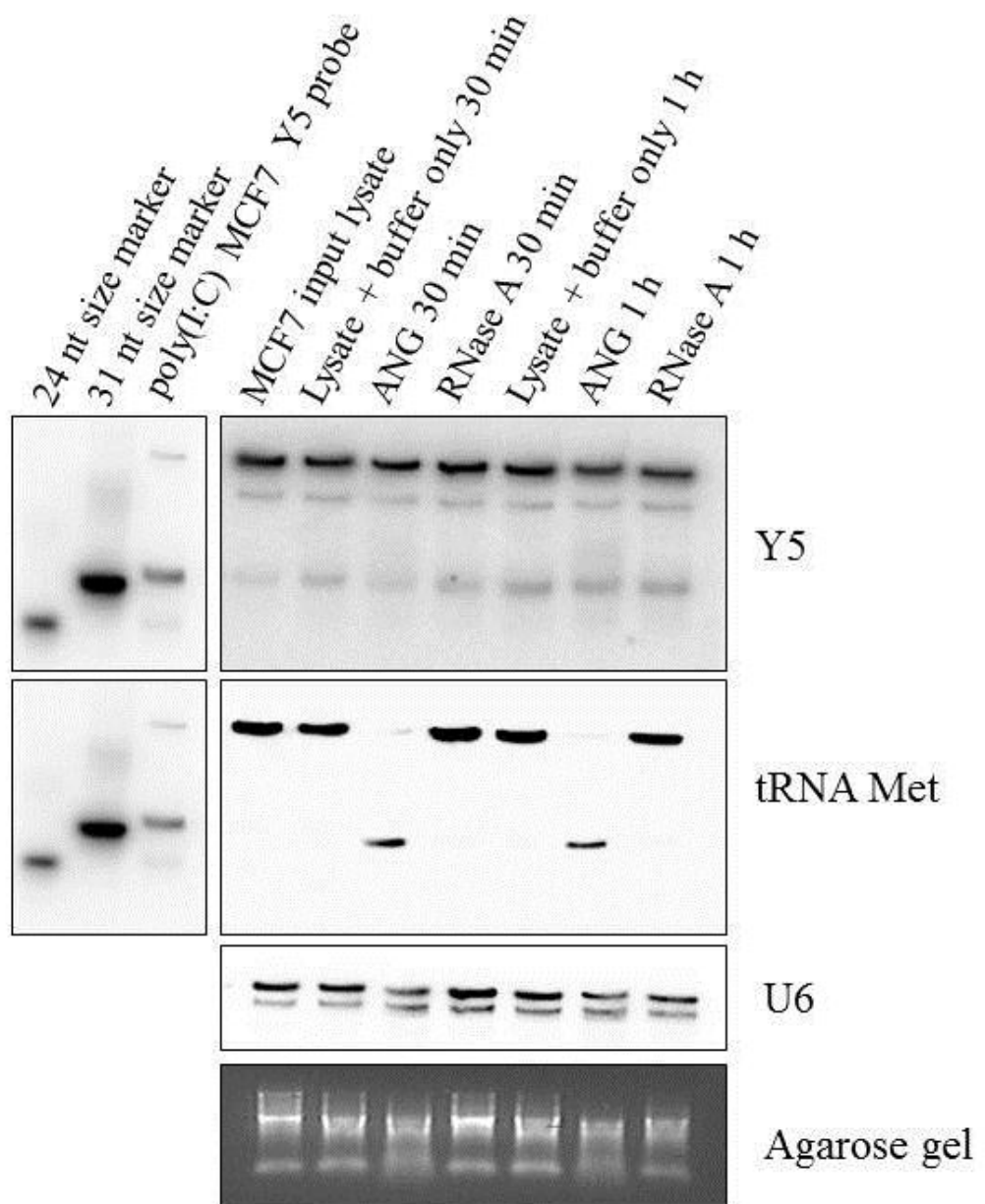
Figure 5.6 MCF7 cell lysate subjected to angiogenin and RNase A *in vitro* cleavage.

Total cell lysate was prepared from MCF7 cells and incubated with either recombinant human angiogenin or RNase A in reaction buffer. Northern blot analysis showed that full length Y5 RNA was present in all samples but was at a reduced level in the 1 h ANG and RNase A time point. The 32 nt Y5sRNA increased slightly during 1hr incubation with ANG, but this was also the case for RNase A treatment, and with incubation with reaction buffer alone for 1h. The tRNA halves from tRNA Methionine showed specific generation at a high level for both ANG 30 min and 1 h incubations, but not for any RNase A incubations. U6 showed slight degradation in all reactions, and so did rRNA (as demonstrated by agarose gel analysis).

5.2.2 Ro60 is essential for YsRNA biogenesis

Considering the integral role of Ro60 and La in forming the Ro-ribonucleoprotein complex, these proteins were investigated in relation to Y RNA fragmentation. First, Ro60 was investigated in a pilot experiment using siRNAs against Ro60 (siRo60) (Figure 5.7). MCF7 cells were transfected with Ro60 specific, or Ago2 specific (control) siRNAs prior to poly(I:C) treatment, or treated with poly(I:C) alone. Ago2 knockdown was confirmed by miR-21 levels. Ro60 protein levels were not assessed in this pilot experiment, although siRo60 efficiency was confirmed by Western blot in later experiments (Figures 5.8-5.10). Northern blot analysis showed that full length Y5 was present at a constant level in all three poly(I:C) treated samples. However, in cells where an siRNA against Ro60 was transfected prior to poly(I:C) treatment, Y5sRNA levels were greatly reduced compared to poly(I:C) treated cells alone or where a siRNA against Ago2 was used prior to poly(I:C) treatment (Figure 5.7). This suggested that Ro60 affected YsRNA biogenesis.

A more detailed investigation was then undertaken with respect to Ro60, and also La. Both Ro60 and La were targeted by siRNA in separate samples prior to poly(I:C) treatment. The experiment was repeated three times (Figures 5.8-5.10). Protein depletion was confirmed by Western blot for both proteins, and all four Y RNAs were analysed by Northern blot. Where siRNAs against La were used prior to poly(I:C) treatment, there was no noticeable change in YsRNA production from Y1, Y3 or Y4. However, in the case of Y5sRNA, the 24 nt fragment was detected at a considerably higher level in La depleted poly(I:C) treated cells compared to

where a negative control siRNA was used prior to poly(I:C) treatment, with a slight reduction in the longer 32 nt fragment. Strikingly, when an siRNA against Ro60 was used prior to poly(I:C) treatment, YsRNAs from all four Y RNAs were drastically reduced in all three repeat experiments compared to where a negative control siRNA, or an siRNA against La was used prior to poly(I:C) treatment.

It was noticed that the amount of full length Y RNA in poly(I:C) treated samples was at a lower level compared to control cells. Therefore, it was possible that the rate of Y RNA cleavage was the same in the absence of Ro60 as in the control cells but that there was simply a reduction in the amount of full length Y RNA available for cleavage. To test this, the intensity of the signals of full length Y RNA and respective YsRNA were quantified in the three independent experiments for all four Y RNAs. Following normalisation using U6 RNA for each lane for each of the three repeat experiments, the ratio of full length Y RNA to the respective principle YsRNA produced in each lane was calculated. This ratio was almost identical in the two poly(I:C) control samples (no siRNA or negative control siRNA) for all Y RNAs: Y1 (9.7 and 9.3 respectively), Y3 (18.0 and 15.0 respectively), Y4 (7.1 and 7.3 respectively) and Y5 (14.6 and 14.5 respectively). However, the ratio of Y RNA full length to YsRNA in Ro60 knockdown poly(I:C) treated samples was significantly higher compared to negative control siRNA poly(I:C) treated samples for all four Y RNAs: Y1 (33.9 ; one-tailed Z-test $p=0.002$), Y3 (43.0 ; one-tailed Z-test $p=0.05$), Y4 (21.2 ; one-tailed Z-test $p=0.03$) and Y5 (20.6 ; one-tailed Z-test $p=0.13$). This raised the possibility that Ro60 had a significant effect on Y RNA cleavage, and not just on the overall level of full

length Y RNA. It was therefore decided to investigate the role of Ro60 in YsRNA biogenesis further.
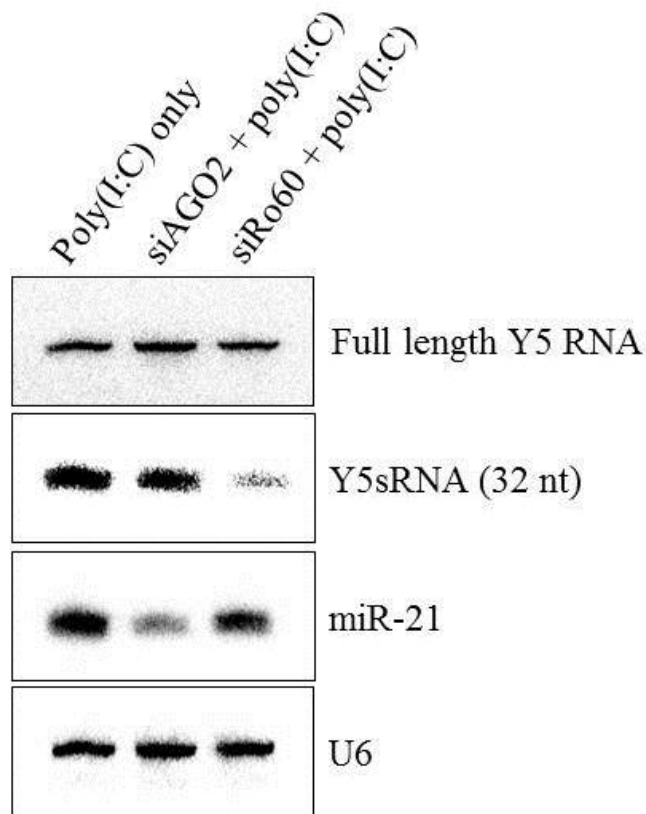
Figure 5.7 Ro60 targeting by siRNA reduces Y5sRNA following poly(I:C) treatment (pilot experiment).

MCF7 cells were treated with siRNAs targeting Ago2 or Ro60 prior to poly(I:C) treatment, or poly(I:C) alone. Northern blot analysis of miR-21 confirmed Ago2 knockdown. Ro60 protein level was not assessed in this pilot experiment, but the Ro60 siRNA (siRo60) efficiency was confirmed by Western blot in later experiments. The Y5sRNA fragment was produced at high levels for poly(I:C) only and siAgo2 poly(I:C) treated cells. However when an siRNA against Ro60 was transfected, poly(I:C) treatment only mildly induced Y5sRNA production.
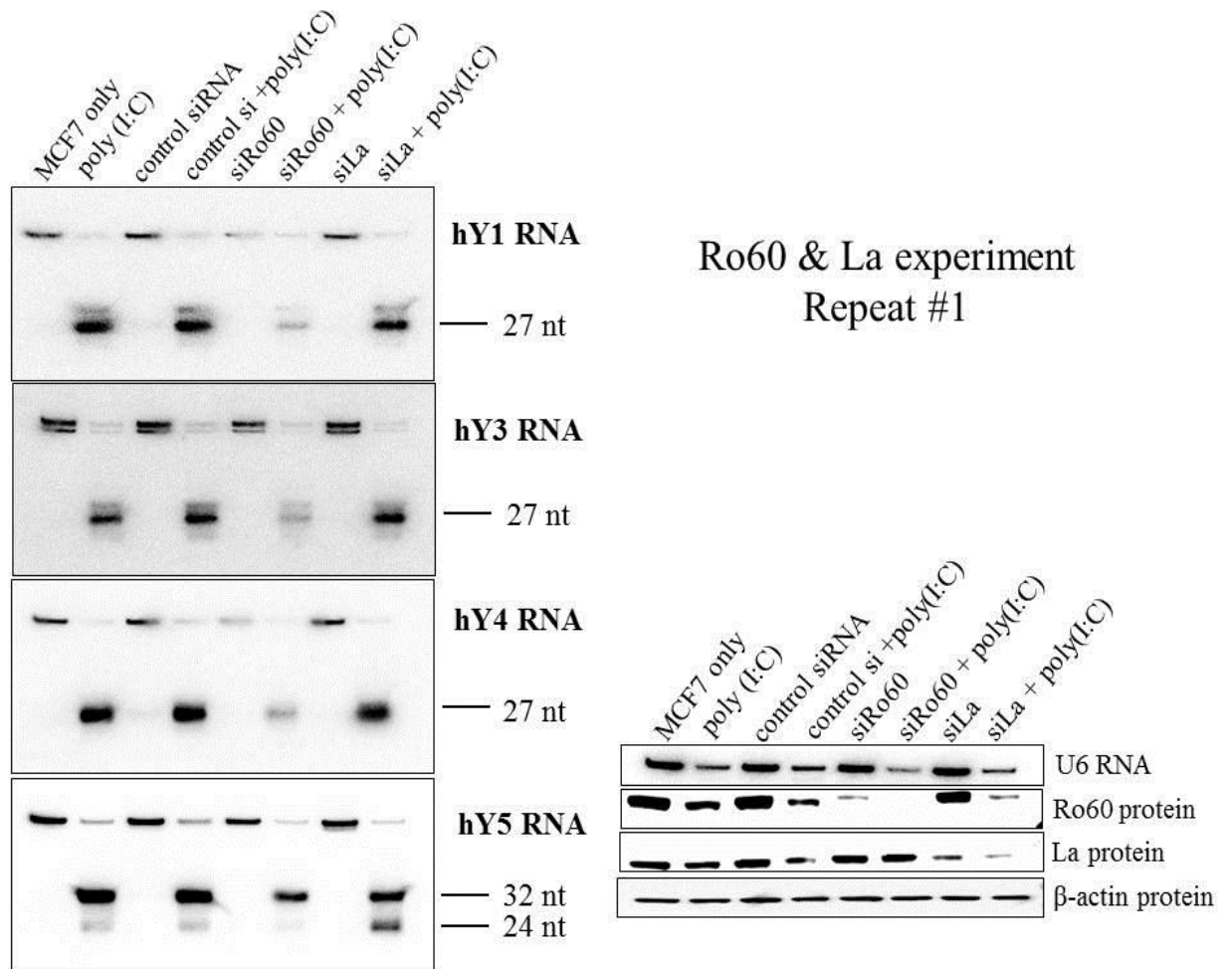
Figure 5.8 Knockdown of Ro60 and La affects the generation of YsRNAs: Repeat number 1.

Expression of Ro60 and La proteins was inhibited by RNAi mediated knockdown using siRNAs siRo60 and siLa respectively. A negative control siRNA was used as a comparison. Accumulation of YsRNAs was measured by Northern blot analysis before and after poly(I:C) treatment, with or without prior siRNA transfection. Use of siLa only affected Y5sRNA production in poly(I:C) treated cells: the 24 nt Y5sRNA fragment was detected at a considerably higher level compared to where a negative control siRNA was used, with a slight reduction in the longer 32 nt Y5sRNA fragment. When Ro60 was knocked down in poly(I:C) treated cells however, YsRNAs from all four Y RNAs were significantly reduced compared to control samples. The Northern blot membrane was stripped and re-probed with a U6 specific probe to show equal loading. The efficiency of the siRNAs was tested by Western blot using Ro60 and La specific antibodies. The Western blot membrane was re-probed with a β-actin antibody to show equal loading.

Figure 5.9 Knockdown of Ro60 and La affects the generation of YsRNAs: Repeat number 2.

Expression of Ro60 and La proteins was inhibited by RNAi mediated knockdown using siRNAs siRo60 and siLa respectively. A negative control siRNA was used as a comparison. Accumulation of YsRNAs was measured by Northern blot analysis before and after poly(I:C) treatment, with or without prior siRNA transfection. Use of siLa only affected Y5sRNA production in poly(I:C) treated cells: the 24 nt Y5sRNA fragment was detected at a considerably higher level compared to where a negative control siRNA was used, with a slight reduction in the longer 32 nt Y5sRNA fragment. When Ro60 was knocked down in poly(I:C) treated cells however, YsRNAs from all four Y RNAs were significantly reduced compared to control samples. The Northern blot membrane was stripped and re-probed with a U6 specific probe to show equal loading. The efficiency of the siRNAs was tested by Western blot using Ro60 and La specific antibodies. The Western blot membrane was re-probed with a β-actin antibody to show equal loading.
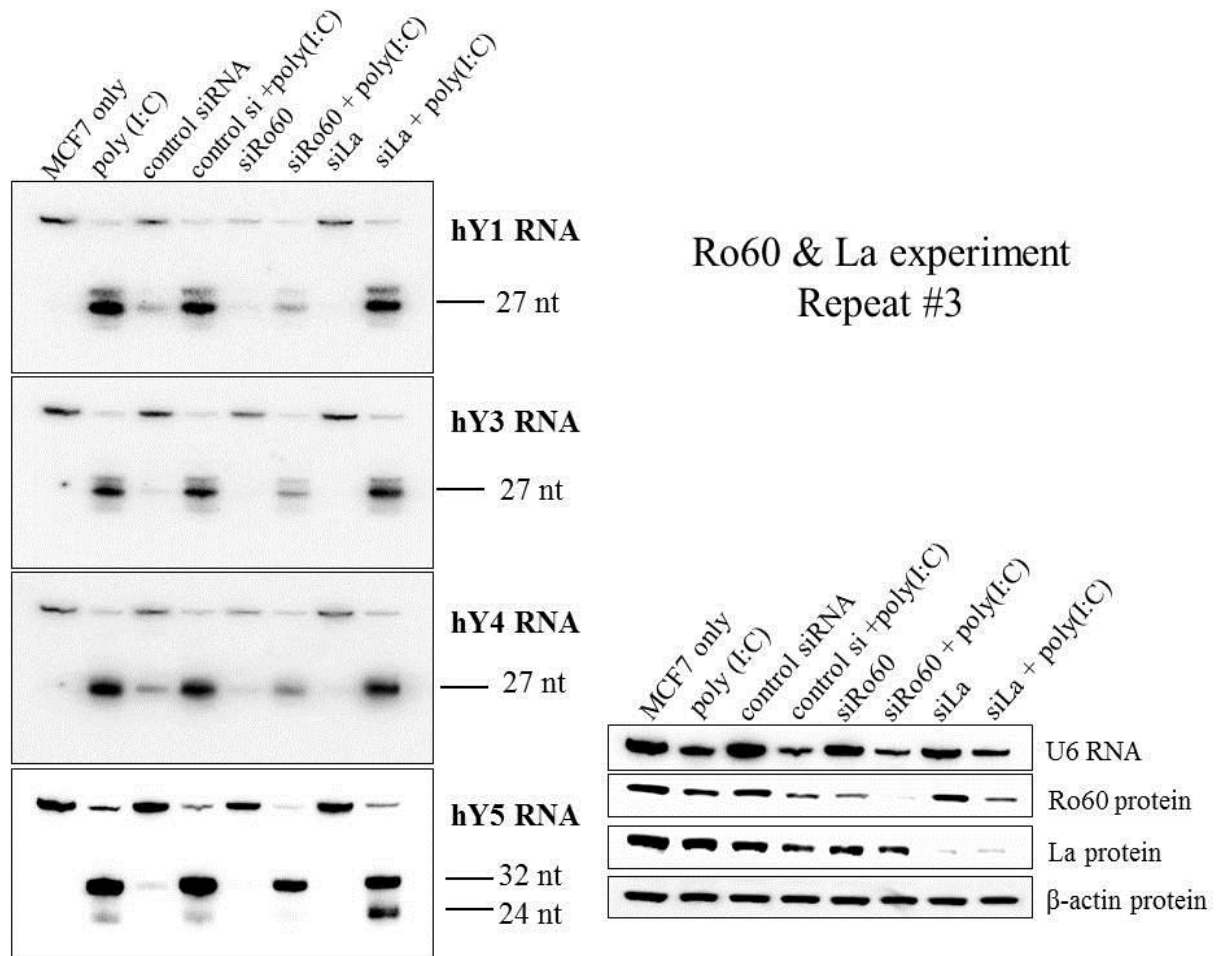
154

Figure 5.10 Knockdown of Ro60 and La affects the generation of YsRNAs: Repeat number 3.

Expression of Ro60 and La proteins was inhibited by RNAi mediated knockdown using siRNAs siRo60 and siLa respectively. A negative control siRNA was used as a comparison. Accumulation of YsRNAs was measured by Northern blot analysis before and after poly(I:C) treatment, with or without prior siRNA transfection. Use of siLa only affected Y5sRNA production in poly(I:C) treated cells: the 24 nt Y5sRNA fragment was detected at a considerably higher level compared to where a negative control siRNA was used, with a slight reduction in the longer 32 nt Y5sRNA fragment. When Ro60 was knocked down in poly(I:C) treated cells however, YsRNAs from all four Y RNAs were significantly reduced compared to control samples. The Northern blot membrane was stripped and re-probed with a U6 specific probe to show equal loading. The efficiency of the siRNAs was tested by Western blot using Ro60 and La specific antibodies. The Western blot membrane was re-probed with a β-actin antibody to show equal loading.

To obtain definitive proof for the involvement of Ro60 in YsRNA generation, a Ro60$^{-/-}$ mouse embryonic stem cell line (mES) was employed (Chen *et al.*, 2003) to investigate how this would affect Y RNA processing. As mice only possess the Y1 and Y3 RNA genes, mES wild-type and Ro60$^{-/-}$ cells were transfected with a Y5 RNA expression construct (pY5) so that Y5sRNA biogenesis could be monitored. The effect of poly(I:C) treatment was then analysed by Northern blot (Figure 5.11). As expected, there was no Y5 RNA present in non-pY5 transfected mES wild-type and Ro60$^{-/-}$ cells either in the presence or absence of poly(I:C), whereas Y1 and Y3 RNAs were expressed. In all cells transfected with the pY5 expression construct, the full length Y5 RNA was detectable by Northern blot but Y5sRNAs were detected only in poly(I:C) treated wild-type and not Ro60$^{-/-}$ cells. Similarly, this was the case for Y1sRNAs, which were only generated in wild-type and not Ro60 knockout cells (Figure 5.11). Y3sRNAs were only detected at weak levels following poly(I:C) treatment in wild-type cells, and not at all in Ro60$^{-/-}$ cells. This demonstrated that the presence of Ro60 is essential for YsRNA biogenesis.
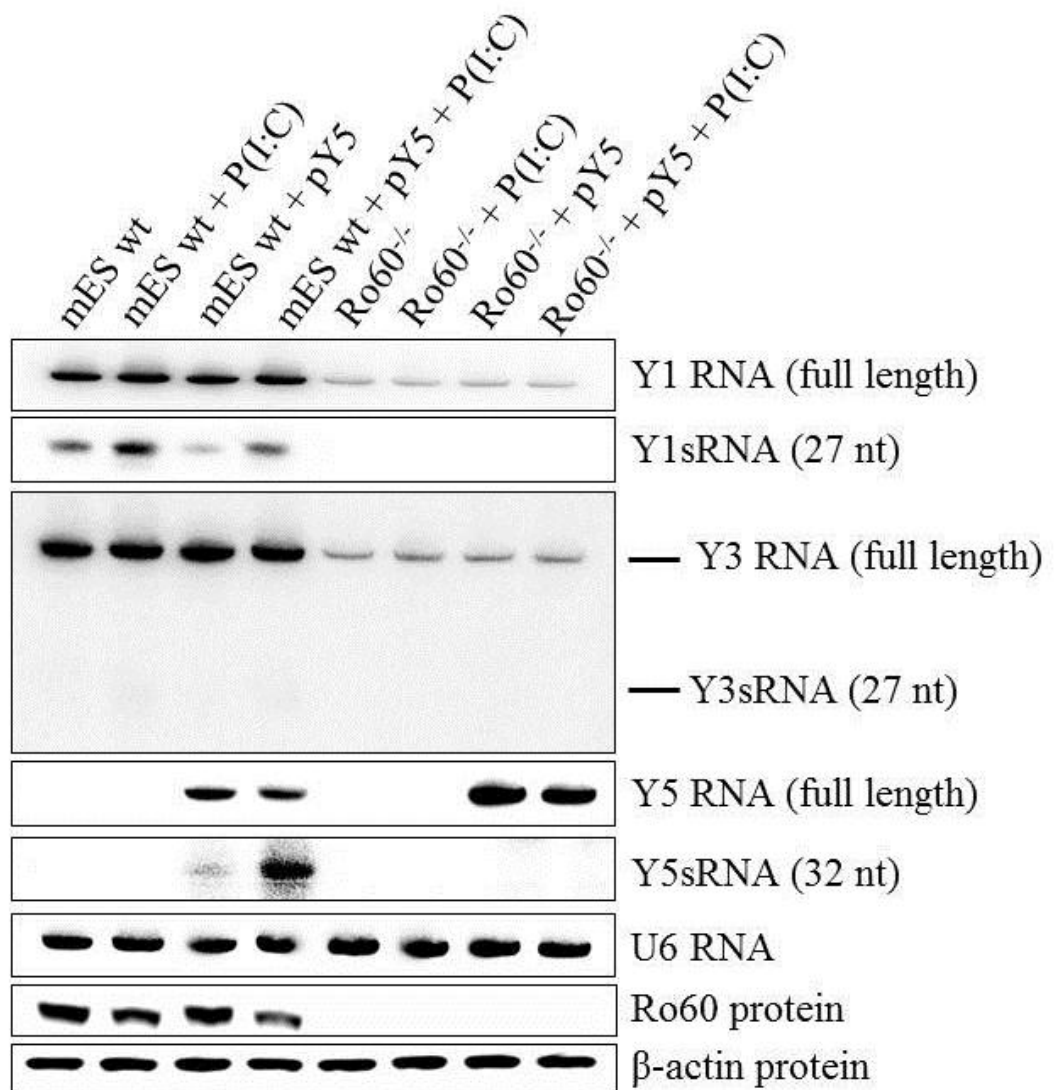
Figure 5.11 YsRNA generation is Ro60 dependent.

Wild-type (wt) and Ro60$^{-/-}$ mouse embryonic stem cells (mES) were transfected with a Y5 RNA expressing plasmid (pY5) and treated with poly(I:C) (P(I:C)). Full length Y5 RNA was detected only in the pY5 transfected cells because mouse lacks the endogenous Y5 gene. Y5sRNA was detected only in the wild-type but not in the Ro60$^{-/-}$ cells. This was also the case for Y1sRNA and Y3sRNAs (although Y3sRNAs were only weakly detectable in mES wt). The Northern blot membrane was stripped and re-probed with a U6 specific probe to show equal loading. The absence of Ro60 in the knockout cells was confirmed by Western blot and a β-actin antibody was used to show equal loading. For Y1 and Y5 RNAs, different panels are shown for the corresponding YsRNAs as different contrast settings were required to show bands clearly to avoid signal saturation from the full length molecule.

In order to investigate whether Ro60-dependent YsRNA production was owing to RNase protection by Ro60, cell lysate *in vitro* cleavage reactions were performed using wild-type mouse embryonic stem cells (mES) and mES Ro60$^{-/-}$ cells. The two cell lines were genotyped by Western blot (Figure 5.12). Lysates from wild-type mES cells were incubated with RNases under the same conditions as already described for MCF7 lysates (Figure 5.13). Y1 RNA was analysed by Northern blot as mouse cells do not have the gene for Y5 RNA. Full length Y1 RNA was detected at a high level in all samples except in incubations with angiogenin for 30 min and 1 h where it was greatly reduced in abundance and almost completely degraded after 1 h. The 27 nt Y1sRNA (see chapter 3 for YsRNA size descriptions) was present at a very low level in the input lysate used for the cleavage experiments, but was at a higher level in all of the lysate incubations (Figure 5.13). The incubation reaction for 1 h with angiogenin showed a very high Y1sRNA production, but was at a similar level between buffer only and RNase A incubations for the same time point. The tRNA halves were generated specifically by angiogenin and not RNase A, but the level of tRNA halves was lower at the 1 h incubation compared to the 30 min time point. Similar to experiments done using MCF7 cell lysates (Figure 5.6), this result suggested that neither angiogenin nor RNase A could specifically generate the YsRNAs but that something in the lysate could produce them instead.
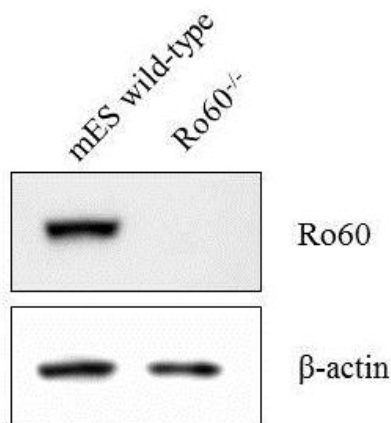
Figure 5.12 Genotyping of wild-type and Ro60 knockout mES cells.

An antibody specific to mouse Ro60 was used to differentiate between the wild type and Ro60 knockout mouse embryonic stem cells (mES) by Western blot.

The same *in vitro* cleavage experiments were then implemented in mES Ro60$^{-/-}$ cells in order to investigate YsRNA generation by RNase protection by Ro60 (Figure 5.14). The full length Y RNAs were detected in all but one lane (ANG 1 h). However, the Y1sRNA (27 nt) which was detectable in wild-type mES cells (Figure 5.13), was not detected in any of the experiments conducted with mES Ro60$^{-/-}$ cells under the same RNase and incubation conditions (Figure 5.14). This result suggested that RNases in the cell lysate do not specifically generate YsRNAs, but generate them via RNase protection owing to Ro60 binding.
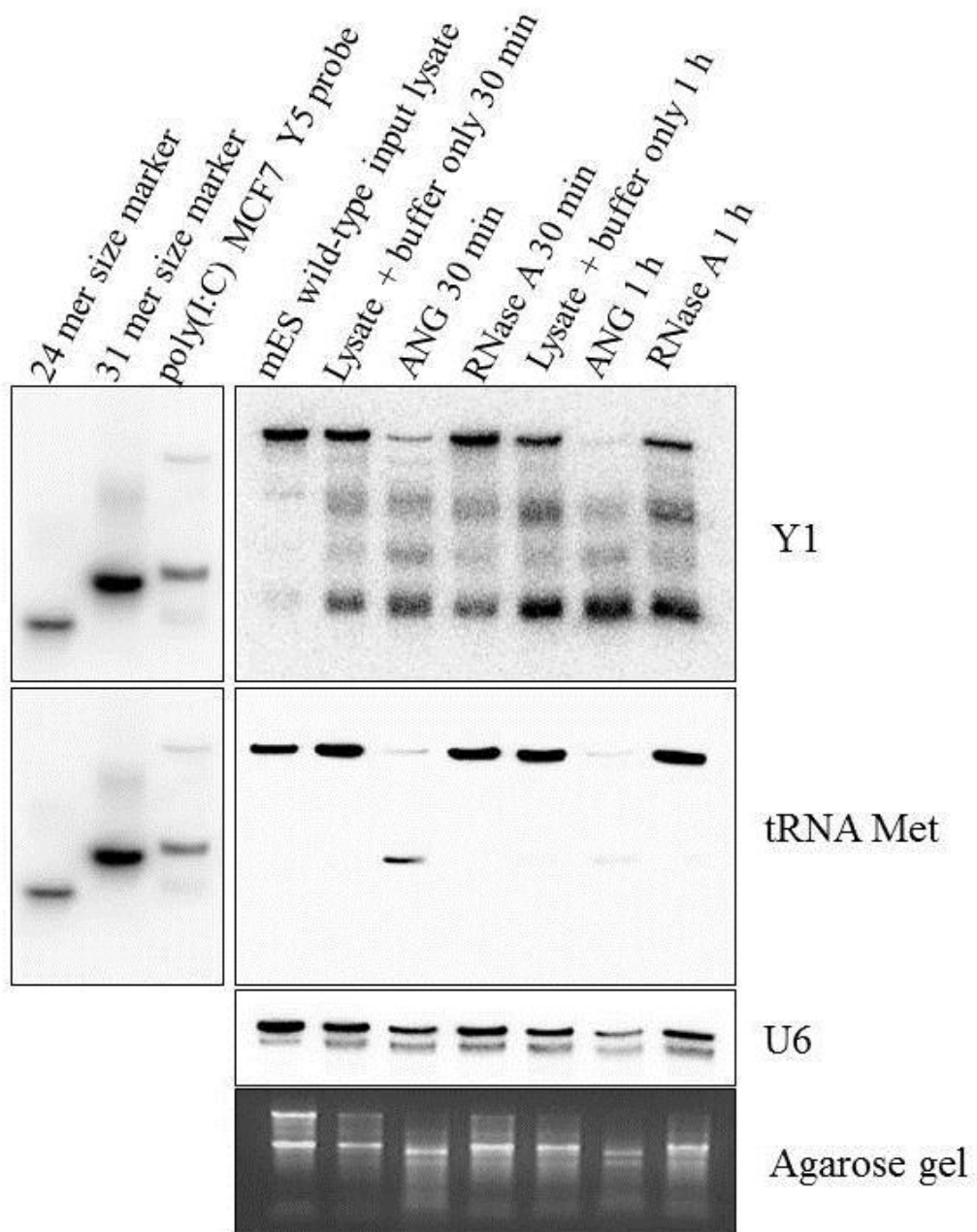
Figure 5.13 Wild-type mouse embryonic stem cell (mES) lysate subjected to angiogenin and RNase A *in vitro* cleavage.

Cell lysates were prepared from wild-type mES cells and incubated with either angiogenin (ANG) or RNase A in reaction buffer. Northern blot analysis showed that full length Y1 RNA was present in most samples, but was at a reduced level in both 30 min and 1 h ANG treated samples. The 27 nt Y1sRNA increased in abundance in reaction samples with increasing incubation time. Y1sRNA levels did not vary between ANG and RNase A treatments for each time point. Generation of tRNA halves from tRNA methionine occurred in ANG 30 min treatment, but reduced at the 1 h time point. RNase A incubation did not give rise to tRNA halves. U6 and rRNA showed signs of mild degradation in all incubations.
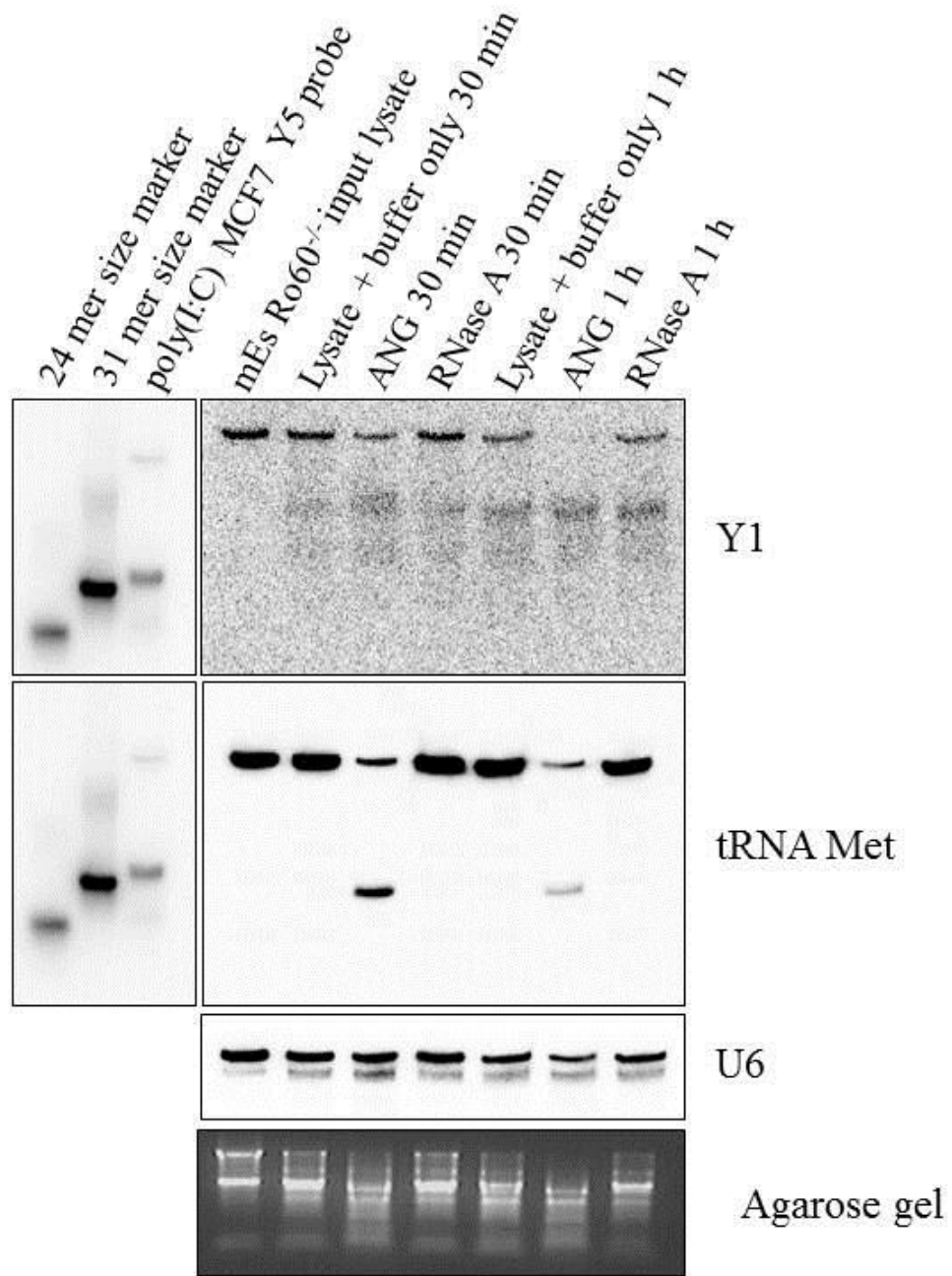
Figure 5.14 Ro60 knockout mouse embryonic stem cell (mES Ro60$^{-/-}$) lysate subjected to angiogenin and RNase A *in vitro* cleavage.

mES Ro60$^{-/-}$ cell lysates were prepared from cells and incubated with either angiogenin (ANG) or RNase A in reaction buffer. RNA was analysed by Northern blot. Full length Y1 RNA was detected in each of the reaction samples, except in the lysate incubated with angiogenin (ANG) for 1 h which showed no signal. No YsRNAs of 24-31 nts were detected in any sample. tRNA halves derived from tRNA methionine were generated in ANG incubated lysates, but not with RNase A. U6 RNA (northern blot) and rRNA (agarose gel) were mildly degraded in all lysates.

5.2.3   RNase L contributes to YsRNA generation

RNase L is an enzyme which is activated as part of the interferon induced antiviral response. As poly(I:C) treatment induces an immune response, RNase L was investigated as a candidate for having a potential role in YsRNA biogenesis. An RNase L knockout, mouse embryonic fibroblast cell line (MEF RNase L$^{-/-}$) was employed which was used in a previous study (Zhou *et al.*, 1997). The wild-type and RNase L$^{-/-}$ cells were genotyped by PCR as described in the materials and methods (Figure 5.15). As expected, the wild-type MEF cells had a 900 bp amplicon generated from the 5' end of the RNase L gene whereas the knockout cells which had an insertion mutation at the 5' end of the gene gave a 2.1 kb amplicon.
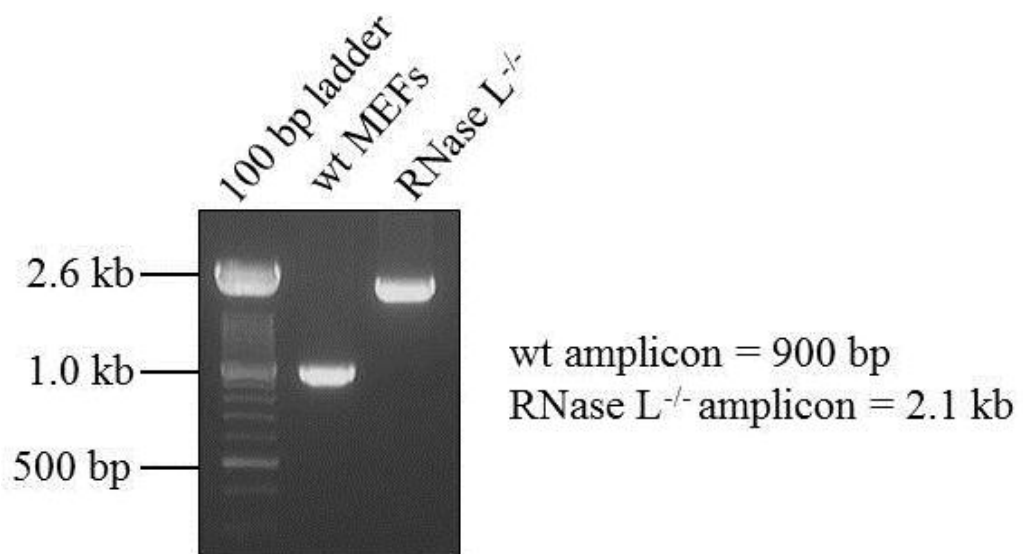
Figure 5.15 PCR genotyping of wild-type and RNase L knockout cell lines.

Mouse embryonic fibroblast (MEF) cells were genotyped by polymerase chain reaction (PCR) amplification of the 5' end of the RNase L gene. Wild-type (wt) cells gave rise to a 900 bp amplicon whereas RNase L$^{-/-}$ cells with an insertion mutation at the 5' end of the gene gave rise to a 2.1 kb amplicon.

Both wild-type and RNase L knockout cells were subjected to poly(I:C) treatment. As mouse cells only have genes for Y1 and Y3 RNA, a Y5 RNA expression construct (pY5) was transfected prior to poly(I:C) treatment so that Y5sRNA biogenesis could be monitored. RNA was then analysed by Northern blot from two independent experiments (Figure 5.16).
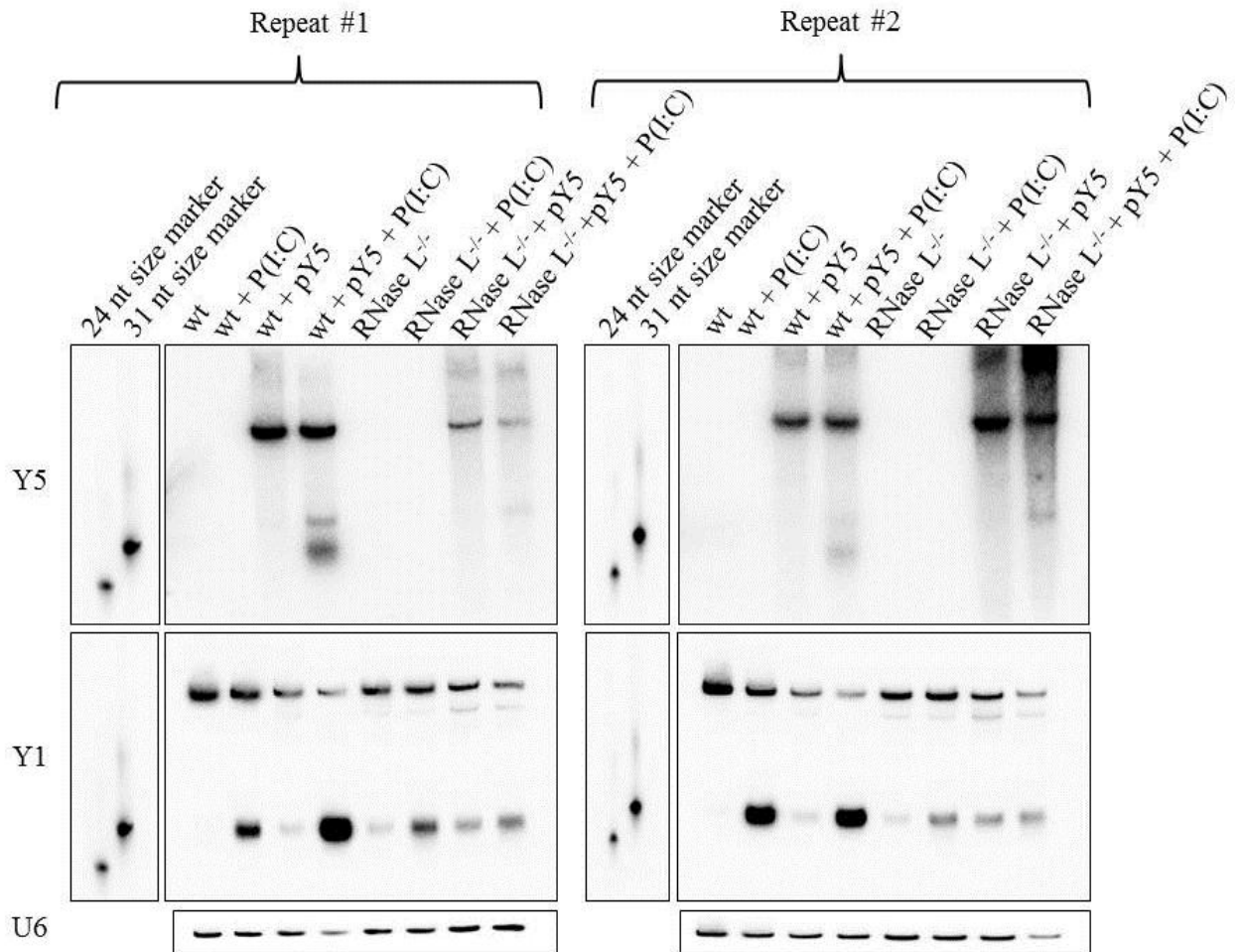
Figure 5.16 Analysis of YsRNA biogenesis in RNase L knockout cells.

Both wild-type (wt) and RNase L$^{-/-}$ MEF cells were subjected to poly(I:C) (P(I:C)) treatment. As mouse cells only have genes for Y1 and Y3 RNA, a Y5 expression construct (pY5) was transfected into a subset of cells prior to stress treatment in order to monitor Y5sRNA biogenesis. Experiments were repeated twice. Northern blot analysis showed Y5 RNA was present in cells transfected with pY5 but not in non-pY5 transfected cells. Poly(I:C) treatment only gave Y5sRNAs in wild-type cells which had been transfected with pY5, but not in cells which lacked pY5. Y5sRNAs were produced at noticeably higher levels in poly(I:C) treated pY5 wild-type cells than in RNase L$^{-/-}$ cells under the same conditions. Y1sRNAs were highly expressed in wild-type poly(I:C) treated cells, but were only very weakly produced in RNase L knockout cells.

The full length Y5 RNA was only present in cells transfected with pY5. Wild-type cells with pY5 gave rise to the expected Y5sRNAs following poly(I:C) treatment (Figure 5.16). However the level of Y5sRNAs was lower in poly(I:C) treated RNase L knockout cells with pY5 for both experimental replicates. As the full length Y5 RNA was lower in the RNase L$^{-/-}$ cells for one of the replicates, Y1 RNA was also analysed. Full length Y1 RNA expression was relatively constant between wild-type and RNase L knockout samples. Y1sRNAs however showed a very high level of production in wild-type MEF cells treated with poly(I:C), but were detected at a much lower level in RNase L$^{-/-}$ poly(I:C) treated cells. This data suggested that RNase L contributed to YsRNA generation in mouse.

YsRNA generation was then investigated in relation to RNase L activity in human cells. As no human RNase L knockout was available, RNase L was knocked down in MCF7 cells using RNAi (Figure 5.17). Intriguingly, poly(I:C) treated cells where RNase L had been depleted by siRNA had similar levels of Y5sRNA and Y1sRNA compared to cells where a negative control siRNA was used. However, the full length Y1 and Y5 RNAs both appeared more stable in RNase L knockout cells following poly(I:C) treatment compared to cells with normal RNase L levels. This suggested that RNase L did not affect YsRNA production in human MCF7 cells as it did in mouse MEF cells, and could therefore indicate either cell line or species specific differences in YsRNA biogenesis.
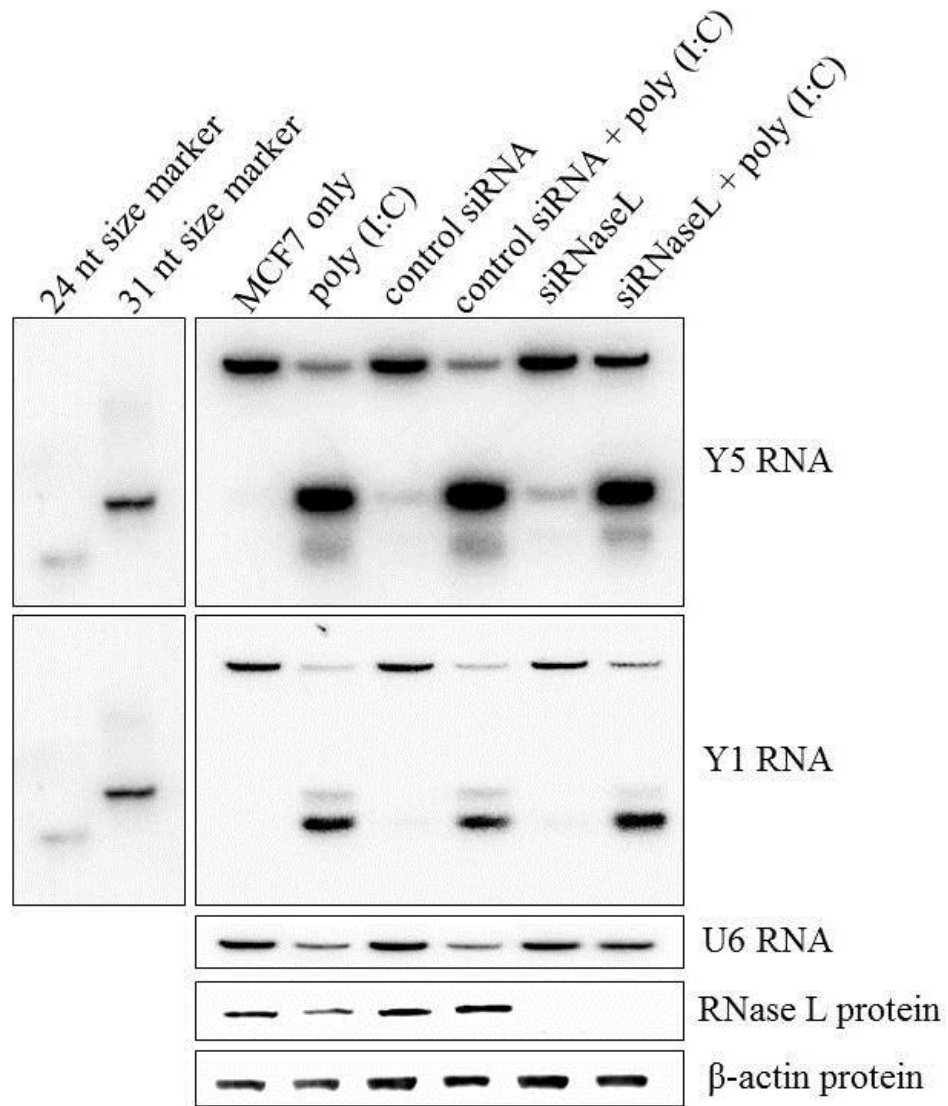
Figure 5.17 YsRNA biogenesis in MCF7 cells following RNase L depletion by RNAi.

MCF7 cells were transfected with either no siRNA, a negative control siRNA or an RNase L specific siRNA. RNase L knockdown was confirmed by Western blot. Cells were then either treated or not treated with poly(I:C) following siRNA transfection. Y1sRNAs and Y5sRNAs were still generated in poly(I:C) cells where RNase L was targeted by RNAi at levels comparable to where only a negative control siRNA was used. Equal loading was confirmed by a U6 specific probe for the Northern blot and a β-actin antibody for the Western blot.

5.2.4   Nucleolin does not affect the biogenesis of YsRNAs

As Ro60 and La appeared to contribute to YsRNA production, experiments then focused on another Y RNA binding protein as a potential candidate for involvement into the maturation of these fragments. Nucleolin has been shown to bind to the pyrimidine-rich loop of hY1 and hY3 RNAs (Fouraux *et al.*, 2002). Here, nucleolin was targeted using siRNAs prior to poly(I:C) treatment (Figure 5.18). YsRNA generation was then assessed by Northern blot. Y1 RNA was analysed (which has been shown to bind to nucleolin), and so was Y5 RNA. Northern analysis showed that both YsRNAs were still generated in Nucleolin depleted cells following poly(I:C) treatment at comparable levels to where a negative control siRNA was used (Figure 5.18).
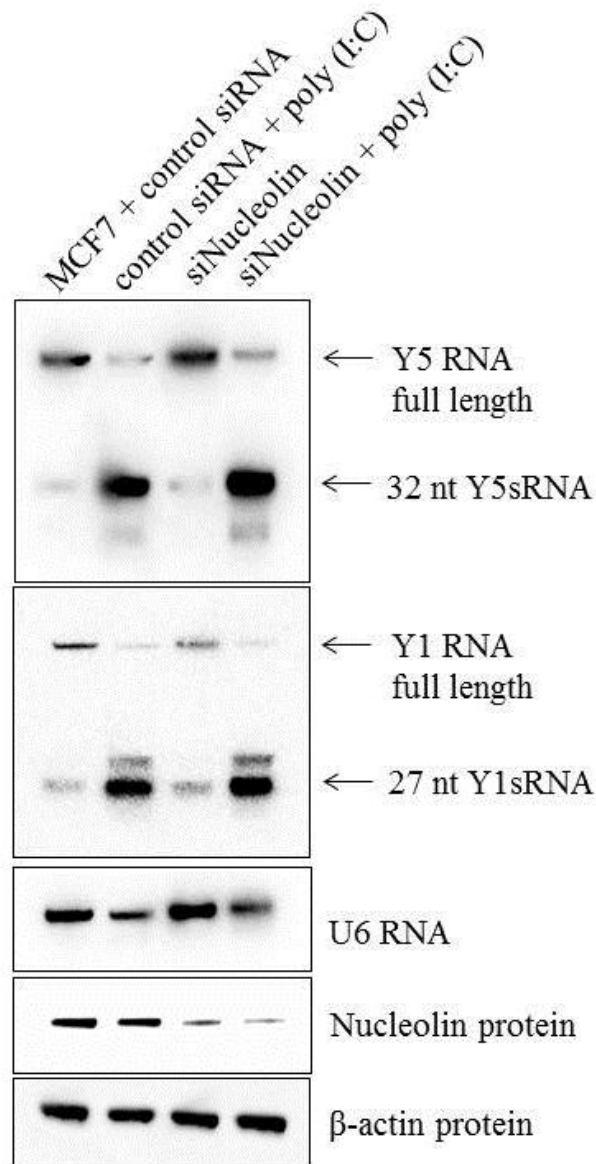
Figure 5.18 The RNA-binding protein and Ro-RNP subunit, nucleolin, does not affect Y RNA cleavage.

Nucleolin mRNA was targeted by RNAi prior to poly(I:C) treatment. Reduced nucleolin levels were confirmed by Western blot. The level of Y1sRNA and Y5sRNA was analysed by Northern blot. Following nucleolin depletion and poly(I:C) stress treatment, YsRNAs from both Y1 and Y5 were still generated. Equal loading of RNA and protein was confirmed using U6 and β-actin probes respectively. SiNucleolin indicates the Nucleolin targeting siRNA.

5.2.5 Investigating the mechanism of YsRNA maturation via Y RNA stem-loop detection

Earlier experiments which utilised cell lysates for *in vitro* cleavage analysis with wild-type and Ro60 knockout cell lines (Figures 5.13-5.14), as well as RNase L knockout experiments (Figure 5.16) hinted at the possibility that YsRNAs are generated via Ro60 protection of the lower region of the full length Y RNAs from non-specific, stress-induced RNase degradation. In order to address this hypothesis, it was investigated whether the pyrimidine-rich loop region of Y5 RNA could be detected by Northern blot. The rationale for this experiment is that if Y RNAs were generated by a specific cleavage event, then the 'loop' region might be detectable by Northern blot (Figure 5.19A). If on the other hand YsRNAs were generated by non-specific degradation of non-Ro60 protected regions, the loop sequence would not be generated or indeed, be detectable.

The 19 nt Y5 RNA loop sequence was not detected in Solexa sequencing data of poly(I:C) treated cells, but the size of the loop region falls outside of the size-class selected for library preparation. Therefore, a highly sensitive LNA probe was used to detect the loop sequence in a range of RNA samples from poly(I:C) treated MCF7 cells (Figure 5.19). Only full length Y5 RNA was detected using the LNA loop probe in non-treated and poly(I:C) treated cells, whereas an oligonucleotide corresponding to the exact sequence of the 19 nt loop region could be detected at amounts as low as 1.0 femtomole (Figure 5.19B). A standard DNA oligonucleotide probe specific to the Y5 3' end was able to detect the stress induced Y5sRNAs. To investigate whether the loop sequence could be detected in

the presence or absence of Ro60, siRNAs were used to knockdown Ro60 in poly(I:C) treated and non-treated cells and RNA was then analysed by Northern blot (Figure 5.19C). Again, no sequence corresponding to the Y5 RNA loop region could be detected.
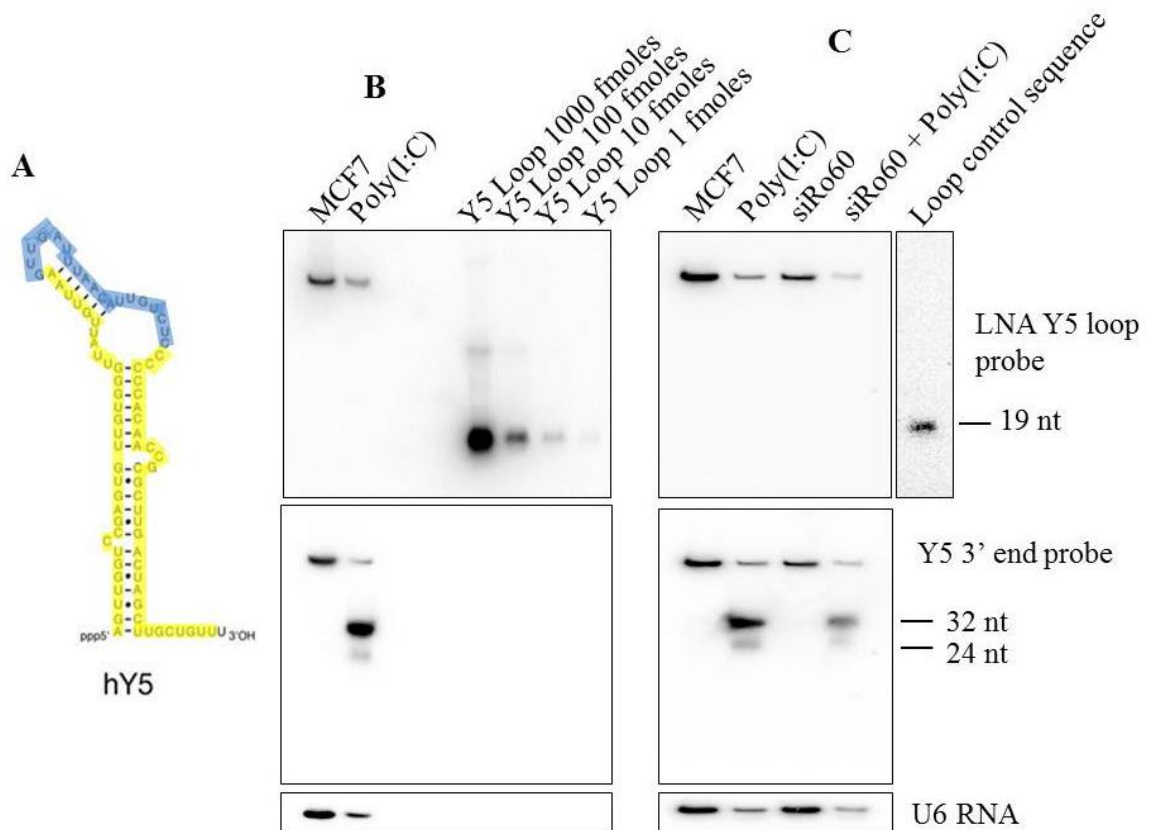
Figure 5.19 Y5 RNA stem-loop analysis.

(A) Annotated diagram of Y5 RNA structure. If Y RNAs were specifically cleaved by an enzyme to generate the 5' and 3' YsRNAs (shown in yellow), the pyrimidine rich loop region would be produced (shown in blue). (B) Attempt to detect the Y5 RNA loop region by Northern blot using LNA 'loop' probe (top panel). MCF7 cells were either treated or not treated with poly(I:C) and RNA was analysed by Northern blot, along with known concentrations of oligonucleotide sequences corresponding to the exact region of the expected loop sequence. Only full length Y5 RNA was detected using the LNA loop probe and no sequence corresponding to the putative 19 nt loop sequence was detected. The poly(I:C) induced Y5sRNAs were detected using Y5 3' end specific DNA oligonucleotide probe (middle panel), and not by the LNA loop probe. (C) MCF7 cells were either treated or not treated with siRNAs against Ro60 (siRo60) prior to poly(I:C) treatment. Again, no Y5 RNA loop derived sequence was detected by LNA loop probe, whereas Y5sRNAs were detected using a Y5 3' end specific probe. A U6 RNA specific probe was used to confirm equal loading (bottom panel).

5.3   Discussion

There are three possible models of how YsRNAs might be produced. Firstly, the YsRNAs could be specifically transcribed from the gene encoding the Y RNA. Although this was not specifically addressed here, based on the fact that the YsRNAs increased in abundance in lysates during *in vitro* cleavage reactions where transcription could not take place owing to reaction conditions (Figures 5.6 & 5.13), and that Y5sRNAs could be produced form a Y5 expression construct in poly(I:C) treated cells using only the promoter for the full length Y5 RNA (Figures 5.11 & 5.16), this scenario seems highly unlikely. Secondly, YsRNAs might be generated by a single enzyme which cleaves the full length Y RNA at specific regions to give the 5' and 3' YsRNAs. And finally, YsRNAs biogenesis could be explained by a simple protein protection model whereby the activity of a cocktail of cellular RNases which have access to only the non-protein bound regions of the Y RNA, would result in Y RNA fragments protected by the proteins being produced. This latter scenario will be referred to as the 'RNase protection hypothesis'. As some cellular RNases are upregulated during cell stress and apoptosis, this might explain the increase in YsRNAs detected under these conditions.

In this chapter, knockout, knockdown and *in vitro* cleavage experiments have demonstrated that Ro60 is essential for YsRNA production for all four Y RNAs (Figures 5.7-5.14). The decrease in YsRNAs was shown to be independent of the corresponding reduction in the stability of the full length Y RNA. Indeed, others have also shown that the full length Y RNA is stabilised by Ro60 (Labbe *et*

al., 1999; Chen *et al.*, 2003; Sim *et al.*, 2009). The Ro60-dependant YsRNA production data shown in this chapter supports the RNase protection hypothesis because if Ro60 is absent, then RNases would have access to all regions of the Y RNA, resulting in complete degradation. However, previous experiments have specifically mapped the Ro60 binding site (Wolin & Steitz, 1984), and the region which gives rise to the YsRNAs is not completely encompassed by Ro60. Nevertheless, this does not rule out that Ro60 binding can prevent RNases from having access to some sequence region which is not directly bound by Ro60. Furthermore, the experiments used to determine the Ro60 binding site had to be repeated several times before they were successfully published, and used an excessive amount of pancreatic RNase to map this binding site (personal communication with Professor Sandra Wolin). This in turn might have given an exaggerated impression of the exact sequence regions which Ro60 protects.

Angiogenin, which specifically generates the biological functional tRNA halves, did not appear to give rise to YsRNAs based on *in vitro* cleavage experiments (Figures 5.5 & 5.6 and Figures 5.13 & 5.14). Interestingly, in MCF7 and mES wt cell lysates, YsRNAs from both Y1 and Y5 were shown to increase following extended incubation with angiogenin, RNase A or reaction buffer alone (Figures 5.6 & 5.13). However, these YsRNA fragments were not detected in Ro60 knockout cell lysates (Figure 5.14). This data would also support the RNase protection hypothesis for YsRNA generation, as RNases within the cellular lysate were clearly producing the YsRNAs only in the presence of Ro60, even in the absence of RNase A or angiogenin.

RNase L specifically had an effect on YsRNA biogenesis, but only in mouse cells (Figure 5.16). It is possible that RNAi targeting in the human MCF7 cells was sufficient to show an RNase L depletion by Western blot (Figure 5.17), but some low level of cellular RNase L may have still been present at functionally active but undetectable levels. Alternatively, this data could imply a differential YsRNA biogenesis pathway between human and mouse. As RNase L depletion only reduced YsRNA generation and did not completely eliminate it, this supports the hypothesis that a cocktail of RNases contributes to YsRNA production, and no one enzyme alone. Unfortunately, the experiments done here have not revealed all the RNase(s) involved. This could be investigated in future studies by using a comprehensive, genome wide RNAi screen against all the individual cellular RNases.

The relative abundance of the YsRNAs is reflective on the approximate molar ratios of full length Y RNAs within the cell, as determined in previous studies (Gendron *et al.*, 2001; Christov *et al.*, 2006). Y5 RNA makes up ~ 60 % of the total Y RNA population and gives the strongest signal for YsRNA fragments on a Northern blot (Figures 5.8-5.10, and see Chapter 4). Y4 is the next most abundant Y RNA and also gives rise to the next most abundant YsRNA. Y1 is expressed at a slightly lower level than Y3, but Y1sRNAs are more abundant than Y3sRNAs. The nature of YsRNA abundance reflecting cellular concentrations of full length Y RNAs would further support a case for the RNase protection hypothesis, as non-specific degradation of exposed non-Ro60 protected regions would be equal for all Y RNAs. Failure to detect the pyrimidine-rich loop region of Y5 RNA in poly(I:C) treated cells (Figure 5.19) also supports the proposed model

for YsRNA biogenesis. However, it is possible that if a specific cleavage were to occur in order to generate YsRNAs, the half-life of any resulting by-product (the loop region) might be too short for the purpose of Northern blot detection.

Other Y RNA binding proteins did not appear to have the same essential role for YsRNA generation as Ro60. Nucleolin depletion did not affect YsRNA generation (Figure 5.18). Although, as siRNA targeting did not achieve a complete knockdown of the protein, it is possible that the small amount of nucleolin remaining was enough to be functional.

Depletion of La protein levels only appeared to affect Y5sRNA biogenesis by resulting in an increase of the shorter, 24 nt Y5sRNA. Considering the structures of the Y RNAs, the 3' tail region which La binds is much longer in Y5 RNA compared to the other Y RNAs. The longer (32 nt) and shorter Y5sRNA (24 nt) have identical sequences apart from the 8 nt 3' tail region (Chapter 3). Therefore, one possible explanation is that when La is absent, the 3' tail region is de-protected and therefore vulnerable to degradation from the 3' end, resulting in more 24 mer YsRNA (which remains protected by Ro60). As the other Y RNAs have much shorter 3' tails, depletion of La would not have such a drastic change in YsRNA size, as reflected in figures 5.8-5.10.

Interestingly, YsRNAs were still generated in mouse embryonic stem cells despite the fact that the interferon response has been shown not to occur in ES cells (Yang *et al.*, 2001). Nonetheless, poly(I:C) treatment still triggered YsRNA generation in these cells, which might indicate an immune-independent function of the YsRNAs.

Functional sRNAs are usually generated with specificity. For example, Dicer generates microRNAs and angiogenin generates tRNA halves. MicroRNAs and tRNA halves then go on to serve precise biological functions. If YsRNAs were generated by the RNase protection hypothesis, it would not be unreasonable to assume that the resulting products are not functional based on the chaotic nature of their biogenesis. Having said this, it does not rule out that a biological function has evolved to utilise these YsRNAs, which may have originally arisen in ancestral cells by chance through RNase protection, but have since become functional. Based on the high abundance of YsRNAs within the cell, and the evidence that they are specifically incorporated into immune cell-derived vesicles (Nolte-'t Hoen *et al.*, 2012), this would certainly hint at biological functionality.

Future work should aim to characterise the biological role of these YsRNAs. Work in our laboratory has started to address this question by subjecting proteins in the same biochemical fraction as YsRNAs to mass spectrometry. We identified around 160 proteins, including Ro60, La and nucleolin (data not shown). Interestingly, many proteins related to proteolysis, including E3 ubiquitin ligase, several proteasome subunits and heat shock proteins such as Hsp70 and Hsp90, were also found in the same biochemical fraction as YsRNAs. Additionally, we found that two proteins which have been shown to antagonise apoptotic pathways, nucleophosmin and the 14-3-3 protein (Fu *et al.*, 2000; Dhar & St Clair, 2009), also co-purified with the YsRNAs according to our mass spectrometry data. This preliminary work would suggest that YsRNAs are associated with pathways related to survival following cellular stress, which is perhaps why they show a specific increase following apoptosis and viral infection.

Chapter 6

Next-generation sequencing of a novel class of small

RNAs

6.1 Introduction

Next-generation sequencing (NGS) technologies have allowed high throughput analysis of small RNA populations, and facilitated the subsequent discovery of novel small RNAs. For example, a recent study has discovered 16-28 nt RNAs in chloroplasts following deep-sequencing of *Arabidopsis* (Ruwe & Schmitz-Linneweber, 2012). Interestingly, it was shown that these chloroplast sRNAs were footprints of the translation related pentatricopeptide repeat proteins (PPR) and were generated via RNase protection of intergenic regions of chloroplast transcripts. The question of whether these chloroplast sRNAs are merely footprints or whether they go on to serve a biological function remains unanswered. A different study which analysed non-coding RNAs (ncRNAs) in differentiating mouse ES cells via NGS sequencing (Skreka *et al.*, 2012) identified a class of differentially expressed 18-30nt RNAs, also derived from intergenic regions, most of which were novel ncRNAs lacking characteristic microRNA features. Two other size classes were also found in the study, a 45-47 nt class and a 71-78 nt class.

The reasons why these novel sRNAs were missed in previous years is owing to firstly; the increased number of reads which can now be analysed in a single sequencing reaction (the 'depth' of sequencing), and secondly; that gene silencing-related RNAs are predominantly 18-24 nt, resulting in specific selection of these small RNAs during library preparations, neglecting RNAs of other sizes. Indeed, in relation to this project, early Northern blot analysis of Y RNA fragments (Chapter 3) revealed unexpected bands of 28-33 nt RNAs, which encouraged

specific selection of these additional sRNAs during library preparations in subsequent sequencing experiments. In these later experiments, it was noted that a peak of ~30 nt found following Illumina (Solexa) sequencing of MCF7 cells (Chapter 3, Figure 3.2) contained unknown sRNAs in addition to YsRNAs. These ~30 nt RNAs did not map to the well characterised 24-31 nt Piwi-interacting RNAs (piRNAs) which act in pathways to protect the genome against transposable elements (Siomi *et al.*, 2011), indicating that they were from a different class of sRNAs.

The millions of reads generated during sequencing reactions which aim to detect novel small RNAs make it very challenging to differentiate functional molecules from degradation products. For microRNAs, computer programmes exist which can identify functional molecules using a set of well-established parameters (Moxon *et al.*, 2008). Obviously, the function and therefore characteristic features of putative novel sRNAs is not known prior to their discovery. However, several experimental designs can be employed which help to filter out background degradation products. Analysis of cells and tissues which are treated under different conditions, or are at different stages of development, as well as sequencing only RNAs which are protein bound, can all facilitate the identification of functional sRNAs over degradation products (Skreka *et al.*, 2012).

Solexa (Illumina) sequencing is the most widely used deep sequencing technique for small RNA detection. Indeed, the majority of miRNAs in miRBase have been discovered using Illumina sequencing technology. The current sequencing protocol relies on RNA 'adapter' molecules ligating to the 3' and 5' ends of total RNA extracted from a tissue or cell sample. Ligation of adenylated 3'

adapters to the small RNAs is achieved using a truncated form of T4 RNA ligase 2 (Rnl2) and then 5' adapters are ligated to the small RNAs using T4 RNA ligase 1 (Rnl1). Reverse transcription followed by PCR (RT-PCR) then results in generation of complementary DNA molecules (cDNA) from the ligated product which can then be separated on a polyacrylamide gel. The desired size class of small RNA can then be selected for sequencing by purifying the appropriate cDNA band from the gel using size markers as a guide. However, recent work has demonstrated that a considerable bias exists by which the adapter molecules can ligate to target small RNAs of interest (Linsen *et al.*, 2009; Hafner *et al.*, 2011). The sequencing bias in small RNA libraries is ligase dependent (Nandakumar & Shuman, 2004) and results in failure to detect thousands of non-adapter or reduced-adapter ligated small RNAs in library preparations.

In this chapter, in order to better characterise the population of 30-40 nt novel sRNAs in which YsRNAs are found, Illumina (Solexa) sequencing was employed to identify putative candidates. As poly(I:C) treatment induced the generation of stress related YsRNAs, and as Ro60 appeared to be responsible for their biogenesis, both wild-type and Ro60 knockout cells either treated or not-treated with poly(I:C) were subjected to NGS, selecting for molecules between 30-40 nt. Over 150 sRNAs were identified which were found to be upregulated following poly(I:C) treatment, but which were not produced in the absence of Ro60 under the same conditions, implying that Ro60 was required for the production of these molecules. Intriguingly, most of these sRNAs mapped to exonic regions of the genome. Furthermore, the reads of these novel sequences

showed a high sequence conservation at their 3' end, but had a more variable 5' end.

In addition to testing the Ro60 dependency of the ~30 mer sRNAs, experiments were performed to give an indication of the abundance of this size class in adult mice. Small RNAs of the ~30 mer class were selected for sequencing from a range of six organs in mouse which revealed a specific size peak at 40 nt and also 33 nt. Again, these sRNAs primarily mapped to exons. Bioinformatic analysis of both these data sets is ongoing.

In collaboration with others in the laboratory and with the School of Computing Sciences, University of East Anglia, a technique of reducing the sequencing bias of small RNAs was developed and tested. The new technique utilised adapters with degenerate tags (variable sequences) called 'high definition' adapters (HD adapters). Using quantification of small RNAs by Northern blot as a standard, data obtained using HD adapters correlated better with absolute quantifications than did standard Illumina adapters. The use of HD adapters will be a highly useful tool for future investigations of novel classes of small RNAs.

6.2 Results

6.2.1 Deep sequencing of RNA from Ro60 knockout cells reveals novel Ro60-dependent small RNAs

As the Ro60 autoantigen is essential for YsRNA biogenesis (see chapter 5), this raised the question as to whether other 30-40 nt non-Y RNA related small RNAs were also regulated by Ro60. In order to address this, both wild-type mouse embryonic stem cells (mES) and mES Ro60 knockout cells (Ro60$^{-/-}$) were either treated or not treated with poly(I:C). This was repeated in duplicate, generating eight biological samples. Ro60 knockout cells were genotyped by Western blot as before (see chapter 5). RNA was then isolated from all cells. Small RNA libraries were generated using the latest Solexa protocol (TruSeq Small RNA) using different 'barcodes' (specific PCR primer sequence to allow for multiplexing) for each of the libraries. The Illumina Truseq protocol is optimised for microRNA library preparation. As this new size class of 30-40 nt had not specifically been investigated before, the protocol had to be modified in order to better select for this new size class. It was found that using 2 µg small RNA (rather than 1.0 µg total RNA suggested by Illumina) for adapter ligation was optimal for library preparation of this size class. Therefore, RNA extracted from cells was enriched for the small RNA fraction (RNAs <200 nt) prior to library preparation. Additionally, between 17-19 PCR cycles (as opposed to 11 cycles suggested by Illumina) was optimal for the generation of libraries for 30-40 nt sRNAs. A comparison of the resulting cDNAs generated using either the standard microRNA

or 30-40 nt specific library preparation protocols shows that the method developed

here was highly efficient at selecting for this new size class (Figure 6.1).
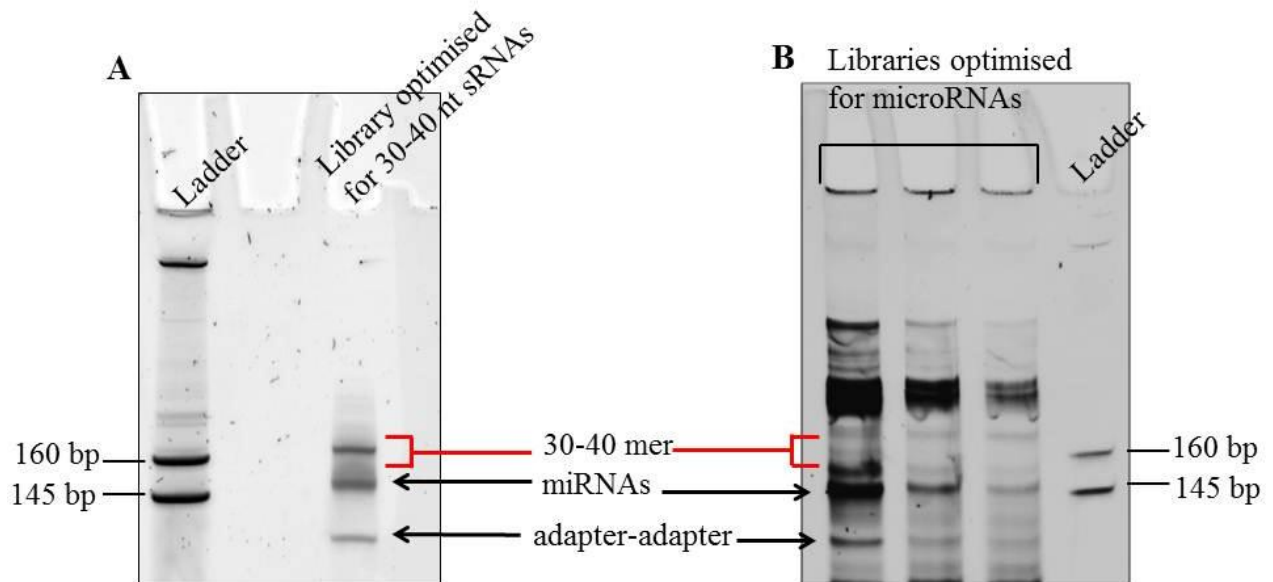
Figure 6.1 Comparison of small RNA cDNA library preparations for microRNAs and 30-40 nt sRNAs.

Small RNA libraries were prepared using the TruSeq small RNA kit (Illumina). (A) Library preparation optimised for 30-40 nt sRNAs (investigated in this chapter) and separated by polyacrylamide gel. Libraries prepared using 2 µg enriched sRNA (<200 nt) and RT-PCR protocol with high PCR cycle numbers (17-19 cycles). The size of the combined barcode and adapter-adapter cDNAs alone is 125 bp. Libraries of 22 nt microRNAs therefore gives a 147 bp band. Gel A shows result of second gel purification after selecting all cDNAs between 140-170 bp in a previous purification. A band representing microRNAs, as well as a larger band of ~160 bp (representing ~35 nt sRNAs) is clearly visible. The red bracket shows cDNA which was purified for sequencing, or the equivalent region on gel B. (B) An example of standard microRNA cDNA library preparation by another researcher in the laboratory, Dr Ping Xu, using the advised reaction conditions stated in the Illumina protocol. A band representing miRNAs is clearly visible in each of the three lanes representing cDNAs generated using different PCR cycles. However, the equivalent ~35 nt sRNA band shown in gel A is only detected at low levels in the first lane of gel B.

Following library preparation, samples were sent for Solexa (Illumina) sequencing where they were first quantified and then each sample (including duplicated repeats) was sequenced on two different lanes on a flow cell using 50 cycles to account for the longer than usual sRNA size. Data was analysed by firstly removing adapter sequences and then removing reads that were less than 16 nt (0.4-3 % of total reads) and greater than 45 nt (no reads were actually greater than 43 nt). Reads mapping to a loci at chromosome 18 representing a highly abundant snoRNA fragment (from SNORD58) were also removed, as this sequence was overrepresented owing to sequencing bias.

The resultant reads were then mapped to the GRCm 37 release (genome reference consortium for mouse) of the mouse genome using sequences with exact matches (Table 6.1). Between 9.9 and 18.1 million redundant reads (non-unique reads) were generated for each sample of which 79-89 % mapped to the genome. With regards to unique reads, between 7.7-13.1 % mapped to the genome. The discrepancy between the percentages of redundant *versus* unique reads which mapped to the genome is explained by the fact that a small number of unique sequences had high read numbers (this is a normal result for sequencing of a specific size class where there is low degradation or background 'noise'). To better explain this, let us take a theoretical data set where we have 10 reads composed of three different sequences: 6 redundant reads with sequence 'A' (which map to the genome), 3 redundant reads with sequence 'B' (which do not map to the genome) and 1 redundant read with sequence 'C' (which does not map to the genome). In this scenario, 60% of redundant reads map to the genome, but only 33 % of unique

reads map to the genome, because only one read represented with sequence 'A' out of the three different sequences in total can map.

A size distribution was then generated for each of the four biological conditions and for each of the biological repeats (Figure 6.2). Results were normalised using quantile normalisation which is more appropriate when comparing different samples than proportional normalisation (or 'reads per million' normalisation) (Garmire & Subramaniam, 2012). Briefly, quantile normalisation is based on ranking sequences in order of abundance for the different samples. This makes sequence distributions more comparable with each other, and does not introduce or reduce differential expression in the same way which proportional normalisation does. Size distributions for all samples showed clear peaks at 28 nt, 33 nt and 38 nt (Figure 6.2). No peak at 22 nt representing microRNAs was found, as this size class was omitted during library preparation.

Table 6.1 Analysis of reads generated from deep sequencing of wild-type and Ro60 knockout mouse embryonic stem (mES) cells either treated or not treated with poly(I:C).

RNA was extracted from cells and enriched for <200 nt molecules. Experiments were repeated twice (biological replicates) represented by #. Following cDNA library preparation using a modified Illumina protocol optimised for 30-40 nt sRNAs, libraries were sequenced and mapped to the GRCm 37 release of the mouse genome. The discrepancy between the percentages of redundant *versus* unique reads which mapped to the genome is normal for sequencing of a specific size class, and is explained by the presence of a small number of unique reads with high read numbers.

	Mapped redundant reads	Mapped unique reads	Unmapped redundant reads	Unmapped unique reads	Percentage redundant reads mapped	Percentage unique reads mapped
mES wt (#1)	8864423	12032	11038296	156118	80.3	7.7
mES wt (#2)	11004442	8977	12379910	78549	88.9	11.4
mES poly(I:C) (#1)	8997915	19153	11370042	159082	79.1	12
mES poly(I:C) (#2)	8555976	14262	9916806	108688	86.3	13.1
Ro60$^{-/-}$ (#1)	10398632	11864	12162696	91841	85.5	12.9
Ro60$^{-/-}$ (#2)	12750874	10785	14935741	96296	85.4	11.2
Ro60$^{-/-}$ poly(I:C) (#1)	10170919	24446	12698282	210363	80.1	11.6
Ro60$^{-/-}$ poly(I:C) (#2)	15437297	15400	18181542	144439	84.9	10.7
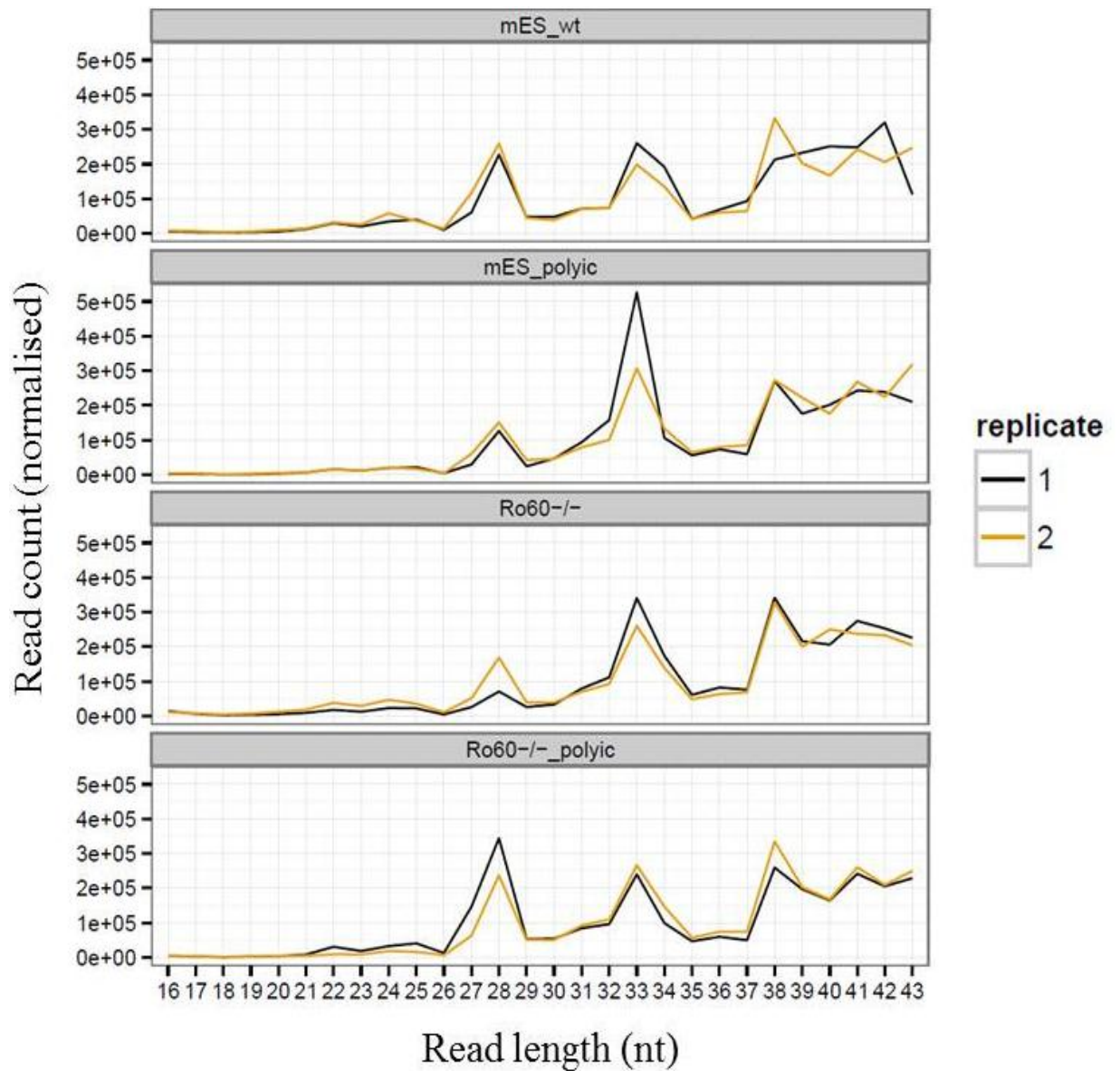
Figure 6.2 Size distribution of sequences from wild-type mouse embryonic stem cell (mES) and mES Ro60 knockout (Ro60$^{-/-}$) cells either treated or not treated with poly(I:C).

Two biological replicates were sequenced using Solexa (Illumina) technology for each of the four experiments. Size distribution showed peaks at 28 nt, 33 nt and 38 nt. As the microRNA size class was omitted during library preparation, there is no visible peak at 22 nt. Data was quantile normalised.

As YsRNA biogenesis is dependent on Ro60, sequences derived from mY1 and mY3 RNAs were first assessed as a positive control to see if the data set could reliably be used to find other Ro60 dependent small RNAs. A 'presence plot' was generated for both Y RNAs which plots the appearance of each nucleotide in all sequencing reads against its position in the genome (Figure 6.3). For example, taking two theoretical sequences which map to the same locus: GATC (with 4 reads) and GATCGA (with 6 reads), the 'abundance' for each base at a particular position in the sequence is calculated as the sum of all reads covering that base. So with the example given, the abundance of 'G' at position one in both theoretical sequences (mapped to the same locus) would be 10 because it appears in both sequences at the same position and the sum of the reads of the two sequences gives an abundance value of 10. In the case of the last 'A' in sequence two; it would only have a value of 6 as it only appears in the second sequence which has 6 reads.

The presence plot revealed that sequences from the 5' and 3' ends of mY1 had a higher abundance in poly(I:C) treated wild-type mES cells than for non-treated wild-type mES cells (Figure 6.3). The abundance of reads from mY1 in Ro60$^{-/-}$ cells (treated and non-treated with poly(I:C)) remained low across the mY1 locus. The presence plot for mY3 was not as conclusive as high abundance was only detected at the 3' end and the overall abundance mY3 was low (see scale on the *y*-axis). Furthermore, the abundance for one of the non-poly(I:C) treated wild-type mES cell repeat samples was much higher than for both poly(I:C) treated wild-type mES cell repeats. Earlier Northern blot data showed that Y3 RNA expression and Y3sRNA production is generally much lower compared to the other Y RNAs and does not seem to be representative of YsRNA biogenesis (for

example see Figure 4.1, 4.2, 5.9 and 5.11 from previous chapters). However, the mY1 presence plot did correlate with Northern blot analysis from Ro60 knockout experiments (see chapter 5) confirming that YsRNAs are dependent on Ro60. This in turn demonstrated that the sequencing data set could reliably be used to find other potential Ro60-dependent small RNAs.
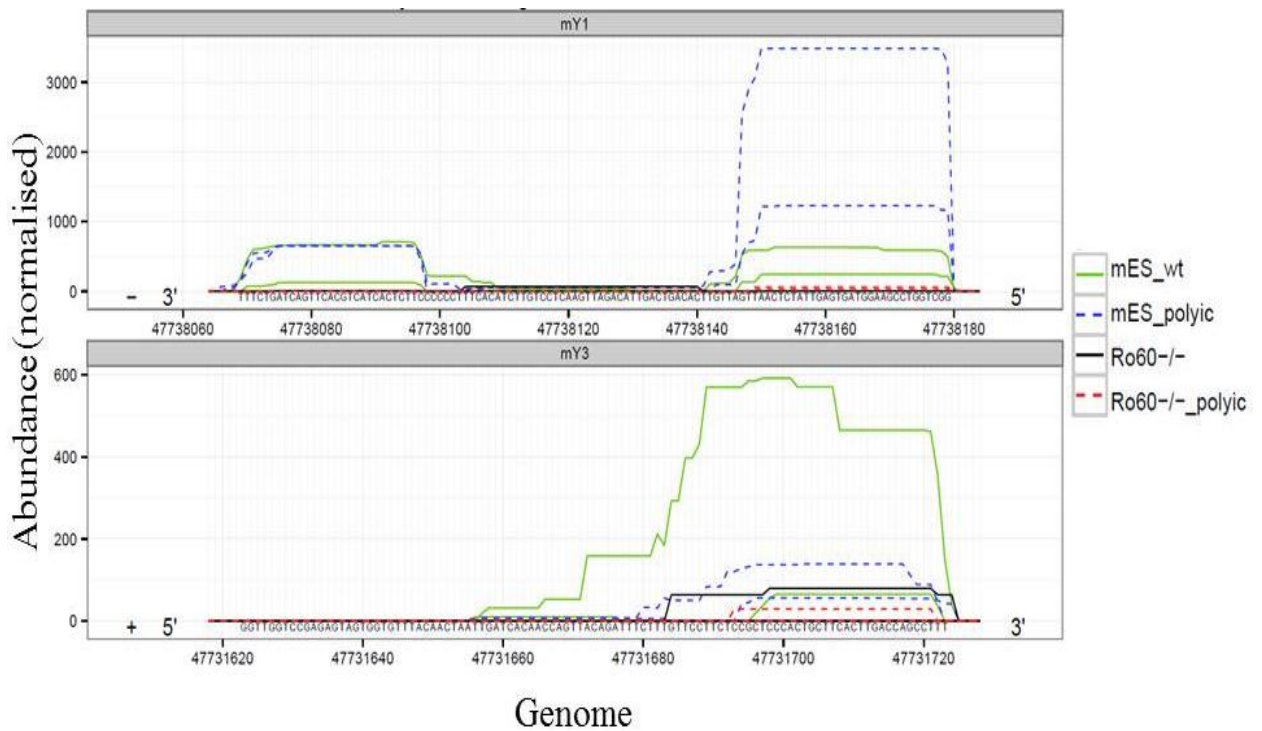
Figure 6.3 Presence plot for Y1 and Y3 RNA from wild-type mouse embryonic stem cells (mES) and mES Ro60 knockout (Ro60$^{-/-}$) cells either treated or not treated with poly(I:C).

The presence of each nucleotide in reads which mapped to the loci shown were analysed for mY1 (top) and mY3 (bottom) for all four experiments using redundant sequences. Data for biological duplicates is also shown. Presence plot for mY1 (top) showed that sequences from the 5' and 3' ends had higher read counts in wild-type poly(I:C) treated cells (mES_polyic) than in non-treated wild-type cells (mES_wt). Read counts for mY1 remained low across the locus for all Ro60 knockout (Ro60$^{-/-}$) cells. Presence plot for mY3 (bottom) shows only fragments derived from the 3' end of Y3 RNA. One of the replicates for non-poly(I:C) treated cells shows much higher read counts compared to wild-type poly(I:C) treated cells. Data was quantile normalised.

In order to see if there were other Ro60-dependent small RNAs which were upregulated during cell stress, differentially produced sequences were filtered from the data set using two different parameters. Sequences were selected which showed a four-fold increase in production ($2\log_2$ offset fold change using an offset value of 20) between wild-type mES cells and poly(I:C) treated wild-type mES cells, but which did not show differential production between Ro60$^{-/-}$ cells and poly(I:C) treated Ro60$^{-/-}$ cells. The most differentially produced sequences mapping to each loci were used in the analysis. Subsequently, 177 sequences were found which showed at least a four-fold increase during stress, but which specifically required Ro60 for their generation. These novel sequences will be referred to as rosRNAs (Ro60-dependent small RNAs).

Interestingly, rosRNAs primarily mapped to exonic regions of the genome, implying that they were derived from mRNA transcripts. The list of 177 candidates did not include tRNA halves (as expected) but did include the 5' end Y1sRNA - which appeared as the 26th most differentially produced rosRNA. In order to further filter the list of candidates, all 25 sequences from loci showing a greater Ro60-dependent differential production than Y1sRNA were chosen for analysis, as well as 8 sequences which were the next most Ro60-dependent after Y1sRNA. These sequences were all between 31-41 nt in length.

The presence plots for all sequences mapping to these 33 loci were then analysed, and only 6 of these sequences were then selected as *bona fide* rosRNAs (Table 6.2) based on more stringent analysis of their differential generation pattern between experiments (Table 6.3). The most differentially produced sequence from each Ro60-dependent locus was selected as the consensus sequence for each

rosRNA. The change in Y1sRNA-5' production was also analysed (Table 6.3). None of the rosRNAs showed a fold change difference between untreated mES wild-type cells and untreated Ro60$^{-/-}$ cells, whereas Y1sRNA production reduced in an untreated Ro60$^{-/-}$ background (-1.6 $\log_2$ offset fold change), as this sRNA is naturally produced in non-stressed cells and is Ro60 dependent (see Chapters 3 and 5 respectively).

Table 6.2 Ro60 dependent small RNA (rosRNA) candidates.

The sequence and origin of the six putative rosRNAs is shown. The candidates all mapped to exonic regions: either the coding regions of genes (CDS) or the untranslated regions (UTR).

	Sequence	size (nt)	Feature	Description
rosRNA-1	GAAAACCCGGCTAGCTGTAGCCAAAACTGGACAATACTC	39	CDS	solute carrier family 25
rosRNA-2	TTCTGTATTTTAGGTCTAGTACTGTTCATGACTA	34	3'-UTR	ataxin 1
rosRNA-3	TGTAGGTTAGAACTCCCCCAGTGCGAGGAAAATGGACT	38	5'-UTR	lipolysis stimulated receptor
rosRNA-4	TTTCGGGGCCTAAGGTCAATGTAGAAGCTCCAAACGTCAACA	42	CDS	AHNAK nucleoprotein
rosRNA-5	TACAATGCTGTCCCTTTCAGCTGGAGGATCAGGGACTACC	40	CDS	ring finger protein 213
rosRNA-6	GTAGTGTAACTTTGGGAACCATCTGCCCATCTA	33	3'-UTR	ubiquitin specific peptidase 47

Table 6.3 Fold change of rosRNA candidates and Y1sRNA between experimental conditions and cell lines.

Table shows the $\log_2$ offset fold change values (using an offset value of 20) of rosRNAs between mES wild-type and Ro60[-/-] cells under poly(I:C) treated or untreated conditions. Candidates were selected on the conditions that they showed a four-fold change between mES wt untreated and mES wt poly(I:C) treated conditions, but did not change between Ro60[-/-] and Ro60[-/-] poly(I:C) treated conditions. This confirmed that they were indeed Ro60 dependent. Fold change differences for Y1sRNA-5' is shown as a comparison.

	mES wt → Ro60[-/-]	mES wt poly(I:C) → Ro60[-/-] poly(I:C)	Ro60[-/-] → Ro60[-/-] poly(I:C)	mES wt → mES wt poly(I:C)
rosRNA-1	0.0	-4.6	0.3	4.9
rosRNA-2	0.0	-4.5	0.0	4.5
rosRNA-3	0.0	-4.3	0.0	4.4
rosRNA-4	0.0	-4.8	-0.9	3.9
rosRNA-5	0.0	-3.0	0.0	3.0
rosRNA-6	0.0	-2.8	0.0	2.8
Y1sRNA-5'	-1.6	-4.5	0.0	2.9

Surprisingly, these six rosRNA sequences shared a common read structure, whereby the 3' end showed high sequence conservation and the 5' end showed greater variation (Figures 6.4-6.5). For example, rosRNA-1 has a very defined 3' end whereas the presence plot for the 5' end is more sloped with a less defined cut-off point. This could be due to the rosRNA being derived from the 3' end of a transcript, or due to evolutionary selective pressure at the 3' end owing to conservation of important sequences, which is similar to the asymmetric end selection seen with animal microRNAs (but with the opposite end). The 'seed' region in microRNAs (nucleotide positions 2-8) is at the 5' end and presence plots for microRNAs shows a fixed 5' end and more variable 3' end.

Upon closer investigation, it was noted that there was a 'kink' or step in the rosRNA presence plots consistently appearing at the 5' end (Figures 6.4-6.5). This kink corresponded with a run of 3-4 T nucleotides at the 5' end of these sRNAs (with the exception of rosRNA-5 where the run of Ts appears at a secondary kink in the centre of the sRNA). This is identical to the La binding motif. Indeed, the differential length of rosRNAs at the 5' end may be due to differential La binding of these transcripts, which is what we have seen for Y5sRNA-3'. With Y5sRNA-3', a 32 nt fragment is produced when La is present, and in its absence, the 24 nt Y5sRNA is preferentially produced (Chapter 5, Figures 5.8-5.10). Differential La binding might explain the 5' end variation seen in rosRNAs.

These rosRNA candidates are currently being subjected to further investigation in the laboratory.
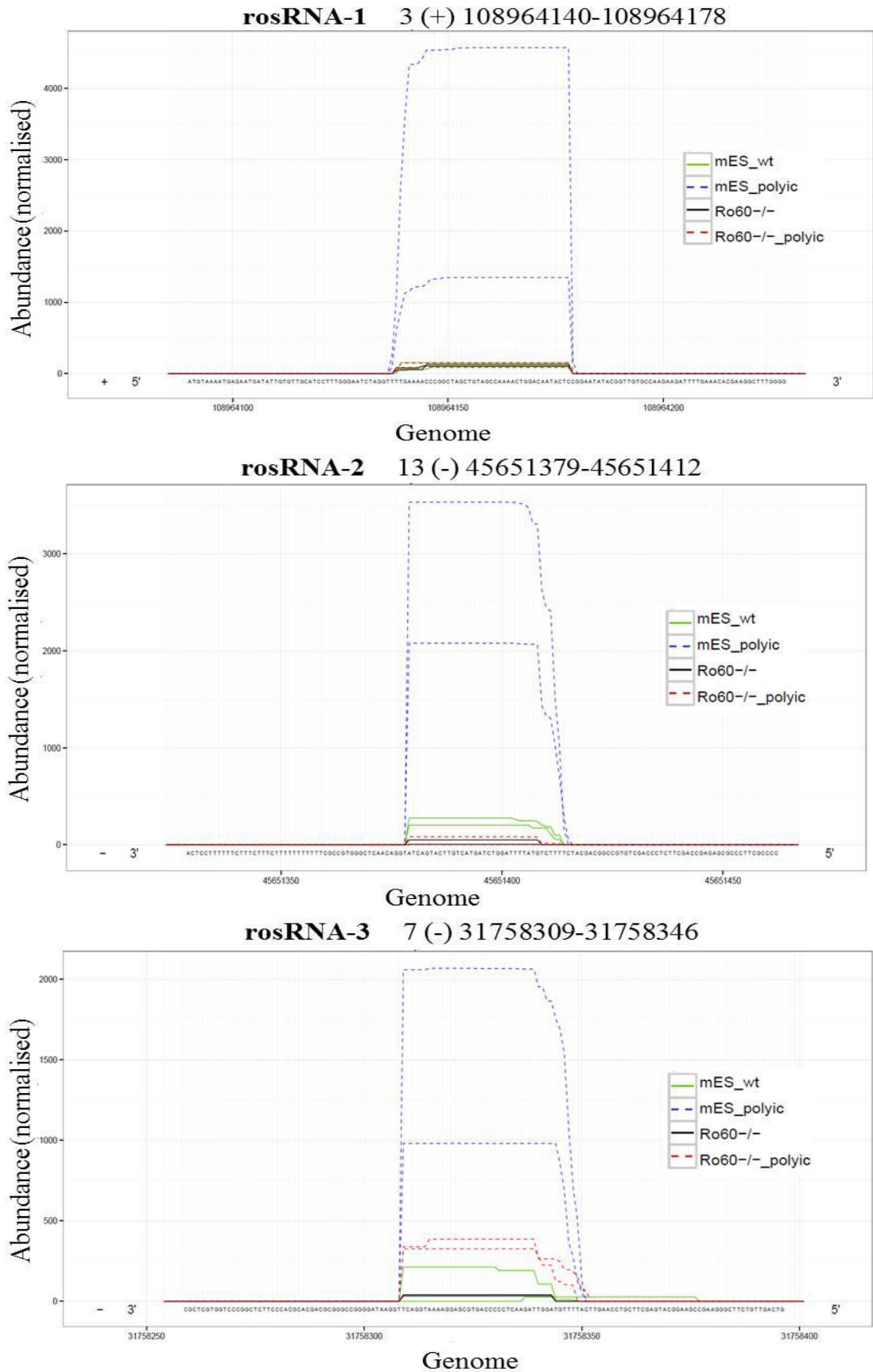
Figure 6.4 Ro60-dependent small RNA (rosRNA) 1-3 presence plots.

Sequences mapping to each locus were analysed. Mapped reads showed a highly conserved 3' end and a more variable 5' end.
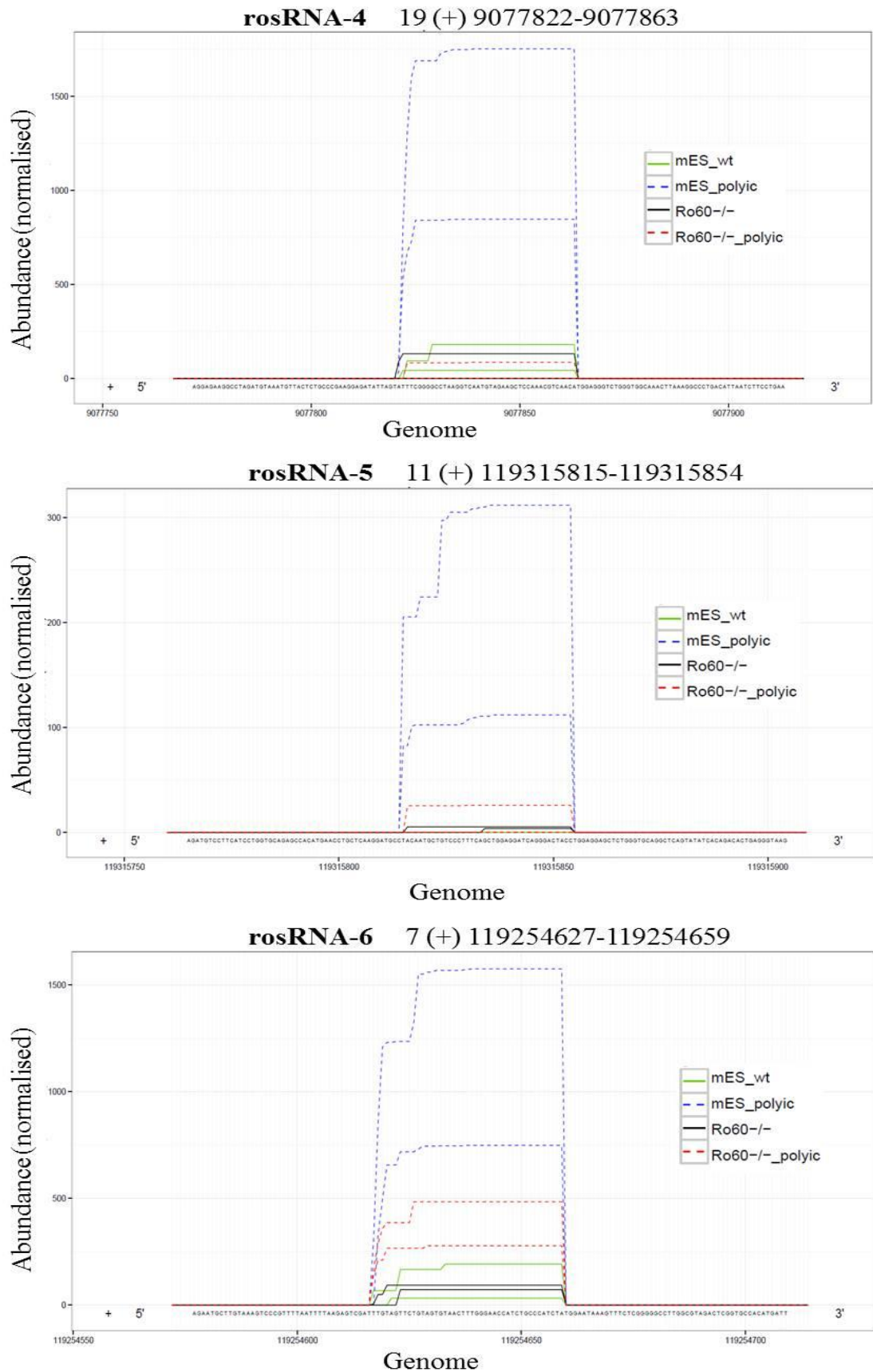
Figure 6.5 Ro60-dependent small RNA (rosRNA) 4-6 presence plots.

Sequences mapping to each locus were analysed. Mapped reads showed a highly conserved 3' end and a more variable 5' end.

6.2.2 Next-generation sequencing of 30-40 nt sRNAs in mouse organs

The 30-40 nt small RNA population studied using mouse embryonic stem cells was highly informative at revealing Ro60-dependent small RNAs. However the experimental design would not reveal the full extent of this novel RNA population in animals as only a single undifferentiated cell type was employed. Therefore, in order to better characterise this class of sRNAs in animals, 30-40 nt RNAs from different organs in mouse were investigated via NGS as described before. The brain, heart, kidney, liver, lungs and femoral muscle were all dissected from two different adult mice (one male and one female) and cDNA libraries were subjected to Illumina sequencing.

As with rosRNAs, the 30-40 nt sRNAs in mouse organs were primarily composed of RNAs derived from exons. Analysis of the redundant size distribution of reads for all organs from these two individuals revealed a significant peak around 40 nt, and smaller peak at 33 nt (Figure 6.6). The kidney sample extracted from the female mouse showed a peak around 22 nt, but this was probably due to microRNA contamination during library preparation. Bioinformatic analysis of this data set is ongoing, but the initial read distribution results clearly demonstrated that a 30-40 nt small RNA population exists in the cell, and shows varied abundance between different organs. Small RNAs from this size class were most abundant in the brain and kidneys, and showed the lowest production in the heart (Figure 6.6). Individual sequences with high read counts showed organ-specific production which was reproduced in the two biological replicates (preliminary data not shown).
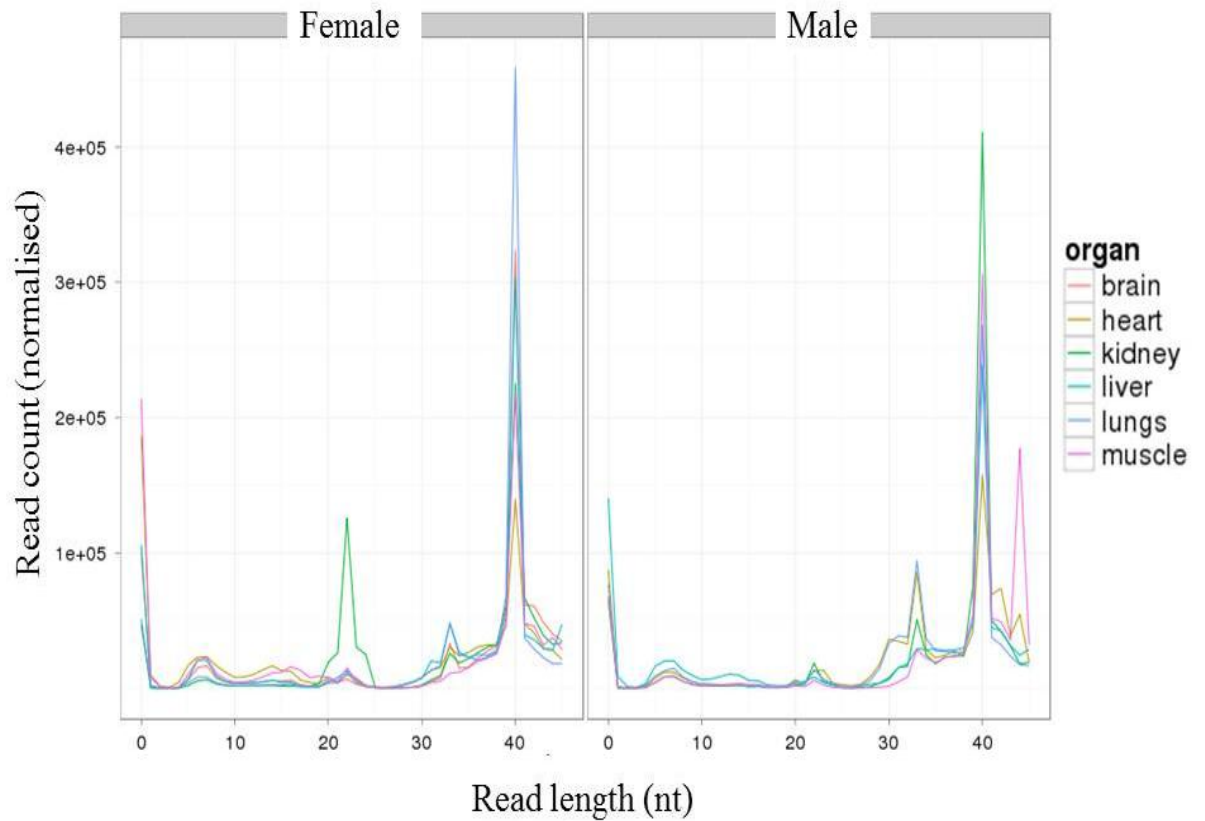
Figure 6.6 A population of 30-40 nt small RNAs in mouse organs.

Six different organs (brain, heart, kidney, liver, lung and muscle) were extracted from two different mice (an adult male and adult female). RNA libraries were prepared using the Illumina protocol, selecting from the 30-40 nt size class. Size distribution of redundant sequences revealed specific peaks at 40 nt and 33 nt. The peak visible at 22 nt for the female kidney sample is probably owing to microRNA contamination in that sample.

6.2.3 Improving next-generation sequencing of small RNAs by reducing sequencing bias

Current sequencing protocols have been shown to be biased with respect to which small RNAs are preferentially sequenced. This in turn could hinder the discovery of novel small RNAs and new classes of RNAs. The sequencing bias is due to RNA ligases having a preference for certain secondary structures at the adapter-small RNA ligation site (Nandakumar & Shuman, 2004). We found that the use of commercial adapters with fixed sequences worsened this problem by restricting the diversity of possible secondary structures generated between the fixed adapters and the small RNA molecules. In order to counteract this problem our laboratory developed and patented adapters with degenerate tags (varied sequences) called high definition adapters (HD adapters) thereby increasing the number of different secondary structures at adapter-small RNA ligation sites for each small RNA. Subsequently, this would increase the ligation efficiency and therefore the representation of certain small RNAs in cDNA libraries. Here, the effectiveness of HD adapters for detecting small RNAs was investigated in biological samples.

RNA was extracted from the colon cancer cell line, DLD-1. Small RNA libraries were then prepared by a collaborator in the laboratory, Dr Karim Sorefan using the Illumina protocol. Libraries were prepared using either Illumina adapters or HD adapters and sent for sequencing. RNA from the same samples was then analysed by Northern blot for eight known microRNAs (miR-10a, miR-19b, miR-21, miR-25, miR-29b, miR-93, miR-375 and let-7i) (Figure 6.7). To quantify these

microRNAs, serial dilutions were made with eight oligonucleotide sequences corresponding to each respective microRNA. Calibration curves were then generated and used to quantify endogenous microRNA concentrations in two technical replicates. Absolute quantifications determined from Northern blots were then compared to read numbers generated following cDNA library sequencing using either HD or Illumina adapters for each microRNA (Figure 6.8). Analysis revealed that libraries made using HD adapters correlated better with the cellular abundance of each microRNA as determined by quantitative Northern blot (R^2=0.70) than libraries made with Illumina adapters (R^2=0.12). For example, using Illumina adapters, miR-29b had 2,002 reads per million (RPM) whereas using HD adapters the read number was calculated as 74,339 RPM in the same sample. The high cellular abundance of miR-29b, quantified via Northern blot as 1.68 femtomoles/µg total RNA, therefore reflected HD adapter data better. However, absolute quantification for one of the microRNAs, miR-10a, correlated better with Illumina adapter-generated reads than with HD adapter-generated reads. Overall, this result demonstrated that HD adapters are better at detecting small RNAs than commercially available adapters and will therefore be more effective in detecting novel small RNAs in future studies.
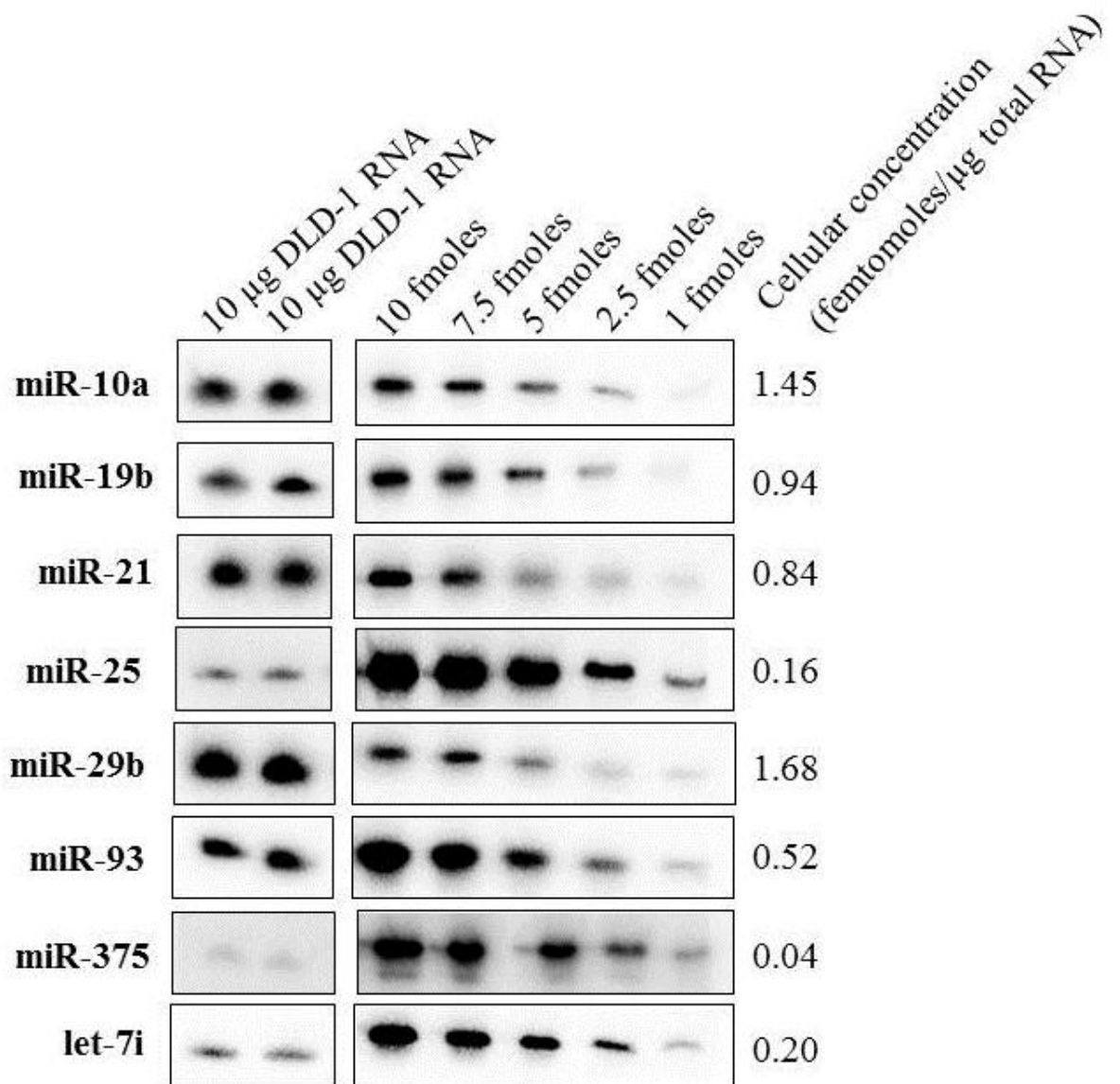
Figure 6.7 Quantification of known microRNAs in DLD-1 cells (epithelial adenocarcinoma cells).

Total RNA (10μg) from DLD-1 cells was analysed via Northern blot in two different lanes. Serial dilutions of synthetic oligonucleotide sequences corresponding with the microRNA of interest were also generated for each of miR-10a, miR-19b, miR-21, miR-25, miR-29b, miR-93, miR-375 and let-7i. A calibration curve was then created and used to quantify the endogenous level of the corresponding microRNA in the two DLD-1 samples. The average calculated concentration for each microRNA is shown as femtomoles/μg total RNA. For miRNA quantification, both experimental and serial dilution membranes were hybridised together and signal was developed and exposed under the same conditions.
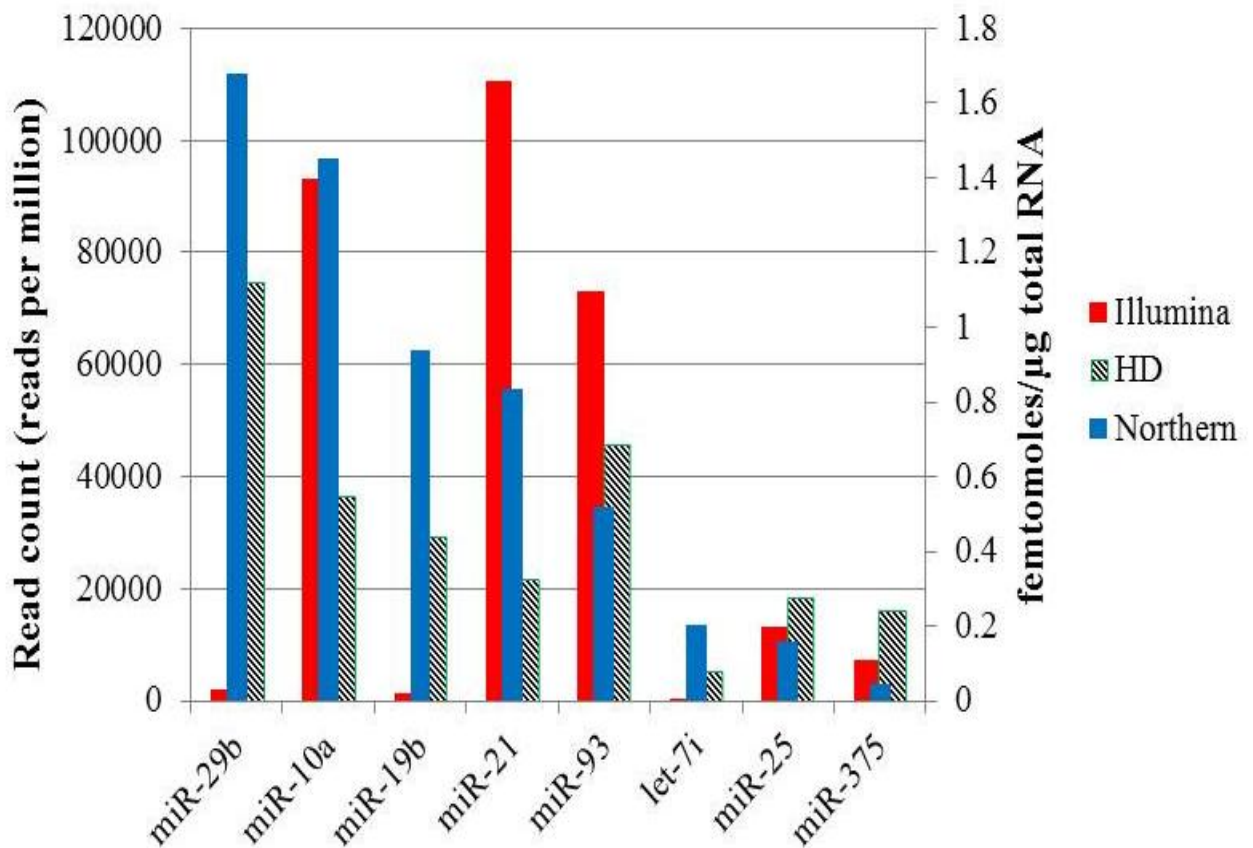
Figure 6.8 Comparison of HD and Illumina adapter-derived read numbers with Northern blot-quantification of microRNAs.

Absolute quantification of eight known microRNAs was achieved by Northern blot and values were then compared with the number of times these microRNAs were sequenced using either Illumina or HD adapters in the DLD-1 cell line. The data obtained using HD adapters (black stripes) correlated better with absolute quantification (blue) (R^2=0.70) than with Illumina adapters (red) (R^2=0.12). The *y*-axis on the left represents the normalised read numbers calculated for HD and Illumina adapter samples. The *y*-axis on the right represents the absolute quantification values for the Northern blot data.

6.3   Discussion

The 30-40 nt sRNAs revealed in this chapter are 'novel' in the sense that they are Ro60 dependent and this has not been reported previously. However the general class of 30-40 nt RNAs is itself not novel as small RNAs of this size have been reported in multiple studies in recent years. RNA fragments from this 30-40 nt class have been shown to be primarily derived from tRNAs (Thompson *et al.*, 2008; Fu *et al.*, 2009; Yamasaki *et al.*, 2009), snoRNAs (Taft *et al.*, 2009) and rRNAs (Li *et al.*, 2012). The function of some of these molecules has been revealed. For example, tRNA halves have been demonstrated to modulate translation of certain transcripts following stress (Ivanov *et al.*, 2011), whilst other 20-40 nt RNAs from tRNAs, rRNAs and snRNAs can associate with a long form of tRNA endoribonuclease Z (tRNase Z^L) and direct silencing of targets (Elbarbary *et al.*, 2009). Indeed some 30-40 nt sRNA fragments still have the potential to bind Ago, and could then be trimmed to become functional siRNAs (Tuck & Tollervey, 2011).

In the work done here, rosRNAs were found to primarily map to the exonic regions of genes, which is in contrast to the majority of intronic derived sRNAs discovered recently (Li *et al.*, 2012; Ruwe & Schmitz-Linneweber, 2012; Skreka *et al.*, 2012). Nonetheless, there are several studies which have also uncovered small RNAs which are derived from exonic regions. Capped analysis of gene expression (CAGE) experiments have shown that RNA transcripts can undergo cleavage and secondary capping (Fejes-Toth *et al.*, 2009). Many of these reported mRNA-derived sRNAs mapped to internal exons with some crossing exon-exon junctions.

These small RNAs were proposed to arise due to processing of mRNAs whereby spliced transcripts undergo continued cleavage and re-capping, generating a hierarchy of increasingly smaller fragments (Fejes-Toth *et al.*, 2009). Another similar study found evidence of intraexonic post-transcriptional cleavage which generated 15-30 nt RNAs (Mercer *et al.*, 2010). Furthermore, these fragments showed tissue and developmental stage specific expression which was not Dicer or Ago dependent (Mercer *et al.*, 2010). The study showed that small RNAs produced as a result of this post-transcriptional cleavage did not appear to share any common structural or sequence motifs.

Recently, two models have been proposed which explain how some of these mRNA-derived sRNAs are produced. Transcription start site-associated (TSSa) RNAs arise as a result of RNA polymerase II (RNAPII) stalling, and subsequent exonuclease processing by XRN1 and XRN2, generating RNAPII 'footprints' (Valen *et al.*, 2011). The same study also proposed that splice site-associated (SSa) RNAs are generated owing to decay of upstream exons which fail to undergo splicing correctly.

It seems unlikely that rosRNAs are produced owing to RNAPII stalling as TSSa RNAs are typically ~17 nt long, reflecting the size of the RNA which fits inside RNAPII (Valen *et al.*, 2011), and the rosRNAs are 30-40 nt. If the rosRNAs are produced as a result of splicing errors then this would imply that Ro60 has a role in transcript maturation. Indeed, preliminary mapping analysis indicated that rosRNAs are derived from the end of exons where splicing occurs. It is possible that when cells undergo stress and subsequent cell death, Ro60 might be recruited

to sRNAs which arise due to aberrant splicing errors because these may resemble misfolded RNAs which Ro60 has been shown to regulate (O'Brien & Wolin, 1994).

Ro60 displays increased nuclear localisation following stress treatment (Sim *et al.*, 2009). As the rosRNAs are essentially mRNA fragments, and could therefore be remnants of transcript splicing, one exciting hypothesis is that Ro60 might enter the nucleus following stress and modulate splicing of a subset of genes.

Some of the exon derived rosRNAs were derived specifically from the untranslated regions of transcripts. One possible role for Ro60 in these transcripts could be that Ro60 masks a cellular localisation signal in the UTR which is modulated during stress. Ro60 binding would affect whether this signal was protected or cleaved and destroyed by RNases. This in turn would control and direct where these transcripts are translated in the cell. Looking at the list of UTR derived candidates (rosRNAs-2, 3 and 6), the genes seem to have diverse roles, including in stress response. For example, the ubiquitin specific peptidase 47 (from which rosRNA-6 is derived) catalyses the removal of ubiquitin in order to facilitate cell survival (Peschiaroli *et al.*, 2010). Relocation of the translation of this protein due to modulated Ro60 binding might be a response to stress.

As with YsRNAs, rosRNAs might be produced as a result of Ro60-mediated protection from RNases upregulated during cell stress. This assumption is not only based on YsRNA biogenesis, but also on the biogenesis of other recently described sRNAs (Valen *et al.*, 2011; Ruwe & Schmitz-Linneweber, 2012). This hypothesis would imply that the transcripts from which rosRNAs are produced are naturally bound by Ro60, or perhaps only during poly(I:C) treatment. Furthermore, the string of T nucleotides at the 5' end of rosRNAs implies the

presence of a potential La binding site, and this might explain the variable 5' end seen on rosRNA presence plots; differential La binding will produce transcripts of different length owing to protection/de-protection as seen with Y5sRNA-3'.

The rosRNAs need to be analysed by Northern blot to validate them as specifically produced sequences and not mere degradation products. One other possible experiment could be to synthesise and transfect these small RNAs with biotin tags and then immunoprecipitate them to investigate possible protein-RNA interactions. This might uncover components of any potential biological pathway in which the rosRNAs are involved. Ro60 binding should also be validated using similar experiments.

Although work into the mouse organ 30-40 nt sRNAs is ongoing, it is interesting to note that the sequences identified in this experiment were also derived from exons. Individual sequences showed highly varied abundances between organs which was conserved between replicates. The brain and kidneys had the highest number of 30-40 nt sRNAs, with the heart having the lowest. This also reflects the complexity of these organs. The differential pattern of production of these sRNAs between organs implies functionality.

Poly(I:C) treatment of human cells results in the increased production of a population of RNAs which are principally 31 nt in length (see Figure 3.2). However, in mouse cells, the population of small RNAs (in either stressed or non-stressed cells) appear to have a principal size of 33 nt (Figures 6.2 and 6.6). It is therefore possible that the size of stress related small RNAs is species specific.

In this chapter, techniques were developed and optimised so that identification of novel sRNAs will be more efficient in the future. The Illumina

protocol was modified to better select for 30-40 nt sRNAs. Interestingly, during library preparation it was noted that several sized bands were present following gel electrophoresis of cDNAs which were not adapter-adapter, microRNA or 30-40 nt associated. These could represent even more distinct sRNA size classes which have yet to be investigated.

The work done here also contributed to the development of HD adapters which significantly reduce sequencing bias of small RNAs, as demonstrated in figure 6.8. The data in this chapter contributed to the publication of this technique by our laboratory (Sorefan *et al.*, 2012). The HD protocol could double read coverage and uncover previously unidentified microRNAs (Sorefan *et al.*, 2012). HD adapters serve as a valuable tool for future investigations into small RNAs.

Chapter 7

Summary and general discussion

The aim of this thesis was to characterise the biogenesis of small RNA fragments produced from Y RNA molecules (YsRNAs). The purpose of this characterisation was to enable further understanding of an emerging class of small RNAs which are slightly longer than traditional gene silencing small RNAs.

In chapter 3, it was found that YsRNAs were produced from both the 5' and 3' ends of all Y RNA molecules in a cancer-independent manner. This is similar to how other recently discovered small RNAs are produced, some of which are also derived from the ends of longer molecules (Li *et al.*, 2012). Although YsRNAs were produced at high levels following stress treatment, they were also present in unstressed cells at levels which are similar to microRNAs.

In chapter 4, Dicer knockout, Ago immunoprecipitation and anion exchange chromatography experiments showed that the YsRNAs did not appear to have gene silencing potential. This indicated that YsRNAs do not fall into the same category as small RNA fragments derived from miRNA precursors, tRNAs and snoRNAs which do have gene silencing activity (Ender *et al.*, 2008; Li *et al.*, 2012).

In chapter 5, potential enzymes involved in YsRNA biogenesis were investigated. A natural assumption was that the enzyme which produces similar sized fragments from tRNAs, angiogenin, was responsible for YsRNA production. This was shown not to be the case. RNase L did appear to contribute to YsRNA production, but only in mouse. However, using a variety of techniques in mouse and human cells, the autoimmune protein Ro60 was shown to be essential for YsRNA generation. The La protein also seemed to affect the production of one of the YsRNAs. Experiments from this chapter appeared to support a model of YsRNA biogenesis through RNase protection by Ro60 and indeed, by La. This

model is in line with how some other novel small RNAs have been shown to be made (Valen *et al.*, 2011; Ruwe & Schmitz-Linneweber, 2012).

In chapter 6, after establishing that Ro60 was integral to YsRNA biogenesis, it was investigated whether other small RNAs in the same 30-40 nt size class were also Ro60 dependent. This was indeed found to be the case, and deep sequencing experiments led to the discovery of a new, highly abundant class of Ro60 dependent small RNAs (rosRNAs). This chapter demonstrated that analysis of the biogenesis of YsRNAs has resulted in a better understanding of the small RNA class in which it resides. Not only was this one of the major aims of the thesis, but it increased the impact of the work across many different research disciplines, beyond the Y RNA community. Indeed, the additional discovery in chapter 6 of other exon derived small RNAs which are differentially produced in mouse organs would be of interest to developmental biologists and molecular geneticists alike.

Although this work did not uncover the biological function of YsRNAs, it has uncovered a class of other related molecules, the rosRNAs. Therefore, studying these rosRNAs might facilitate the discovery of a biological function in the near future.

Following the RNA 'renaissance' which was marked by the discovery of gene silencing related molecules, sequencing technologies have unveiled a plethora of other small RNAs which were once mistakenly considered to be degradation products. The last part of this thesis involved testing a new deep sequencing protocol developed in the laboratory using biological samples. The 'HD' protocol significantly reduced sequencing bias and better reflected absolute quantification of small RNAs. As a result, HD adapters will now facilitate the discovery of other

small RNAs in the future. Studies have already managed to determine the function of some of these small RNAs, such as tRNA halves (Ivanov *et al.*, 2011). However others, such as YsRNAs, remain elusive in their function. The small RNA journey of discovery continues.

References

Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., & Axtell, M. J. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current Biology* **18:** 758-762.

Aftab, M. N., Ikram-ul-Haq, & Javed, M. M. (2012) Depletion of Y associated proteins by RNAi down-regulates the Y RNA in Caenorhabditis elegans. *Pakistan Journal of Botany* **44:** 199-202.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Asselbergs, F. A., & Widmer, R. (2003) Rapid detection of apoptosis through real-time reverse transcriptase polymerase chain reaction measurement of the small cytoplasmic RNA Y1. *Analytical Biochemistry* **318:** 221-229.

Bai, C., Li, Z., & Tolias, P. (1994) Developmental Characterization of a Drosophila Rna-Binding Protein Homologous to the Human Systemic Lupus Erythematosus-Associated La/ss-B Autoantigen. *Molecular and cellular biology* **14:** 5123-5129.

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

Ben-Chetrit, E., Chan, E. K., Sullivan, K. F., & Tan, E. M. (1988) A 52-kD protein is a novel component of the SS-A/Ro antigenic particle. *The Journal of experimental medicine* **167:** 1560-1571.

Bernstein, B., & ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57-74.

Birney, E., & ENCODE Project Consortium. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799-816.

Bohnsack, M. T., Czaplinski, K., & Gorlich, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA (New York, N.Y.)* **10:** 185-191.

Boire, G., & Craft, J. (1990) Human Ro ribonucleoprotein particles: characterization of native structure and stable association with the La polypeptide. *The Journal of clinical investigation* **85:** 1182-1190.

Boire, G., & Craft, J. (1989) Biochemical and Immunological Heterogeneity of the Ro Ribonucleoprotein-Particles - Analysis with Sera Specific for the Rohy5 Particle. *Journal of Clinical Investigation* **84:** 270-279.

Boire, G., Gendron, M., Monast, N., Bastin, B., & Menard, H. A. (1995) Purification of antigenically intact Ro ribonucleoproteins; biochemical and immunological evidence that the 52-kD protein is not a Ro protein. *Clinical and experimental immunology* **100:** 489-498.

Borer, R. A., Lehner, C. F., Eppenberger, H. M., & Nigg, E. A. (1989) Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell* **56:** 379-390.

Boria, I., Gruber, A. R., Tanzer, A., Bernhart, S. H., Lorenz, R., Mueller, M. M., Hofacker, I. L., & Stadler, P. F. (2010) Nematode sbRNAs: homologs of vertebrate Y RNAs. *Journal of Molecular Evolution* **70:** 346-358.

Bouffard, P., Barbar, E., Briere, F., & Boire, G. (2000) Interaction cloning and characterization of RoBPI, a novel protein binding to human Ro ribonucleoproteins. *RNA (New York, N.Y.)* **6:** 66-78.

Boulanger, C., Chabot, B., Menard, H. A., & Boire, G. (1995) Autoantibodies in Human Anti-Ro Sera Specifically Recognize Deproteinized Hy5 Ro Rna. *Clinical and experimental immunology* **99:** 29-36.

Bracken, C. P., Szubert, J. M., Mercer, T. R., Dinger, M. E., Thomson, D. W., Mattick, J. S., Michael, M. Z., & Goodall, G. J. (2011) Global analysis of the mammalian RNA degradome reveals widespread miRNA-dependent and miRNA-independent endonucleolytic cleavage. *Nucleic acids research* **39:** 5658-5668.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., & Gruber, J. (2011) Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic acids research* **39:** 675-686.

Brennecke, J., Stark, A., Russell, R. B., & Cohen, S. M. (2005) Principles of microRNA-target recognition. *PLoS biology* **3:** e85.

Brunner, E., Ahrens, C. H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., Deutsch, E. W., Panse, C., de Lichtenberg, U., Rinner, O. *et al. &* Aebersold, R. (2007) A high-quality catalog of the Drosophila melanogaster proteome. *Nature biotechnology* **25:** 576-583.

Burroughs, A. M., Ando, Y., de Hoon, M. J. L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., & Daub, C. O. (2011) Deep-sequencing of human argonaute-associated small RNAs provides insight into miRNA sorting and reveals argonaute association with RNA fragments of diverse origin. *Rna Biology* **8:** 158-177.

Casciola-Rosen, L. A., Anhalt, G., & Rosen, A. (1994) Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *The Journal of experimental medicine* **179:** 1317-1330.

Chambers, J., Kenan, D., Martin, B., & Keene, J. (1988) Genomic Structure and Amino-Acid Sequence Domains of the Human La Auto-Antigen. *Journal of Biological Chemistry* **263:** 18043-18051.

Cheloufi, S., Dos Santos, C. O., Chong, M. M., & Hannon, G. J. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465:** 584-589.

Chen, X., & Wolin, S. L. (2004) The Ro 60 kDa autoantigen: insights into cellular function and role in autoimmunity. *Journal of Molecular Medicine (Berlin, Germany)* **82:** 232-239.

Chen, X., Quinn, A. M., & Wolin, S. L. (2000) Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation. *Genes & development* **14:** 777-782.

Chen, X., Smith, J. D., Shi, H., Yang, D. D., Flavell, R. A., & Wolin, S. L. (2003) The Ro autoantigen binds misfolded U2 small nuclear RNAs and assists mammalian cell survival after UV irradiation. *Current biology : CB* **13:** 2206-2211.

Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., & Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436:** 740-744.

Christov, C. P., Trivier, E., & Krude, T. (2008) Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *British journal of cancer* **98:** 981-988.

Christov, C. P., Gardiner, T. J., Szuts, D., & Krude, T. (2006) Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Molecular and cellular biology* **26:** 6993-7004.

Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G. J., Lawson, N., Wolfe, S., & Giraldez, A. J. (2010) A Novel miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. *Science (New York, N.Y.)*

Cole, C., Sobala, A., Lu, C., Thatcher, S. R., Bowman, A., Brown, J. W. S., Green, P. J., Barton, G. J., & Hutvagner, G. (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *Rna-a Publication of the Rna Society* **15:** 2147-2160.

Collart, C., Christov, C. P., Smith, J. C., & Krude, T. (2011) The Midblastula Transition Defines the Onset of Y RNA-Dependent DNA Replication in Xenopus laevis. *Molecular and cellular biology* **31:** 3857-3870.

Collins, F., & International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931-945.

Dalmay, T. (2008) MicroRNAs and cancer. *Journal of internal medicine* **263:** 366-375.

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

Deutscher, S., Harley, J., & Keene, J. (1988) Molecular Analysis of the 60-Kda Human Ro Ribonucleoprotein. *Proceedings of the National Academy of Sciences of the United States of America* **85:** 9479-9483.

Dhar, S. K., & St Clair, D. K. (2009) Nucleophosmin Blocks Mitochondrial Localization of p53 and Apoptosis. *Journal of Biological Chemistry* **284:** 16409-16418.

Djuranovic, S., Nahvi, A., & Green, R. (2012) miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. *Science* **336:** 237-240.

Doma, M. K., & Parker, R. (2007) RNA quality control in eukaryotes. *Cell* **131:** 660-668.

Elbarbary, R. A., Takaku, H., Uchiumi, N., Tamiya, H., Abe, M., Takahashi, M., Nishida, H., & Nashimoto, M. (2009) Modulation of Gene Expression by Human Cytosolic tRNase Z(L) through 5 '-Half-tRNA. *Plos One* **4:** e5908.

Elbashir, S., Lendeckel, W., & Tuschl, T. (2001) RNA interference is mediated by 21-and 22-nucleotide RNAs. *Genes & development* **15:** 188-200.

Elmen, J., Lindow, M., Schutz, S., Lawrence, M., Petri, A., Obad, S., Lindholm, M., Hedtjarn, M., Hansen, H. F., Berger, U., Gullans, S., Kearney, P., Sarnow, P., Straarup, E. M., & Kauppinen, S. (2008) LNA-mediated microRNA silencing in non-human primates. *Nature* **452:** 896-899.

Ender, C., Krek, A., Friedlaender, M. R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., & Meister, G. (2008) A Human snoRNA with MicroRNA-Like Functions. *Molecular cell* **32:** 519-528.

Fabian, M. R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J. T., Svitkin, Y. V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J. A., Chen, C. Y., Shyu,

A. B., Yates, J. R.,3rd, Hannon, G. J., Filipowicz, W., Duchaine, T. F., & Sonenberg, N. (2009) Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Molecular cell* **35:** 868-880.

Fabini, G., Rutjes, S. A., Zimmermann, C., Pruijn, G. J., & Steiner, G. (2000) Analysis of the molecular composition of Ro ribonucleoprotein complexes. Identification of novel Y RNA-binding proteins. *European journal of biochemistry / FEBS* **267:** 2778-2789.

Fabini, G., Raijmakers, R., Hayer, S., Fouraux, M. A., Pruijn, G. J., & Steiner, G. (2001) The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the ro ribonucleoprotein-associated Y RNAs in vitro and in vivo. *The Journal of biological chemistry* **276:** 20711-20718.

Fan, H., Goodier, J. L., Chamberlain, J. R., Engelke, D. R., & Maraia, R. J. (1998) 5' processing of tRNA precursors can Be modulated by the human La antigen phosphoprotein. *Molecular and cellular biology* **18:** 3201-3211.

Farris, A. D., O'Brien, C. A., & Harley, J. B. (1995) Y3 is the most conserved small RNA component of Ro ribonucleoprotein complexes in vertebrate species. *Gene* **154:** 193-198.

Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G., Kapranov, P., Foissac, S., Willingham, A., Duttagupta, R., Dumais, E., & Gingeras, T. (2009) Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature* **457:** 1028-1032.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature reviews.Genetics* **9:** 102-114.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806-811.

Fouraux, M. A., Bouvet, P., Verkaart, S., van Venrooij, W. J., & Pruijn, G. J. (2002) Nucleolin associates with a subset of the human Ro ribonucleoprotein complexes. *Journal of Molecular Biology* **320:** 475-488.

Francoeur, A., Chan, E., Garrels, J., & Mathews, M. (1985) Characterization and Purification of Lupus Antigen La, an Rna-Binding Protein. *Molecular and cellular biology* **5:** 586-590.

Frank, M., McCubbin, V., & Heldermon, C. (1995) Expression and Dna-Binding of the Human 52-Kda Ro/ssa Autoantigen. *Biochemical Journal* **305:** 359-362.

Frohn, A., Eberl, H. C., Stohr, J., Glasmacher, E., Rudel, S., Heissmeyer, V., Mann, M., & Meister, G. (2012) Dicer-dependent and -independent Argonaute2

protein interaction networks in mammalian cells. *Molecular and cellular proteomics* [Epub ahead of print].

Fu, H., Subramanian, R., & Masters, S. (2000) 14-3-3 proteins: Structure, function, and regulation. *Annual Review of Pharmacology and Toxicology* **40:** 617-647.

Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., Xing, R., Sun, Z., & Zheng, X. (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS letters* **583:** 437-442.

Fuchs, G., Stein, A. J., Fu, C., Reinisch, K. M., & Wolin, S. L. (2006) Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. *Nature structural & molecular biology* **13:** 1002-1009.

Garcia, E. L., Onafuwa-Nuga, A., Sim, S., King, S. R., Wolin, S. L., & Telesnitsky, A. (2009) Packaging of host mY RNAs by murine leukemia virus may occur early in Y RNA biogenesis. *Journal of virology* **83:** 12526-12534.

Gardiner, T. J., Christov, C. P., Langley, A. R., & Krude, T. (2009) A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. *RNA (New York, N.Y.)* **15:** 1375-1385.

Garmire, L. X., & Subramaniam, S. (2012) Evaluation of normalization methods in mammalian microRNA-Seq data. *Rna-a Publication of the Rna Society* **18:** 1279-1288.

Gendron, M., Roberge, D., & Boire, G. (2001) Heterogeneity of human Ro ribonucleoproteins (RNPS): nuclear retention of Ro RNPS containing the human hY5 RNA in human and mouse cells. *Clinical and experimental immunology* **125:** 162-168.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R., Belle, A., Dephoure, N., O'Shea, E., & Weissman, J. (2003) Global analysis of protein expression in yeast. *Nature* **425:** 737-741.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., & Zamore, P. D. (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA (New York, N.Y.)* **16:** 43-56.

Gilbert, W. (1986) Origin of Life - the RNA World. *Nature* **319:** 618-618.

Ginisty, H., Sicard, H., Roger, B., & Bouvet, P. (1999) Structure and functions of nucleolin. *Journal of cell science* **112:** 761-772.

Gottlieb, E., & Steitz, J. A. (1989) Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. *The EMBO journal* **8:** 851-861.

Gregory, R. I., Chendrimada, T. P., Cooch, N., & Shiekhattar, R. (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123:** 631-640.

Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:** 835-U66.

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J. T. G., Nusbaum, J. D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., & Tuschl, T. (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *Rna-a Publication of the Rna Society* **17:** 1697-1712.

Hamilton, A. J., & Baulcombe, D. C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science (New York, N.Y.)* **286:** 950-952.

Hanahan, D., & Weinberg, R. A. (2000) The hallmarks of cancer. *Cell* **100:** 57-70.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A. Z., & Kay, M. A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *Rna-a Publication of the Rna Society* **16:** 673-695.

Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R., & Steitz, J. A. (1981) Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Molecular and cellular biology* **1:** 1138-1149.

Hogg, J. R., & Collins, K. (2007) Human Y5 RNA specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. *Genes & development* **21:** 3067-3072.

Holcik, M., & Korneluk, R. G. (2000) Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation. *Molecular and cellular biology* **20:** 4648-4657.

Huttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G., Condeelis, J., & Singer, R. (2005) Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* **438:** 512-515.

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Menard, S., Palazzo, J. P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G. A., Querzoli, P., Negrini, M., & Croce, C. M. (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer research* **65:** 7065-7070.

Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P., & Anderson, P. (2011) Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Molecular cell* **43:** 613-623.

Jalali, S., Jayaraj, G., & Scaria, V. (2012) Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. *Biology Direct* **7:** [Epub ahead of print].

Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., Minna, J. D., & Corey, D. R. (2006) Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Structural & Molecular Biology* **13:** 787-792.

Kawashima, C. G., Matthewman, C. A., Huang, S., Lee, B., Yoshimoto, N., Koprivova, A., Rubio-Somoza, I., Todesco, M., Rathjen, T., Saito, K., Takahashi, H., Dalmay, T., & Kopriva, S. (2011) Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in Arabidopsis. *Plant Journal* **66:** 863-876.

Kawashima, C. G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y. N., Saito, K., Takahashi, H., & Dalmay, T. (2009) Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant Journal* **57:** 313-321.

Kelekar, A., Saitta, M. R., & Keene, J. D. (1994) Molecular composition of Ro small ribonucleoprotein complexes in human cells. Intracellular localization of the 60- and 52-kD proteins. *The Journal of clinical investigation* **93:** 1637-1644.

Krude, T., Christov, C. P., Hyrien, O., & Marheineke, K. (2009) Y RNA functions at the initiation step of mammalian chromosomal DNA replication. *Journal of cell science* **122:** 2836-2845.

Krude, T., Jackman, M., Pines, J., & Laskey, R. A. (1997) Cyclin/Cdk-dependent initiation of DNA replication in a human cell-free system. *Cell* **88:** 109-119.

Krude, T. (2006) Initiation of chromosomal DNA replication in mammalian cell-free systems. *Cell Cycle* **5:** 2115-2122.

Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D., & Tollervey, D. (2000) Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Molecular and cellular biology* **20:** 5415-5424.

Labbe, J. C., Hekimi, S., & Rokeach, L. A. (1999) The levels of the RoRNP-associated Y RNA are dependent upon the presence of ROP-1, the Caenorhabditis elegans Ro60 protein. *Genetics* **151:** 143-150.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., & Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyade nylation complex. *Cell* **121:** 713-724.

Lander, E., & International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Langley, A. R., Chambers, H., Christov, C. P., & Krude, T. (2010) Ribonucleoprotein particles containing non-coding Y RNAs, Ro60, La and nucleolin are not required for Y RNA function in DNA replication. *PloS one* **5:** e13673.

Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science (New York, N.Y.)* **294:** 858-862.

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

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., & Kim, V. N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* **23:** 4051-4060.

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

Lerner, M. R., & Steitz, J. A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences of the United States of America* **76:** 5495-5499.

Lerner, M. R., Boyle, J. A., Hardin, J. A., & Steitz, J. A. (1981) Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science (New York, N.Y.)* **211:** 400-402.

Leung, A. K. L., & Sharp, P. A. (2010) MicroRNA Functions in Stress Responses. *Molecular cell* **40:** 205-215.

Levitz, R., Chapman, D., Amitsur, M., Green, R., Snyder, L., & Kaufmann, G. (1990) The Optional Escherichia-Coli Prr Locus Encodes a Latent Form of Phage T4-Induced Anticodon Nuclease. *Embo Journal* **9:** 1383-1389.

Li, Z., Ender, C., Meister, G., Moore, P. S., Chang, Y., & John, B. (2012) Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic acids research* **40:** 6787-6799.

Lin-Marq, N., & Clarkson, S. (1998) Efficient synthesis, termination and release of RNA polymerase III transcripts in Xenopus extracts depleted of La protein. *Embo Journal* **17:** 2033-2041.

Linsen, S. E. V., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., Fritz, B., Wyman, S. K., de Bruijn, E., Voest, E. E., Kuersten, S., Tewari, M., & Cuppen, E. (2009) Limitations and possibilities of small RNA digital gene expression profiling. *Nature Methods* **6:** 474-476.

Liu, Y., Tan, H., Tian, H., Liang, C., Chen, S., & Liu, Q. (2011) Autoantigen La Promotes Efficient RNAi, Antiviral Response, and Transposon Silencing by Facilitating Multiple-Turnover RISC Catalysis. *Molecular cell* **44:** 502-508.

Lucas, R., McMichael, T., Smith, W., & Armstrong, B. (2006) *Solar ultraviolet radiation: Global burden of disease from solar ultraviolet radiation.* Geneva, World Health Organization, 258 pp.

Mamula, M., Silverman, E., Laxer, R., Bentur, L., Isacovics, B., & Hardin, J. (1989) Human Monoclonal Anti-La Antibodies - the La Protein Resides on a Subset of Ro Particles. *Journal of Immunology* **143:** 2923-2928.

Maraia, R., Sakulich, A. L., Brinkmann, E., & Green, E. D. (1996) Gene encoding human Ro-associated autoantigen Y5 RNA. *Nucleic acids research* **24:** 3552-3559.

Maraia, R. J., Kenan, D. J., & Keene, J. D. (1994a) Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III. *Molecular and cellular biology* **14:** 2147-2158.

Maraia, R. J., Sasaki-Tozawa, N., Driscoll, C. T., Green, E. D., & Darlington, G. J. (1994b) The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. *Nucleic acids research* **22:** 3045-3052.

Mattick, J. S., & Makunin, I. V. (2006) Non-coding RNA. *Human molecular genetics* **15 Spec No 1:** R17-29.

Meerovitch, K., Svitkin, Y., Lee, H., Lejbkowicz, F., Kenan, D., Chan, E., Agol, V., Keene, J., & Sonenberg, N. (1993) La Autoantigen Enhances and Corrects Aberrant Translation of Poliovirus Rna in Reticulocyte Lysate. *Journal of virology* **67:** 3798-3807.

Meiri, E., Levy, A., Benjamin, H., Ben-David, M., Cohen, L., Dov, A., Dromi, N., Elyakim, E., Yerushalmi, N., Zion, O., Lithwick-Yanai, G., & Sitbon, E. (2010) Discovery of microRNAs and other small RNAs in solid tumors. *Nucleic acids research* **38:** 6234-6246.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell* **15:** 185-197.

Mercer, T. R., Dinger, M. E., Bracken, C. P., Kolle, G., Szubert, J. M., Korbie, D. J., Askarian-Amiri, M. E., Gardiner, B. B., Goodall, G. J., Grimmond, S. M., & Mattick, J. S. (2010) Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. *Genome research* **20:** 1639-1650.

Metzker, M. L. (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* **11:** 31-46.

Mosig, A., Guofeng, M., Stadler, B. M., & Stadler, P. F. (2007) Evolution of the vertebrate Y RNA cluster. *Theory in biosciences* **126:** 9-14.

Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D. J., & Moulton, V. (2008) A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* **24:** 2252-2253.

Nandakumar, J., & Shuman, S. (2004) How an RNA ligase discriminates RNA versus DNA damage. *Molecular cell* **16:** 211-221.

Narayanan, D. L., Saladi, R. N., & Fox, J. L. (2010) Review: Ultraviolet radiation and skin cancer. *International Journal of Dermatology* **49:** 978-986.

Nicolas, F. E., Hall, A. E., Csorba, T., Turnbull, C., & Dalmay, T. (2012) Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway. *FEBS letters* **586:** 1226-1230.

Nolte-'t Hoen, E. N., Buermans, H. P., Waasdorp, M., Stoorvogel, W., Wauben, M. H., & 't Hoen, P. A. (2012) Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. *Nucleic acids research* **40:** 9272-9285.

O'Brien, C. A., & Wolin, S. L. (1994) A possible role for the 60-kD Ro autoantigen in a discard pathway for defective 5S rRNA precursors. *Genes & development* **8:** 2891-2903.

O'Brien, C. A., & Harley, J. B. (1990) A subset of hY RNAs is associated with erythrocyte Ro ribonucleoproteins. *The EMBO journal* **9:** 3683-3689.

Ohlsson, M., Jonsson, R., & Brokstad, K. A. (2002) Subcellular redistribution and surface exposure of the Ro52, Ro60 and La48 autoantigens during apoptosis in human ductal epithelial cells: a possible mechanism in the pathogenesis of Sjogren's syndrome. *Scandinavian journal of immunology* **56:** 456-469.

Okamura, K., Ishizuka, A., Siomi, H., & Siomi, M. (2004) Distinct roles for argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & development* **18:** 1655-1666.

Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L., & Hamilton, A. (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic acids research* **35:** e60.

Pederson, T. (2010) Regulatory RNAs derived from transfer RNA? *RNA (New York, N.Y.)* **16:** 1865-1869.

Peek, R., Pruijn, G. J., van der Kemp, A. J., & van Venrooij, W. J. (1993) Subcellular distribution of Ro ribonucleoprotein complexes and their constituents. *Journal of cell science* **106 (Pt 3):** 929-935.

Perreault, J., Perreault, J. P., & Boire, G. (2007) Ro-associated Y RNAs in metazoans: evolution and diversification. *Molecular biology and evolution* **24:** 1678-1689.

Perreault, J., Noel, J. F., Briere, F., Cousineau, B., Lucier, J. F., Perreault, J. P., & Boire, G. (2005) Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. *Nucleic acids research* **33:** 2032-2041.

Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A., & Rovira, C. (2009) The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nature cell biology* **11:** 1268-U265.

Peschiaroli, A., Skaar, J. R., Pagano, M., & Melino, G. (2010) The ubiquitin-specific protease USP47 is a novel beta-TRCP interactor regulating cell survival. *Oncogene* **29:** 1384-1393.

Pfeifle, J., Anderer, F., & Franke, M. (1987) Multiple Phosphorylation of Human Ss-B/la Autoantigen and its Effect on Poly(u) and Autoantibody Binding. *Biochimica et biophysica acta* **928:** 217-226.

Pospiech, H., Grosse, F., & Pisani, F. M. (2010) The Initiation Step of Eukaryotic DNA Replication. *Subcellular Biochemistry* **50:** 79-104.

Pruijn, G. J., Slobbe, R. L., & van Venrooij, W. J. (1991) Analysis of protein--RNA interactions within Ro ribonucleoprotein complexes. *Nucleic acids research* **19:** 5173-5180.

Pruijn, G. J., Wingens, P. A., Peters, S. L., Thijssen, J. P., & van Venrooij, W. J. (1993) Ro RNP associated Y RNAs are highly conserved among mammals. *Biochimica et biophysica acta* **1216:** 395-401.

Rao, P., & Johnson, R. (1970) Mammalian Cell Fusion . Studies on Regulation of Dna Synthesis and Mitosis. *Nature* **225:** 159-164.

Rinke, J., & Steitz, J. A. (1982) Precursor molecules of both human 5S ribosomal RNA and transfer RNAs are bound by a cellular protein reactive with anti-La lupus antibodies. *Cell* **29:** 149-159.

Rosa, M., Gottlieb, E., Lerner, M., & Steitz, J. (1981) Striking Similarities are Exhibited by 2 Small Epstein-Barr Virus-Encoded Ribonucleic-Acids and the Adenovirus-Associated Ribonucleic-Acids Vai and Vaii. *Molecular and cellular biology* **1:** 785-796.

Ross, A., Oleynikov, Y., Kislauskis, E., Taneja, K., & Singer, R. (1997) Characterization of a beta-actin mRNA zipcode-binding protein. *Molecular and cellular biology* **17:** 2158-2165.

Rother, S., & Meister, G. (2011) Small RNAs derived from longer non-coding RNAs. *Biochimie* **93:** 1905-1915.

Ruby, J. G., Jan, C. H., & Bartel, D. P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* **448:** 83-86.

Rutjes, S. A., van der Heijden, A., Utz, P. J., van Venrooij, W. J., & Pruijn, G. J. (1999a) Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. *The Journal of biological chemistry* **274:** 24799-24807.

Rutjes, S. A., Utz, P. J., van der Heijden, A., Broekhuis, C., van Venrooij, W. J., & Pruijn, G. J. (1999b) The La (SS-B) autoantigen, a key protein in RNA biogenesis, is dephosphorylated and cleaved early during apoptosis. *Cell death and differentiation* **6:** 976-986.

Ruwe, H., & Schmitz-Linneweber, C. (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic acids research* **40:** 3106-3116.

Scherly, D., Stutz, F., Linmarq, N., & Clarkson, S. (1993) La Proteins from Xenopus-Laevis - Cdna Cloning and Developmental Expression. *Journal of Molecular Biology* **231:** 196-204.

Sharp, P. A. (2009) The centrality of RNA. *Cell* **136:** 577-580.

Shi, H., OBrien, C. A., VanHorn, D. J., & Wolin, S. L. (1996) A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens. *Rna-a Publication of the Rna Society* **2:** 769-784.

Silverman, R. H. (2007) Viral encounters with 2 ',5 '-oligoadenylate synthetase and RNase L during the interferon antiviral response. *Journal of virology* **81:** 12720-12729.

Sim, S., Weinberg, D. E., Fuchs, G., Choi, K., Chung, J., & Wolin, S. L. (2009) The subcellular distribution of an RNA quality control protein, the Ro autoantigen, is regulated by noncoding Y RNA binding. *Molecular biology of the cell* **20:** 1555-1564.

Sim, S., Yao, J., Weinberg, D. E., Niessen, S., Yates, J. R.,III, & Wolin, S. L. (2012) The zipcode-binding protein ZBP1 influences the subcellular location of the Ro 60-kDa autoantigen and the noncoding Y3 RNA. *Rna-a Publication of the Rna Society* **18:** 100-110.

Simons, F. H., Pruijn, G. J., & van Venrooij, W. J. (1994) Analysis of the intracellular localization and assembly of Ro ribonucleoprotein particles by microinjection into Xenopus laevis oocytes. *The Journal of cell biology* **125:** 981-988.

Simons, F. H., Rutjes, S. A., van Venrooij, W. J., & Pruijn, G. J. (1996) The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. *RNA (New York, N.Y.)* **2:** 264-273.

Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology* **12:** 246-258.

Skreka, K., Schafferer, S., Nat, I., Zywicki, M., Salti, A., Apostolova, G., Griehl, M., Rederstorff, M., Dechant, G., & Huettenhofer, A. (2012) Identification of differentially expressed non-coding RNAs in embryonic stem cell neural differentiation. *Nucleic acids research* **40:** 6001-6015.

Slobbe, R. L., Pluk, W., van Venrooij, W. J., & Pruijn, G. J. (1992) Ro ribonucleoprotein assembly in vitro. Identification of RNA-protein and protein-protein interactions. *Journal of Molecular Biology* **227:** 361-366.

Slobbe, R. L., Pruijn, G. J., Damen, W. G., van der Kemp, J. W., & van Venrooij, W. J. (1991) Detection and occurrence of the 60- and 52-kD Ro (SS-A) antigens and of autoantibodies against these proteins. *Clinical and experimental immunology* **86:** 99-105.

Sommer, G., Dittmann, J., Kuehnert, J., Reumann, K., Schwartz, P. E., Will, H., Coulter, B. L., Smith, M. T., & Heise, T. (2011) The RNA-binding protein La contributes to cell proliferation and CCND1 expression. *Oncogene* **30:** 434-444.

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

Sorefan, K., Pais, H., Hall, A. E., Kozomara, A., Griffiths-Jones, S., Moulton, V., & Dalmay, T. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3:** (4) 1-11.

Sorrentino, S. (2010) The eight human "canonical" ribonucleases: Molecular diversity, catalytic properties, and special biological actions of the enzyme proteins. *FEBS letters* **584:** 2194-2200.

Stefano, J. E. (1984) Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell* **36:** 145-154.

Stein, A. J., Fuchs, G., Fu, C., Wolin, S. L., & Reinisch, K. M. (2005) Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. *Cell* **121:** 529-539.

Taft, R. J., Glazov, E. A., Lassmann, T., Hayashizaki, Y., Carninci, P., & Mattick, J. S. (2009) Small RNAs derived from snoRNAs. *Rna-a Publication of the Rna Society* **15:** 1233-1240.

Takeda, D., & Dutta, A. (2005) DNA replication and progression through S phase. *Oncogene* **24:** 2827-2843.

Tang, F., Lao, K., & Surani, M. A. (2011) Development and applications of single-cell transcriptome analysis. *Nature Methods* **8:** S6-S11.

Taylor, R. C., Cullen, S. P., & Martin, S. J. (2008) Apoptosis: controlled demolition at the cellular level. *Nature reviews.Molecular cell biology* **9:** 231-241.

Teunissen, S. W., Kruithof, M. J., Farris, A. D., Harley, J. B., Venrooij, W. J., & Pruijn, G. J. (2000) Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. *Nucleic acids research* **28:** 610-619.

Thompson, D. M., Lu, C., Green, P. J., & Parker, R. (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *Rna-a Publication of the Rna Society* **14:** 2095-2103.

Topfer, F., Gordon, T., & McCluskey, J. (1993) Characterization of the Mouse Autoantigen La (Ss-B) - Identification of Conserved Rna-Binding Motifs, a Putative Atp Binding-Site and Reactivity of Recombinant Protein with Poly(u) and Human Autoantibodies. *Journal of Immunology* **150:** 3091-3100.

Tuck, A. C., & Tollervey, D. (2011) RNA in pieces. *Trends in Genetics* **27:** 422-432.

Valen, E., Preker, P., Andersen, P. R., Zhao, X., Chen, Y., Ender, C., Dueck, A., Meister, G., Sandelin, A., & Jensen, T. H. (2011) Biogenic mechanisms and

utilization of small RNAs derived from human protein-coding genes. *Nature Structural & Molecular Biology* **18:** 1075-1082.

van Gelder, C. W., Thijssen, J. P., Klaassen, E. C., Sturchler, C., Krol, A., van Venrooij, W. J., & Pruijn, G. J. (1994) Common structural features of the Ro RNP associated hY1 and hY5 RNAs. *Nucleic acids research* **22:** 2498-2506.

Vanhorn, D., Eisenberg, D., Obrien, C., & Wolin, S. (1995) Caenorhabditis-Elegans Embryos Contain Only One Major Species of Ro Rnp. *Rna-a Publication of the Rna Society* **1:** 293-303.

Verhagen, A. P. M., & Pruijn, G. J. M. (2011) Are the Ro RNP-associated Y RNAs concealing microRNAs? Y RNA-derived miRNAs may be involved in autoimmunity. *Bioessays* **33:** 674-682.

Wass, M., McCann, K., & Bagshawe, K. (1978) Isolation of Antibodies to Hcg-Lh from Human Sera. *Nature* **274:** 368-370.

White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L. *et al. &* Fraser, C. M. (1999) Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. *Science (New York, N.Y.)* **286:** 1571-1577.

Wittmann, J., & Jaeck, H. (2010) New Surprises from the Deep - The Family of Small Regulatory RNAs Increases. *Thescientificworldjournal* **10:** 1239-1243.

Wolin, S. L., & Cedervall, T. (2002) The La protein. *Annual Review of Biochemistry* **71:** 375-403.

Wolin, S. L., & Steitz, J. A. (1984) The Ro small cytoplasmic ribonucleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. *Proceedings of the National Academy of Sciences of the United States of America* **81:** 1996-2000.

Wolin, S. L., & Steitz, J. A. (1983) Genes for two small cytoplasmic Ro RNAs are adjacent and appear to be single-copy in the human genome. *Cell* **32:** 735-744.

Wurtmann, E. J., & Wolin, S. L. (2010) A role for a bacterial ortholog of the Ro autoantigen in starvation-induced rRNA degradation. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 4022-4027.

Xue, D., Shi, H., Smith, J. D., Chen, X., Noe, D. A., Cedervall, T., Yang, D. D., Eynon, E., Brash, D. E., Kashgarian, M., Flavell, R. A., & Wolin, S. L. (2003) A lupus-like syndrome develops in mice lacking the Ro 60-kDa protein, a major lupus autoantigen. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 7503-7508.

Xue, D., Rubinson, D., Pannone, B., Yoo, C., & Wolin, S. (2000) U snRNP assembly in yeast involves the La protein. *Embo Journal* **19:** 1650-1660.

Yamasaki, S., Ivanov, P., Hu, G., & Anderson, P. (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. *Journal of Cell Biology* **185:** 35-42.

Yang, S., Tutton, S., Pierce, E., & Yoon, K. (2001) Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. *Molecular and cellular biology* **21:** 7807-7816.

Yoo, C. J., & Wolin, S. L. (1997) The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell* **89:** 393-402.

Yoo, C., & Wolin, S. (1994) La Proteins from Drosophila-Melanogaster and Saccharomyces-Cerevisiae - a Yeast Homolog of the La Autoantigen is Dispensable for Growth. *Molecular and cellular biology* **14:** 5412-5424.

Zhang, A. T., Langley, A. R., Christov, C. P., Kheir, E., Shafee, T., Gardiner, T. J., & Krude, T. (2011) Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *Journal of cell science* **124:** 2058-2069.

Zhou, A., Paranjape, J., Brown, T., Nie, H., Naik, S., Dong, B., Chang, A., Trapp, B., Fairchild, R., Colmenares, C., & Silverman, R. (1997) Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. *Embo Journal* **16:** 6355-6363.

Zhuang, F., Fuchs, R. T., & Robb, G. B. (2012) Small RNA Expression Profiling by High-Throughput Sequencing: Implications of Enzymatic Manipulation. *Journal of Nucleic Acids* **2012:** Article ID 360358.