Sequence and structure requirements of Y RNA-derived small RNA biogenesis

Carly Turnbull

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

University of East Anglia
Norwich

School of Biological Sciences

May 2014

ABSTRACT

Numerous small non-coding RNAs have been identified in mammalian cells, including microRNAs and piwi-interacting RNAs. The patterns of gene expression within cells can be altered in response to cellular stress. To examine the effects of cellular stress on all small RNA types, a number of human cell lines were treated with Poly(I:C), a mimic of viral infection, and the levels of small RNAs were examined by next generation sequencing. Surprisingly, we did not find many differentially expressed microRNAs, but we discovered a new class of small RNAs that were 30-35 nucleotides and showed up-regulation following Poly(I:C) treatment. These slightly longer small RNAs were derived from many different types of non-coding RNA and only very few small RNAs were derived directly from messenger RNAs. Small RNAs derived from various types of RNA were validated by northern blot. Further sequencing libraries were prepared for Poly(I:C)-treated and untreated MCF7 cells, as well as Poly(I:C)-treated and untreated SW1353 cells. The human Y5 RNA gene was chosen as an example of a Poly(I:C)-induced small RNA-producing gene. This gene was cloned into an expression construct and systematically mutated to alter the sequence or secondary structure of the resulting Y RNA. These mutant plasmids were expressed in mouse cells and the effect on small RNA production determined. These individual mutants together helped to determine a region vital for cleavage, and that it is the structure rather than the sequence within this region that is important. A high-throughput method was also implemented, involving the generation of large pools of plasmids containing all possible sequences within a particular region of the RNA gene. These pools were expressed in mouse cells and the mutants that were expressed and processed into small RNAs were sequenced. These experiments showed that the formation of the large internal loop determines the internal cleavage site.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Tamas Dalmay, for all his help and advice during this project, and for controlling the many ideas I had whilst undertaking it. I am grateful to Guy Wheeler for his help and advice and for forcing me to take regular breaks for coffee. I am very grateful to Ping Xu and Sara Lopez-Gomollon for their careers advice and their honest opinions on my work, as well as providing me with the enthusiasm and encouragement to continue in science.

Many thanks to all those in the Munsteberg lab for providing me with samples and helping me to extract them, especially those that were sent from other labs.

I am very grateful to my fiance Lee, for moving my entire house for me while I was doing my last few weeks of lab work. You have given me an amazing amount of support and I am not sure I would have finished without it!

ABBREVIATIONS

30mer 30-35 nucleotide small RNA
AGO Argonaute
ANG Angiogenin APS Ammonium persuphate
ARG Analytical reagent grade
ATP Adenosine triphosphate
cDNA Complementary DNA
CNBP Cellular nucleic acid binding protein
CRM1 Chromosome region maintenance 1 (also exportin1 or Xpo1)
Δ3' Mutant with 3 nucleotide in 3 tail deleted
Δ1S Mutant with one side of the stem deleted
Δ2S Mutant with both sides of the stem deleted
ΔACC Mutant with ACC nucleotides deleted
ΔRo Mutant with Ro60 binding site deleted
DGCR8 DiGeorge Syndrome Critical Region 8
DMEM Dulbecco's modified eagle medium
dsRNA Double-stranded RNA
dsDNA Double-stranded DNA
DTT Dithiothreitol
dNTPs Deoxyribonucleotides
drY *D. radiodurans* homolog of Y RNA
EBER Epstein-Barr virus-encoded small RNA
EDC 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
EDTA Ethylenediaminetetraacetic acid
eIF4F Eukaryotic initiation factor 4F complex
eIF4G/A Eukaryotic initation factor 4G/A complex
Endo-siRNA Endogenous small interfering RNA
FAM Fluorescein amidite
FBS Fetal bovine serum
gDNA Genomic DNA
HDAC Histone deacetylase
hnRNP Heterogeneous nuclear ribonucleoprotein particle
hY Human Y RNA
IFIT5 Interferon-induced protein with tetratricopeptide repeats 5
IRES Internal ribosome entry site
IPTG Isopropyl -D-1-thiogalactopyranoside
LB Luria-Bertani
LNA Locked nucleic acid

lncRNA Long intergenic non-coding RNA
MBq Megabecquerel
MCPIP1 Monocyte chemotactic protein-induced protein
mESCs Mouse embryonic stem cells
MIDAS Metal ion-dependent adhesion site
miRNA MicroRNA
MOV10 Moloney leukaemia virus 10, homolog (mouse)
mRNA Messenger RNA
mRo60 Mouse Ro60
mY Mouse Y RNA
ncRNA Non-coding RNA
NGS Next generation sequencing
nt Nucleotides
OptiMEM OptiMEM I reduced serum media
PAGE Polyacrylamide gel electrophoresis
PASR Promoter-associated small RNA
PBS Phosphate buffered saline
PCR Polymerase chain reaction
PI Propidium iodide
piRNA Piwi-interacting RNA
Piwi P-element induced wimpy testes in *Drosophila*
PKR Protein kinase RNA-activated
PNK T4 polynucleotide kinase
PNPase Polynucleotide phosphorylase
Poly(I:C) Polyinosinic:polycytidylic acid
Puf60 Poly(U)-binding splicing factor 60 kDa
RISC RNA-induced silencing complex
RNase Ribonuclease
RNASET2 Human orthologue of Rny1p
RNH1 Ribonuclease inhibitor of angiogenin
RNP Ribonucleoprotein particle
Rny1p Ribonuclease in yeast 1
RoBP1 Ro60-binding protein 1
RPA Replication protein A
Rpm Revolutions per minute
RRM RNA recognition motif
rRNA Ribosomal RNA
Rsr Ro-sixty related (*D. radiodurans* orthologue of Ro60)
RT-PCR Reverse transcriptase PCR

QC Quality Control
SAP Shrimp alkaline phosphatase
sbRNA Stem bulge RNA
SDS Sodium dodecyl sulfate
siRNA Small interfering RNA
sitRNA Stress-induced tRNA-derived RNA
SLE Systemic lupus erythematosus
sno-miRNA Small nucleolar RNA with microRNA-like activity
snoRNA Small nucleolar RNA
snRNA Small nuclear RNA
sRNA Small RNA
SS Sjögrens Syndrome
SSC Sodium chloride - sodium citrate buffer
STS Staurosporine
Sub1S Mutant with one side of the stem substituted
Sub2S Mutant with both sides of the stem substituted
svRNA Small vault RNA
TBE Tris-borate-EDTA
TEMED Tetramethylethylenediamine
tiRNA Transcription initiation RNA
TOP Terminal oligopyrimidine tract
tRNA Transfer RNA
Tsix Antisense RNA transcribed from Xist locus
TTSaRNA Transcription start site-associated RNA
UV Ultra violet
vWFA Von Willebrand factor A-like motif
X-gal 5-Bromo-4-chloro-3-indolyl -D-galactopyranoside
XIAP X-linked inhibitor of apoptosis
XiRNA X-inactivation RNA
Xist X-chromosome inactivation-specific transcript
xRo60 *Xenopus sp.* Ro60
xY *Xenopus sp.* Y RNA
YB-1 Y box-binding protein 1
ydRNA Y RNA-derived RNA
ZBP1 Zipcode-binding protein 1

Contents

List of Figures

List of Tables

1 INTRODUCTION

When the human genome sequence was completed many scientists struggled to explain how so little of the genome could contain protein-coding genes. From this point on an examination of the 'junk' DNA in the genome began and an expanding area of science resulting from this is the study of the human 'transcriptome'. This transcriptome is made up not only of messenger RNAs (mRNAs), but also of many different types of non-coding RNA (ncRNA). A number of classes of ncRNA, such as transfer RNA (tRNA) and ribosomal RNA (rRNA), were discovered over fifty years ago but other classes are still being discovered. Research into novel types of ncRNA is progressing at a phenomenal pace, as a result of continuing advancements in sequencing technologies. Below are detailed some of the more recently discovered classes of ncRNA identified in mammals.

1.1 Long non-coding RNAs

This class of RNAs consists of all ncRNAs that are longer than 200 nucleotides. The majority of long ncRNAs (lncRNAs) are thought to be produced by RNA polymerase II (Guttman et al., 2009). Wang and Chang describe examples where lncRNAs act as "signals, decoys, guides, and scaffolds" or often combinations of these functions (Wang and Chang, 2011). These RNAs are sometimes further classified into long intergenic ncRNAs that are present in the genome between protein-coding genes, and long intronic ncRNAs (Sharma and Misteli, 2013)

1.2 Small non-coding RNAs

ncRNAs shorter than 200 nucleotides are considered small ncRNAs. They are grouped based on their biogenesis and function.

1.2.1 MicroRNAs

MicroRNAs (miRNAs) are short RNA species that are around 22 nucleotides in length in their mature form (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). They are expressed endogenously and interact with target mRNA transcripts to block their expression (Lee et al., 1993). miRNA binding can either prevent translation (Wightman et al., 1993), or de-stabilize the mRNA transcript resulting in it being targeted for degradation (Lim et al., 2005). miRNAs are generally either located within the introns of previously annotated genes (Aravin et al., 2003; Lagos-Quintana et al., 2003; Lai et al., 2003; Lim et al., 2003) or are clustered together within regions of the genome that do not contain protein-coding genes (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). Some

miRNAs are expressed in many parts of the organism (Pasquinelli et al., 2000; Lagos-Quintana et al., 2003), while others are expressed in only one or two tissues (Lee and Ambros, 2001) As well as having a specific spatial expression pattern, many miRNAs are also expressed temporally (Reinhart et al., 2000).

1.2.2 Small interfering RNAs

Unlike miRNAs, small interfering RNAs (siRNAs) are produced as a result of processing of a long double-stranded RNA (dsRNA) molecule (Hammond et al., 2000; Zamore et al., 2000). This dsRNA was originally thought to be produced exogenously, entering the cell through viral infection or being transcribed from transposable elements that have inserted into the genome (Hamilton and Baulcombe, 1999; Ketting et al., 1999; Tabara et al., 1999). Several years after siRNAs were first discovered, a subclass of endogenous siRNAs (endo-siRNAs) was identified (Hamilton et al., 2002; Ambros et al., 2003; Zilberman et al., 2003). endo-siRNAs are produced when long dsRNA molecules are cut by the Dicer protein (Bernstein et al., 2001; Grishok et al., 2001; Ketting et al., 2001). As opposed to its action in miRNA biogenesis, Dicer produces endo-siRNAs by moving along the dsRNA and cutting at various positions (Myers et al., 2003; Provost et al., 2002). endo-siRNAs have been found in only a few mammalian tissues to date.

1.2.3 Piwi-interacting RNAs

The main function of piwi-interacting RNAs (piRNAs) is their involvement in the prevention of transposon mobility, preserving the integrity of the genome in germline cells (Vagin et al., 2006; Brennecke et al., 2007). These RNAs interact with the Piwi (originally P-element induced wimpy testis in *Drosophila*) clade of Argonaute proteins (Grivna et al., 2006; Aravin, 2006; Girard et al., 2006; Lau et al., 2006). There appears to be a "ping-pong" method of biogenesis, where a primary piRNA is produced and aids the biogenesis of a secondary piRNA (Brennecke et al., 2007; Gunawardane et al., 2007). This secondary piRNA is often antisense to target transposon mRNAs and it hybridizes to them, blocking their expression. This biogenesis is not dependent upon an RNase III enzyme, as miRNAs and siRNAs are (Houwing et al., 2007; Vagin et al., 2006; Brennecke et al., 2007). piRNAs were initially thought to only be expressed in the germline, but we now know that they are expressed in somatic cells (Li et al., 2009; Malone et al., 2009) and in a number of cancers (Cheng et al., 2011; Cui et al., 2011; Lu et al., 2010).

1.2.4 Promoter-associated RNAs

The class of promoter-associated RNAs encompasses a number of different groups of RNAs that have been discovered; including promoter-associated short RNAs (PASRs), transcription start site-associated RNAs (TSSaRNAs), and transcription initiation RNAs (tiRNAs). All of these RNAs are produced from highly expressed genes (Seila et al., 2008; Taft et al., 2009c,a). PASRs are usually capped (Affymetrix ENCODE Transcriptome Project and Cold Spring Harbor Laboratory ENCODE Transcriptome Project, 2009), they are between 20 and 200 nucleotides in length, and their 5' ends map to the transcription start sites of the genes they regulate (Kapranov et al., 2007). TSSaRNAs are 20 to 90 nucleotides in length and are transcribed from the -250 to +50 region of the gene (Seila et al., 2008). The shorter 18nt tiRNAs are produced from position -60 to +120 in relation to the transcription start site of the gene. RNA polymerase II stalls at a nucleosome at the +1 position, which results in a backtracking and release of these short transcripts (Taft et al., 2009c).

1.2.5 X-inactivation RNAs

Prior to and post X-chromosome inactivation, the X-inactivation specific transcript (Xist) and the Xist antisense transcript (Tsix) RNAs are expressed from both chromosomes, but during X-chromosome inactivation only one RNA is expressed from each chromosome (Lee et al., 1999). Ogawa *et al.* have shown that these two RNAs can form a dsRNA duplex, some of these dsRNAs are less stable than others and are degraded into shorter RNAs of varying length called X-inactivation RNAs (xiRNAs) (Ogawa et al., 2008). xiRNAs are only produced during X-inactivation and not before or after. These xiRNAs are proposed to repress Xist on the active X, but their exact mode of action has yet to be confirmed (Ogawa et al., 2008).

1.2.6 Small vault RNAs

Vault RNAs were discovered in the 1980s (Kedersha and Rome, 1986) and later were linked to drug resistance (Mossink et al., 2003). Vault RNAs associate with a number of different proteins to form vault particles (Kedersha and Rome, 1986). In 2009 human vault RNAs were found to be processed into at least six different small vault RNAs (svRNAs), between 23 and 40 nucleotides in size. At least one of these acts like a miRNA, binding to Argonaute proteins and targeting the mRNA of the CYP3A4 gene, which produces a protein key in drug metabolism (Persson et al., 2009). The levels of svRNAs vary between tissue types, but this does not seem to correlate with the levels of vault proteins in these tissues (Persson et al., 2009).

1.2.7 Transfer RNA-derived RNAs

Transfer RNAs can be cleaved by a number of different enzymes to produce miRNA-sized fragments. The RNase Z (ELAC2) enzyme is involved in the removal of the tRNA 3' trailer sequence during normal tRNA maturation (Takaku et al., 2003). Lee *et al.* detected a number of small tRNA fragments in human cells, some of which were produced by RNase Z and were derived from these 3' trailers (Lee et al., 2009). One of these RNase Z-dependent fragments was found to be essential for proliferation (Lee et al., 2009). Other groups detected miRNA-sized tRNA fragments that were Dicer-dependent (Cole et al., 2009; Babiarz et al., 2008).

1.2.8 SnoRNA-derived RNAs

Small nucleolar RNAs (snoRNAs) can be divided into two classes; the box C/D snoRNAs and the H/ACA snoRNAs, depending upon the sequence motifs they contain. Both of these classes of snoRNA have been shown to produce miRNA-like fragments that have gene-silencing activity in human cells (Brameier et al., 2011), although the H/ACA box snoRNAs more commonly produce microRNAs (Ender et al., 2008; Taft et al., 2009b). A number of sequences already entered into miRBase, the miRNA registration database (Griffiths-Jones, 2004), have been re-classified as sno-miRNAs (Brameier et al., 2011).

1.3 Stress-induced small ncRNAs

Cells within the body are under constant stress from the external environment and need to respond rapidly to stimuli. They may detect many different threats, such as viral infection, changes in temperature or changes in oxygen levels. Cells must decide how to react to these new conditions, and whether to warn surrounding cells of the threat. Cells alter their metabolism in order to adapt to these changes in the environment. They may decide to try and repair any damage caused by the stress, or they may decide that the damage is too great and undergo apoptosis or necrosis. It is now widely accepted that ncRNAs play an important role in many cellular processes and this project aims to investigate their role in cellular response to stress.

Since the commencement of this project a number of groups have published evidence of stress-induced small RNAs derived from longer RNAs, many of which were seen in our dataset. Here is a summary of this very recent data. Unlike the miRNA-sized fragments derived from snoRNAs and tRNAs already discussed, these groups focused on longer small RNAs that were produced in response to stress.

1.3.1 Transfer RNA 'halves'

Transfer RNAs were first found to be bisected in *E. coli* cells in the presence of the plasmid-encoded colicins E5 and D (Ogawa et al., 1999; Tomita et al., 2000). These longer tRNA-derived fragments were denoted tRNA 'halves' after cleavage was found to occur roughly in the middle of the molecule (Ogawa et al., 1999). The E5 colicin produces 5' and 3' halves derived from $tRNA^{Tyr}$, $tRNA^{His}$, $tRNA^{Asp}$ and $tRNA^{Asn}$ and these tRNAs are cleaved in the anticodon loop, between the nucleotides at positions 34 and 35 (Ogawa et al., 1999). The D colicin targets $tRNA^{Arg}$ specifically and cuts these molecules in a different place, cutting between positions 38 and 39 (Tomita et al., 2000). Both of these colicins are toxic to bacterial cells, as the tRNA halves produced block translation (Ogawa et al., 1999; Tomita et al., 2000).

The first stress-induced tRNA halves were found in *Tetrahymena thermophila* during nutrient starvation (Lee and Collins, 2005). A deficiency in any one of the ten essential amino acids led to cleavage of tRNAs into fragments of between 30-35 nucleotides in size. A similar reaction was seen when these cells were heat-shocked, or entered stationary phase of the cell cycle (Lee and Collins, 2005). The fragments detected were derived from mature tRNAs, the majority of which were nuclear-encoded rather than mitochondrial (Lee and Collins, 2005). This group also noted that these fragments were only seen in early responses to starvation and were absent in later stages.

After tRNA half discovery in *T. thermophila*, the race was on to identify these small RNAs in other organisms and three years later they were discovered in single-celled organisms, such as the fungus *Aspergillus fumigatus* (Jöchl et al., 2008) and the yeast *Sacchromyces cerrevisiae* (Thompson et al., 2008), as well as the simple eukaryote *Giardia lamblia* (Li et al., 2008). *Aspergillus* cells produce tRNA halves as they undergo a sporulation stage of their life cycle, when protein synthesis is blocked (Jöchl et al., 2008). Yeast tRNA halves were detected after a number of stresses, most significantly after oxidative stress. Interestingly they were not detected after amino acid starvation (Thompson et al., 2008). Yeast cells appear to produce tRNA halves derived from the 3' and the 5' end of the tRNA at equal levels and these fragments were sized at 40 nucleotides (Thompson et al., 2008). These halves were detected for the first time in unstressed cells and were found at much lower levels than in stressed cells (Thompson et al., 2008). The situation in *G. lamblia* cells is more complex, as two classes of tRNA-derived small RNAs have been detected: tRNA halves and 40 nucleotide fragments named stress-induced tRNA-derived RNAs (sitRNAs) (Li et al., 2008). SitRNAs were present as these cells changed from the vegetative state into cysts, whereas tRNA halves were not detectable under these conditions (Li et al., 2008). Both of these types of tRNA

fragments were detectable in *G. lamblia* cells undergoing serum starvation (Li et al., 2008). SitRNAs are likely a unique adaptation of this early eukaryote in response to particular stress situations.

In 2008 tRNA halves were also detected in human cells (Kawaji et al., 2008; Thompson et al., 2008) and plant cells (Thompson et al., 2008). In human cells tRNA halves were found to be 30-40 nucleotides in length and were induced by phosphate buffered saline (PBS) treatment and serum starvation (Fu et al., 2009). Kawaji *et al.* noticed that the abundance of these fragments did not correlate with the level of the full length tRNA from which they were derived, or the number of copies of that tRNA gene present in the genome, suggesting that either certain tRNAs are targeted for cleavage over others, or certain fragments are more stable after cleavage than others (Kawaji et al., 2008). The 5' and 3' halves produced from the same tRNA molecule were often present at different levels (Kawaji et al., 2008). Transfer RNA halves are expressed in many different human tissue types after stress (Fu et al., 2009) and each tRNA half shows a unique tissue-specific expression pattern (Kawaji et al., 2008). Figure 1.1 shows the levels of various tRNA halves present in HepG2 cells after growth in PBS or in DMEM without serum. Thirty minutes of growth in PBS produced the highest level of 5' tRNA halves for all tRNAs, except $tRNA^{Tyr}$, which produced no detectable halves (Fu et al., 2009). Fu *et al.* suggested that this is because the secondary structure of $tRNA^{Tyr}$ is different to other tRNAs, as it is longer (Fu et al., 2009). Based on their own results and those of Kawaji group, Thompson *et al.* suggested that the purpose of tRNA cleavage is unlikely to be to reduce full length tRNAs, as the levels of full length tRNAs were hardly affected and levels of each cleavage product were much lower than the corresponding full length tRNA. They hypothesized that the fragments are likely to play some sort of direct role in blocking translation, although they found no evidence for this at the time (Thompson et al., 2008).

In 2009 further work by Thomspson *et al.* uncovered the enzyme responsible for cleavage in yeast. Rny1p, an endonuclease normally stored in the vacuoles of yeast cells, is released into the cytosol in cells undergoing stress and is then able to target the cytoplasmic tRNAs (Thompson and Parker, 2009a). Two groups simultaneously proved that the enzyme responsible for stress-induced tRNA half production in human cells is Angiogenin (Yamasaki et al., 2009; Fu et al., 2009). This protein had previously been identified as a tRNase, but cleavage conditions had not been well characterised (Saxena et al., 1992). Angiogenin is present in the nucleus, nucleolus and the cytoplasm of human cells (Moroianu and Riordan, 1994; Tsuji et al., 2005). Cytoplasmic Angiogenin is held in an inactive form by its binding partner Ribonuclease/Angiogenin inhibitor 1 (RNH1) (Shapiro and Vallee, 1987). Under conditions such as oxidative stress Angiogenin could become active by two mechanisms. Firstly,

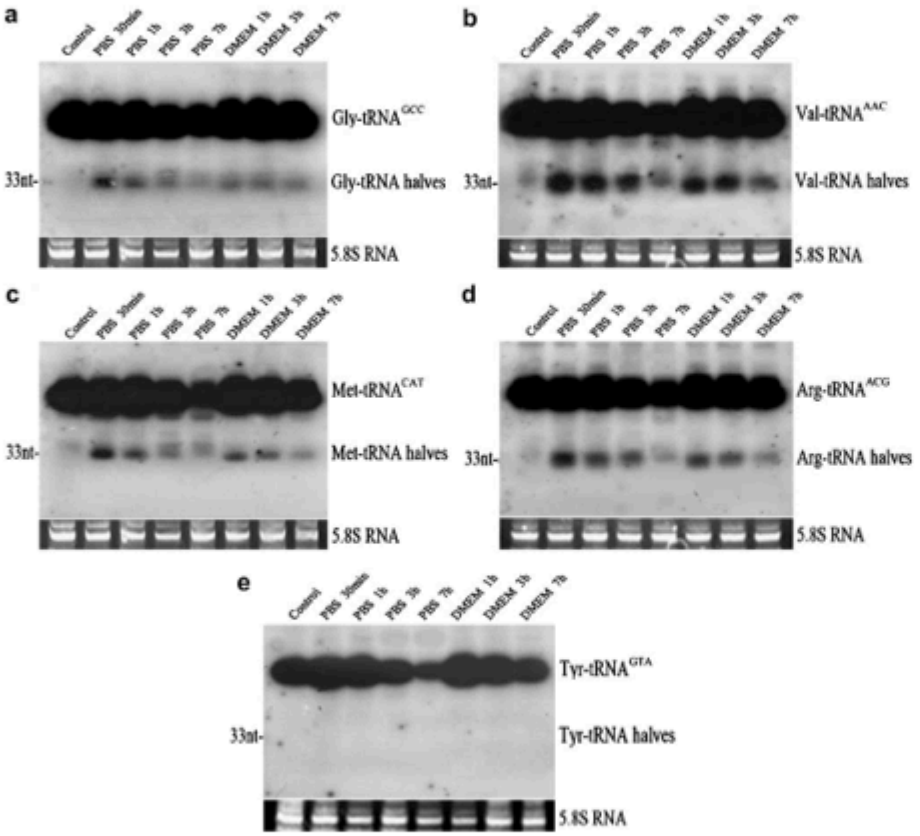


Figure 1.1: The cleavage of tRNAs.
In experiments by Fu *et al.* (2009), HepG2 cells were treated with PBS or grown in DMEM without serum. Northern membranes were probed for the 5' tRNA half of various tRNAs. Tyr-tRNA halves were not detected in these experiments. 5.8S rRNA stained with ethidium bromide was used as a control for RNA loading.

the associated RNH1 could be released, allowing cytoplasmic Angiogenin to become active and cleave tRNAs. Secondly, Angiogenin could be released from the nucleus or nucleolus and act on tRNAs in the cytoplasm. An inverse relationship between the levels of RNH1 and Angiogenin suggests that it is mainly cytoplasmic Angiogenin that is responsible for tRNA cleavage (Yamasaki et al., 2009). Yamasaki *et al.* showed that the 5' halves, when purified from a polyacrylamide gel and transfected back into cells, blocked translation but the 3' halves did not. They indicated that the 5' halves most likely have a monophosphate at the 5' end and a 2', 3' cyclic phosphate at the 3' end (Yamasaki et al., 2009). The 3' fragments are unlikely to have a 5' monophosphate or a 2', 3'cyclic phosphate at the 3' end, and are therefore less likely to be functionally active (Yamasaki et al., 2009). These results should be viewed objectively however, as it is impossible to rule out that other functional small RNAs were not purified alongside the 3' or 5' tRNA fragments. In their review in 2009, Thompson *et al.* suggest that the anticodon loop is targeted for cleavage as it is the most accessible point on the tRNA molecule (Thompson and Parker, 2009b). It was Ivanov *et al.* who more fully characterized the role of these tRNA fragments in translation. They found that only a small proportion of tRNA 5' halves had this role and that those that did had 4-5 guanine bases at their 5' ends (Ivanov et al., 2011a). These are essential in inhibiting translation as they interact with the initiation factors eIF4G/A and the initiation complex eIF4F, causing dissociation from mRNAs. Transfer RNA halves likely inhibit translation through an interaction with Y box-binding protein 1 (YB-1), which has also been shown to have an involvement in this process (Ivanov et al., 2011a). $tRNA^{Ala}$ and $tRNA^{Cys}$ are most active in this process and it is mainly uncapped translation initiation that is blocked by these tRNA halves (Ivanov et al., 2011a).

In summary, the process of tRNA half generation appears to be a conserved response to cellular stress, despite the fact that cleavage positions and stress conditions differ between organisms, these differences are probably the result of a range of nucleases acting on tRNAs (Li et al., 2008). Why tRNA halves are generated from so many different tRNA types in mammalian cells, when only a small proportion block translation, requires further investigation. It seems likely that they are involved in a host of other stress-related processes that have yet to be uncovered.

1.3.2 Ribosomal RNA-derived RNAs

As with tRNAs, ribosomal RNA (rRNA) was found to be cleaved in *E. coli* after the expression of a toxin. Colicin E3 is specific for bacterial 16S rRNA and cleaves towards the 3' end to give a fragment of 50 nucleotides. This cleavage event blocks translation at the elongation step in *E. coli* cells, resulting in cell death. In

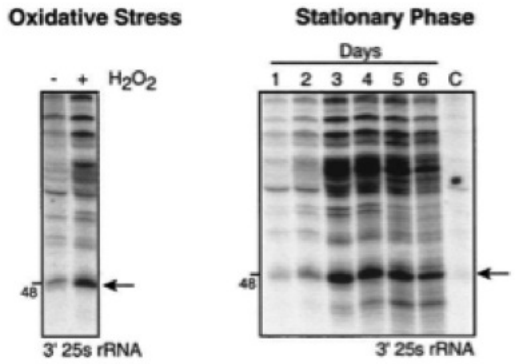


Figure 1.2: The cleavage pattern of 25S ribosomal RNA.
Experiments by Thompson *et al.* (2008), were conducted in yeast cells during stationary phase or under oxidative stress conditions after treatment with hydrogen peroxide. Membranes were probed for various tRNA halves. An arrow marks an sRNA of 50 nt in size, although many other RNA fragments can be seen on both gels.

Tetrahymena RNase T2 enzymes have been shown to cleave rRNAs during starvations conditions (Andersen and Collins, 2012), but stress-induced rRNA cleavage in eukaryotic cells is less well defined. Ribosomal RNA cleavage was identified in yeast cells in response to oxidative stress, but a number of fragments were detected rather than a fragment of one specific size, as for tRNAs (Thompson et al., 2008). The investigation of ribosomal RNA fragment production is impaired by the size of these RNA molecules and the complexity of their secondary structure.

1.4 Origins of the project

In preliminary experiments conducted by Dr. Nicolas, a senior research associate in the Dalmay group, the MCF7 breast cancer cell line was subjected to a number of environmental stresses, before being harvested for RNA extraction. Stress conditions included heat shock, cold shock, starvation and Polyinosinic:polycytidylic acid (Poly[I:C]) treatment. Poly(I:C) is a synthetic double-stranded RNA used to mimic viral infection (Larson et al., 1969). Libraries were prepared for each sample where RNAs of 19-33 nucleotides were selected and sequenced using Illumina next generation sequencing (NGS). The levels of small RNA expression under each condition were compared to those of a control untreated sample.

Initially, the aim was to identify miRNAs that were differentially expressed between the stressed and control samples, but a surprisingly small number of miRNAs displayed altered expression. What was also unexpected was that a large number of sequences derived from other ncRNAs were increased in the Poly(I:C)-treated sample when compared with the control. These differences were not seen in cells

treated with any of the other stress conditions.

In the Poly(I:C) sample 21,837,679 redundant reads were sequenced, 79% of which could be mapped to the human genome. Filtering of the data identified 490,826 non-redundant, or unique sequences, 33% of which could be mapped to the human genome. This suggests that the unique sequences mapping to the genome make up a large percentage of the total reads (since almost 80% of all reads mapped to the genome) but many unique sequences (67%) did not map to the genome, therefore, these must have low read numbers (since only 21% of all reads did not map to the genome).

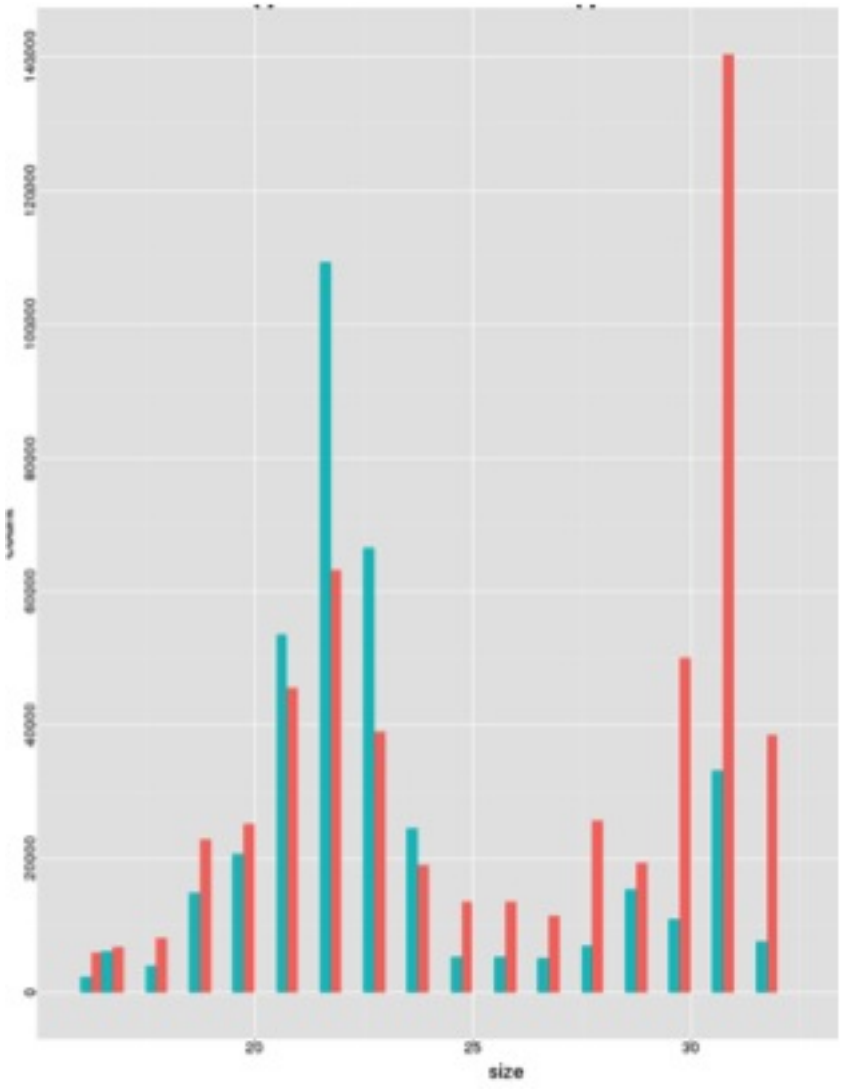


Figure 1.3: The distribution of sequencing reads.
NGS libraries were prepared from control and Poly(I:C)-treated samples and this is a graphical representation of the distribution of the sequencing reads according to size and abundance. Control sample reads are shown in turquoise and Poly(I:C)-treated sample reads are shown in red.

In a graphical representation of the Illumina NGS data (see Figure 1.3), where

the number of reads of each sequence is plotted against the read length, the miRNA fraction can be seen as a peak at around 22 nucleotides. The majority of the ncRNAs altered in the stressed samples are present in a peak of 31 nucleotides. The 22 nucleotide peak is present in the control and Poly(I:C)-treated samples. The 31 nucleotide peak is present in the control sample, but at substantially lower levels than in the Poly(I:C)-treated sample.

One of the most highly up-regulated sequences in the Poly(I:C)-treated sample compared to the control is derived from a human Y RNA. We initially focused on this sequence, as it was originally annotated as a miRNA, but has since been removed from miRBase (Griffiths-Jones, 2004). Work by the Dalmay group has provided much evidence to show that this sequence, as with a number of other Y RNA-derived sequences, is not generated by the miRNA biogenesis machinery and is also not incorporated into the RNA-induced silencing complex (Nicolas et al., 2012).

1.4.1 What are Y RNAs?

Systemic Lupus Erythematosus and Sjögren's Syndrome are diseases characterized by the presence of antibodies against a patient's own antigens (Fennell et al., 1962; Vazquez, 1963). Widespread targets include DNA, RNA and protein. Early experiments characterizing the nuclear targets of these autoantibodies led to the discovery of Y RNAs. The term Y RNA was coined to distinguish the cytoplasmic location of these RNAs from the nuclear U RNAs (Lerner et al., 1981b). Y RNAs are around 100nt in size and are usually found in ribonucleoprotein complexes (RNPs). They are still generally accepted to be mainly cytoplasmic in location. Y RNAs have been conserved throughout evolution and have so far been identified in many vertebrates, as well as bacteria, nematodes and mosquitos (Van Horn et al., 1995; Perreault et al., 2007; Mosig et al., 2007; Chen et al., 2000). Y RNAs differ in sequence, but have a conserved secondary structure. In certain regions of these molecules, however, the sequence is highly conserved, as these regions contain protein binding sites.

Y RNAs were originally identified through their interaction with the autoantigens Ro60 and La (Lerner et al., 1981b). Each Y RNA can bind to either or both of these proteins to form an RNP complex (Hendrick et al., 1981). They are believed to commonly interact with Ro60 and a subset of these Ro RNPs also contain La (Hendrick et al., 1981; Wolin and Steitz, 1984). Each type of Y RNA in humans can bind to a subset of Y RNA-binding proteins depending on its conserved secondary structure, as will be discussed later.

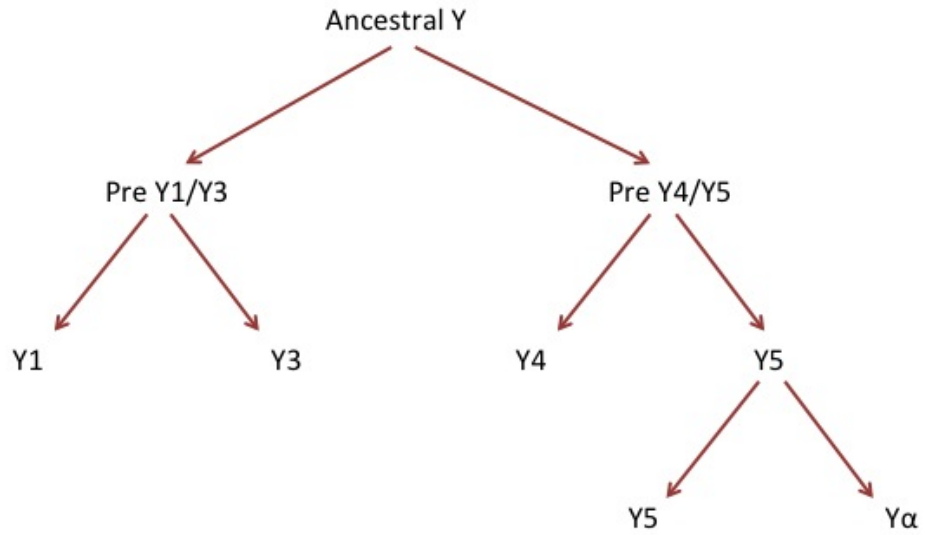


Figure 1.4: The proposed evolution of the Y RNA genes.
Adapted from Mosig *et al.* 2007.

1.4.1.1 Y RNA evolution

All Y RNAs are believed to be paralogs, as they are derived from the same ancestral Y RNA gene, which has been duplicated during evolution (Perreault et al., 2007; Mosig et al., 2007). Figure 1.4 shows a representation of how these RNAs are thought to have developed (Mosig et al., 2007). During the evolution of placental, or eutherian, mammals the Y RNA gene cluster at some point contained four different Y RNA genes, as they are all present to some degree in each taxonomic group within this lineage (Mosig et al., 2007). Figure 1.5 displays the structure of the Y RNA gene cluster in this eutherian ancestor. In genomes that contain Y RNAs, each Y RNA is thought to be expressed from a single gene (Mosig et al., 2007), except for in *Xenopus*, where two versions of *Xenopus* Y3 (xY3) RNA with a one nucleotide difference have been identified, suggesting that there may be two active copies of this gene in the *Xenopus* genome (O'Brien et al., 1993). Y RNA pseudogenes are found in many eukaryotes (Perreault et al., 2007; Mosig et al., 2007), but in primates a large number of psuedogenes have developed (Perreault et al., 2005, 2007). None of these pseudogenes, either in primates or in other eukaryotes, are thought to be expressed (Perreault et al., 2005, 2007; Mosig et al., 2007).

Figure 1.5: The proposed structure of the ancestral vertebrate Y RNA gene cluster.
Adapted from Mosig *et al.* 2007.

In humans all four Y RNA genes are clustered together on chromosome 7, with Y1 inverted in orientation compared to the rest (Wolin and Steitz, 1983; Maraia et al., 1994b). In what is a striking difference between humans and rodents, both mice and rats only have two active Y RNA genes, which are homologous to hY1 and hY3 (Hendrick et al., 1981), while in humans all four genes are active. Frogs have genes for xY3, xY4 and xY5 having lost the homolog of hY1, but gained a paralog Y RNA called Yα (O'Brien et al., 1993). Dogs and cows also have four different functional Y RNA genes (Perreault et al., 2007). Zebrafish and nematode worms were found to have two versions of this family of ncRNAs, so have Y RNAs gained an important tetrapod-specific function, with more copies developing as tetrapods have evolved? In *C. elegans* a large class of RNAs known as stem-bulge, or sbRNAs, have recently been re-classified as Y RNAs (Boria et al., 2010), providing evidence of a lower complexity organism with many more potentially functioning Y RNAs than tetrapods. This family of RNAs may have a vital function which has been conserved through evolution, or certain family members could have evolved to take on specialized functions within particular taxonomic groups.

1.4.2 Y RNA secondary structure

All vertebrate Y RNAs have a similar secondary structure. In Figure 1.6 the common structure has been divided into different regions. Region 1 forms a terminal stem made up of the 3' and 5' ends of the molecule (Wolin and Steitz, 1983). This region is important for Ro60 protein binding, as will be discussed in greater detail later. There is some variation in the length of the 3' tail between different Y RNAs; in humans Y5 has the longest single-stranded overhang (Kato et al., 1982; Wolin and Steitz, 1983). An extra uridylate, marked in Figure 1.6 in brackets, is present at the 3' end of some Y RNA molecules (Wolin and Steitz, 1983). In region 2 there is a conserved cytosine bulge and a larger bulge on the opposite stem. A further stem is formed in region 3. The helix in region 3 has been linked to Y RNA export from the nucleus (Rutjes et al., 2001), as well as being required for the RNA to function in DNA replication (Christov et al., 2006), as discussed later. Regions 1 and 3 are highly conserved between Y RNAs and across species (Pruijn et al., 1997; Teunissen et al., 2000). In region 2 the two bulges are conserved (Green et al., 1998). In region 4 there is an internal loop and in region 5 there may be one or two stem-loops. Regions 4 and 5 are less well conserved than regions 1-3 and have been proposed to be rather dynamic, as "breathing" occurs to allow conformational changes and potentially enable interactions with proteins or other nucleic acids (van Gelder et al., 1994; Teunissen et al., 2000). Comparison of the sequences of all identified Y RNAs suggests that there are five central base pairs that are conserved across all Y RNAs

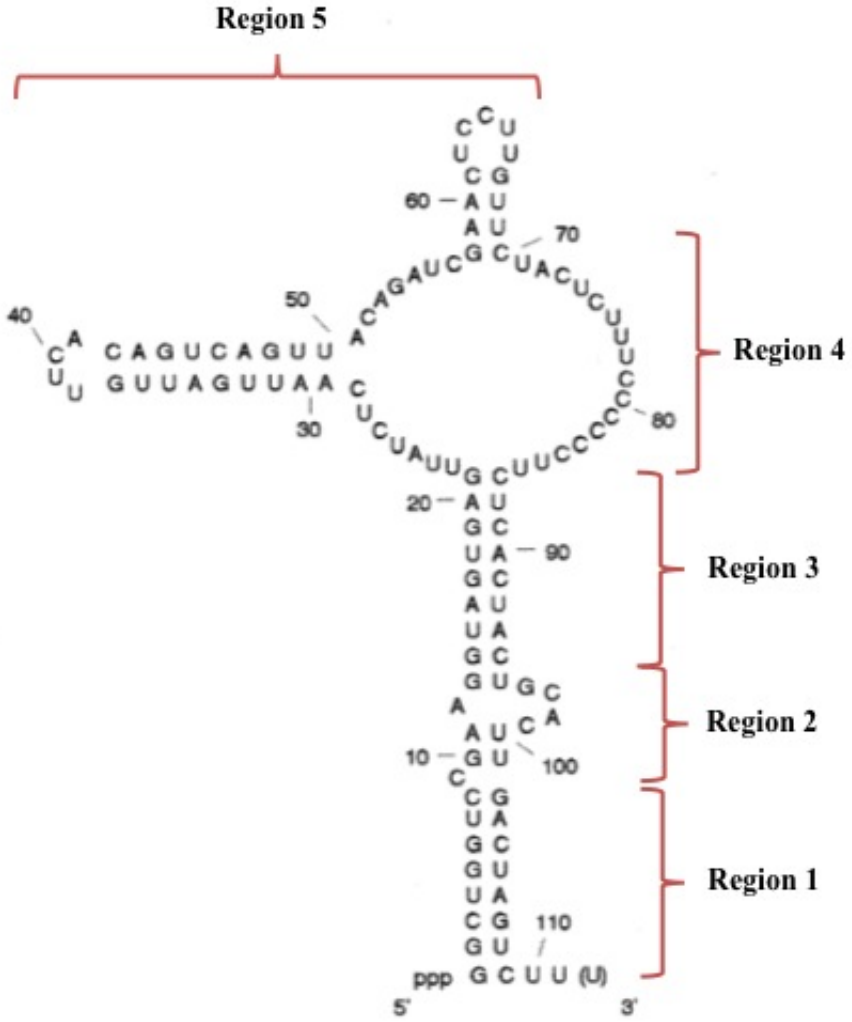


Figure 1.6: The secondary structure of a human Y RNA.
This structure can be divided into particular regions according to sequence conservation and function. Adapted from van Gelder *et al.* 1994.

and these are likely to be essential for a common function of these RNAs (Green et al., 1998).

The secondary structure of all human Y RNAs has been experimentally determined by a number of groups using enzymatic digestion and chemical modification. Comparison of experimentally validated structures with those predicted by folding algorithms highlighted a few differences. Firstly, in region 2 there is some variation in the cysosine that bulges out, but it is more often the cytosine at position 9 in the Ro60-free Y RNA, rather than the predicted cytosine at position 8 that is present in the bulge (van Gelder et al., 1994). Secondly, the larger bulge in region 2 was shown to be closer to the terminal end than predictions suggested. In a proportion of Y RNA molecules these two separate bulges can join to form a small internal loop (van Gelder et al., 1994). Finally, in region 4 early prediction algorithms produced two loops, while experimental data indicated that this region is more likely to form a single large loop. In the larger Y RNAs this large internal loop may form tertiary structures, although this is unlikely for smaller RNAs such as Y5 (van Gelder et al., 1994). This loop is likely to have an altered conformation when interacting with proteins (van Gelder et al., 1994).

1.4.3 Y RNA binding proteins

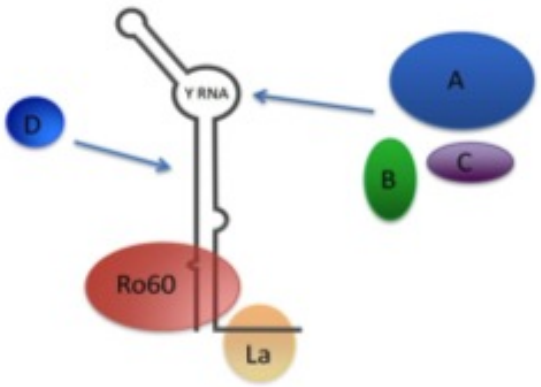


Figure 1.7: Protein binding sites.
The secondary structure of a human Y RNA with the binding sites for Ro60 and La displayed. Proteins A-D represent where proteins other than Ro or La are known to bind.

Many of the RNAs within a cell will associate with RNA-binding proteins to form RNPs. As detailed in Figure 1.7, the Ro60 and La proteins interact with each type of human Y RNA (Hendrick et al., 1981; Wolin and Steitz, 1983). There have been some suggestions that the 52 kDa Ro protein may be incorporated into Ro RNPs, but work by various groups has shown that this is not the case (Kelekar et al., 1994; Boire et al., 1995; Itoh et al., 1992).

1.4.3.1 Ro60 and misfolded ncRNA binding

The binding of Ro60 to misfolded ncRNAs has been well characterized. There is increasing evidence to suggest that these Ro60-ncRNA interactions lead to either the degradation or correct folding of ncRNAs; this supports a quality control function for Ro60 in many different organisms, including bacteria, mice and humans (Chen et al., 2003, 2007; Hogg and Collins, 2007). The Ro60 protein is made up of two domains; a ringed donut-shaped domain made up of α-helical HEAT motifs and a von Willebrand Factor A (vWFA) domain (Stein et al., 2005). The HEAT motif helices fold to make the central cavity. The vWFA domain is comprised of a β sheet and a number of α helices (Stein et al., 2005). Within the β sheet is a metal-ion dependent adhesion site (MIDAS) motif. This motif may be important for structural changes to occur within the Ro60 protein (Stein et al., 2005).

In *Xenopus* oocytes Ro60 (xRo60) interacts with variant pre-5S rRNAs (O'Brien and Wolin, 1994). These RNAs usually have a 3' extension as a result of aberrant RNA polymerase III termination, as well as a number of mutations within the sequence of the RNA (O'Brien and Wolin, 1994). The majority of these mutations have been predicted to result in the destabilisation of existing helices and the formation of an alternative helix (O'Brien and Wolin, 1994; Shi et al., 1996). X-ray crystallography experiments indicate that the 3' single-stranded end of misfolded 5S rRNAs binds to the xRo60 basic inner surface, while a helical domain of these RNAs binds to a basic region on the outer surface of xRo60 (Fuchs et al., 2006).

A 3' tail of at least 5 nucleotides, along with a particular helical domain within the RNA, is key for this type of interaction with xRo60 (Fuchs et al., 2006). Altering the sequence of this helix, or a nearby loop region, appears to have a large effect on Ro60-binding efficiency, as Ro60 probably binds directly to this region (Fuchs et al., 2006). xRo60 is thought to have a large flexible misfolded ncRNA-binding surface, as sequence alterations in the rest of the variant pre-5S rRNA have little effect on xRo60 binding, as long as the secondary structure remains the same (Fuchs et al., 2006). *In vitro*, xRo60 can interact with wild type pre-5S rRNA, but Fuchs *et al.* hypothesize that this RNA would be unlikely to be available for binding *in vivo* and that xRo60 acts as a scavenger, picking up free misfolded RNA (Fuchs et al., 2006). Along with variant 5S rRNA, xRo60 can interact with the viral EBER1 RNA, as

well as *Saccharomyces cerevisiae* tRNA and U6 snRNA *in vitro*; all of these ncRNAs have 3' tails (Fuchs et al., 2006).

In mouse embryonic stem cells mouse Ro60 interacts with variant U2 snRNAs (Chen et al., 2003). 68% of these variant RNAs contain at least one sequence change. Experiments involving injecting mouse variant U2 snRNAs into *Xenopus* oocytes revealed that xRo60 interacts with variant pre-U2 and variant processed U2 snRNA, but does not interact with wild type U2 snRNA (Chen et al., 2003). Like the *Xenopus* variant pre-5S rRNA, Ro60-bound variant U2 snRNA appeared to commonly have two different mutations, resulting in the formation of an alternative helix in the RNA (Chen et al., 2003). In mouse cells Ro60 only associates with a small proportion of variant U2 snRNA, likely as a result of only a small proportion of these RNA molecules forming the required alternative helix (Chen et al., 2003). It is possible that mouse Ro60 also binds to wild type U2 snRNA at low levels to aid its folding (Chen et al., 2003). In *Xenopus* oocytes many 5S rRNAs are present, including variants. Many of these variants are encoded in the genome of *Xenopus* (Peterson et al., 1980). Similarly there are many variant versions of the U2 snRNA gene encoded in the mouse genome (Mouse Genome Sequencing Consortium, 2002). The ability of Ro60 to specifically recognize misfolded RNAs in different organisms is likely to have evolved to protect cells from the presence of high levels of these RNAs produced from variant genes (Chen et al., 2003).

1.4.3.2 Ro60 and Y RNA binding

Ro60 binds to regions 1 and 2 of each Y RNA (Wolin and Steitz, 1984) and a bulged cytosine that projects from the stem is important for this interaction (Simons et al., 1996). Mutation studies and chemical modification experiments give conflicting evidence as to whether the cytosine at position 8 (C8) or the cytosine at position 9 (C9) is important for Ro60 binding. Bearing in mind that this region can form either two bulges or a small internal loop, Green *et al.* hypothesized that either C8 or C9 could be free in the protein-free Y RNA, depending on its conformation. They further hypothesize that C9 is either base-paired or directly interacts with Ro60 when Ro60 binds to the Y RNA. This would explain why this nucleotide appears to be more important for Ro60 interaction than C8 (Green et al., 1998). Work by the Wolin group using xRo60 and *Xenopus* Y3 (xY3) RNA indicates that both bulges, or the alternative small loop structure that they can form, are also required for Ro60 binding (Green et al., 1998). The sequence of this region is not important, only the structure (Stein et al., 2005; Green et al., 1998). This group further suggest that the bulges are essential because they allow the secondary structure of the Y RNA to be open enough for Ro60 to enter and read the sequence in this region (Green et al., 1998). X-ray crystallography experiments have shown that *Xenopus*

Ro60 undergoes a conformational change when either a misfolded RNA or Y RNA associates (Stein et al., 2005). As well as Ro60, each Y RNA also undergoes a change in conformation as a result of this interaction (Green et al., 1998; Stein et al., 2005).

1.4.3.3 La

La associates with many different RNAs including Alu element RNAs, Pre-5S RNAs and pre-tRNAs (Rinke and Steitz, 1982; Shen and Maniatis, 1982). La also interacts with viral RNAs including Epstein-Barr virus non-coding RNAs and virus associated RNAs (Lerner et al., 1981a; Rosa et al., 1981; Lerner et al., 1981b). This protein usually interacts with RNA precursors, rather than RNAs in their mature form. Its ability to bind to a run of uridylates at the 3' end of target RNAs has been well characterized (Francoeur and Mathews, 1982; Reddy et al., 1983). La binding is involved in stabilization of nascent RNA polymerase III transcripts (Maraia et al., 1994a; Yoo and Wolin, 1997). This protein regulates tRNA maturation by binding to the 3' end of pre-tRNAs and preventing their degradation by exonucleases (Fan et al., 1998; Yoo and Wolin, 1997). La is also involved in the correct folding of pre-tRNAs, enabling further processing to produce mature tRNAs (Fan et al., 1998). La interaction can allow internal ribosome entry site (IRES)-mediated translation of viral RNAs (Meerovitch et al., 1993), prevent RNA export from the nucleus (Boelens et al., 1995) or aid RNA integration into RNPs (Gottlieb and Steitz, 1989a).

The La protein is made up of an N terminal domain known as the La motif (Sobel and Wolin, 1999), the nearby RNA recognition motif RRM1(Bandziulis et al., 1989; Query et al., 1989), and a carboxyl (C) terminal domain. The La motif and RRM1 work in combination to recognise and interact with the oligo(U) 3' terminal sequences of target RNAs (Ohndorf et al., 2001). There is also evidence suggesting that La may bind to 3' targets by recognising structural motifs at their 5' ends (Fan et al., 1998). The C terminal part of La has low sequence conservation among orthologs (Chambers et al., 1988; Chan et al., 1986). Vertebrate La proteins appear to have a number of domains within their C terminal ends, including an alternative RRM domain, RRM2 (Topfer et al., 1993; Jacks et al., 2003). This C terminal region of La has a certain amount of flexibility, allowing it to interact with a vast array of coding and non-coding RNAs (Kucera et al., 2011).

La interaction with RNAs that do not contain a run of uridylates at the 3' terminus requires a different mechanism of binding involving the La motif and RRM1, combined with the RRM2 domain (Martino et al., 2012). La has been shown to interact with both primary miRNAs (pri-miRNAs) and precursor-miRNAs (pre-miRNAs) via this 3' oligo(U)-independent binding mechanism (Liang et al., 2013). As with tRNAs, La is thought to protect the vast majority of pre-miRNAs from digestion by enzymes such as monocyte chemotactic protein-induced protein 1 (MCPIP1), al-

lowing theses pre-miRNAs to be exported from the nucleus and be passed to Dicer for processing (Liang et al., 2013). La binds to a minority of pri-miRNAs and is thought to be incorporated into the Drosha/DGCR8 complex as these pri-miRNAs are processed (Liang et al., 2013). Mutation studies indicate that a stem-loop structure is required for 3' oligo(U)-independent binding to pre-miRNAs and Liang *et al.* propose that this might be a more general mechanism for 3' oligo(U)-independent La binding (Liang et al., 2013).

The La protein is present in cells at much higher levels than Ro60 (Gottlieb and Steitz, 1989a) and plays many different roles within each cell. La is mainly located in the nucleus in the cells of many different organisms (Yoo and Wolin, 1994; Deng et al., 1981; Simons, 1994), although it also has functions in the cytoplasm (Meerovitch et al., 1993; Casciola-Rosen et al., 1994a). The human La protein can be phosphorylated on a Serine residue as position 366, altering its functional activity (Fan et al., 1997). Phosphorylated La is active in tRNA processing (Itine et al., 2000), while dephosphorylated La has the ability to stabilize nascent Pol III transcripts (Fan et al., 1997). This protein leaves the nucleus under environmental stress conditions, such as apoptosis (Casciola-Rosen et al., 1994a) or viral infection (Meerovitch et al., 1993). During apoptosis the majority of La is dephosphorylated, probably by a protein phosphatase 2A-like enzyme (Rutjes et al., 1999a). A small proportion of La is also cleaved at an Aspartate residue in the C terminus, resulting in the loss of a nuclear localisation signal and the relocation of La to the cytoplasm (Rutjes et al., 1999a; Shiroki et al., 1999). During apoptosis La cleavage prevents its association with, and the subsequent unwinding of, double-stranded RNAs (Xiao et al., 1994). This leaves these RNAs free to be targeted by protein kinase RNA-activated (PKR)(Rutjes et al., 1999a). La also has many other apoptosis-specific functions, such as enabling IRES-mediated translation of the cellular protein X-linked inhibitor of apoptosis (XIAP) (Holcik and Korneluk, 2000). This protein is important in cellular response to stress (Deveraux et al., 1997).

1.4.3.4 La and Y RNA binding

Y RNAs are RNA polymerase III transcripts and La binds to the 3' oligo(U) sequences in these transcripts and increases their stability, protecting them from degradation (Gottlieb and Steitz, 1989a,b). As well as the 3' end, La may also interact with hY3 and hY5 at other sites. It is currently unknown whether these other sites contain U-rich sequences (Pruijn et al., 1991). Many functions have been assigned to the La protein, but as yet none have been found to require Y RNA. So do Y RNAs work with La to perform a particular function or is this interaction solely important for Ro RNP biogenesis? Human Y RNAs are unusual in that they stay bound to La after the termination of transcription, while most other polymerase III transcripts

dissociate when the La-binding site is removed (Peek et al., 1993; Boire and Craft, 1990). La association is necessary for the binding of a number of other proteins to Y RNAs, so it may also play a role in the correct folding of Ro RNPs (Fabini et al., 2000, 2001; Belisova et al., 2005). It is possible that Y RNA binding to La may block certain functions of La, as La is less active in splicing when bound to hY1 (Belisova et al., 2005). The La protein can undergo a number of modifications and processing events, so it is possible that Y RNAs interact with a form of La that is not active in splicing, and play no role in blocking this function. La and Ro have both been implicated in the quality control of pre-5S rRNA (Shi et al., 1996) and, as already mentioned, hY5 aids in the recruitment of 5S rRNA to Ro60, so it may also aid any function of La in this process.

1.4.3.5 Nucleolin

In human cells Nucleolin binds preferentially to Y1 and Y3 (Fouraux et al., 2002; Langley et al., 2010). There is less conclusive evidence that this protein may bind to the other human Y RNAs (Fabini et al., 2000, 2001; Fouraux et al., 2002; Langley et al., 2010). Mutation studies show that Nucleolin interacts with Y RNAs via region 3, the large internal loop (Fabini et al., 2001; Fouraux et al., 2002). Nucleolin was named as it was originally isolated from the nucleolus, but is also present in other areas of the cell (Orrick et al., 1973; Borer et al., 1989). Nucleolin is involved in DNA repair and replication, chromatin modification and the regulation of rRNA transcription (Seinsoth et al., 2003; Yang et al., 2009; Angelov et al., 2006; Bouche et al., 1984). During cellular stress Nucleolin leaves the nucleolus and binds to Replication protein A (RPA) (Daniely and Borowiec, 2000). RPA is involved in chromosomal DNA replication, but this function is blocked when it interacts with Nucleolin (Daniely and Borowiec, 2000; Collins and Kelly, 1991). Perhaps Nucleolin also binds to Y RNAs during cellular stress. As with RPA, Y RNAs also play a vital role in chromosomal replication, as discussed later.

1.4.3.6 Calreticulin

Calreticulin (CRT) is another protein that has been associated with Ro RNPs (Cheng et al., 1996). This protein has been found to bind directly to Ro60 at two different sites (Staikou et al., 2003). It may also interact with hY RNAs directly (Cheng et al., 1996). CRT is mainly located in the endoplasmic reticulum, but also functions at the cell surface and in the nucleus (Opas et al., 1991; Burns et al., 1994; Dedhar, 1994). It is involved in many cellular processes including the accumulation of calcium and the regulation of gene expression (Opas et al., 1991; Burns et al., 1994; Dedhar, 1994). It can also be utilized by viruses for genome replication (Singh

et al., 1994). CRT causes the release of calcium from mitochondria during apoptosis (Nakamura et al., 2000). A calcium-dependent modification of CRT causes it to relocate to stress granules (Decca et al., 2007). It has also been linked to the autoantibody recognition of various endogenous antigens and may interact with Ro60 for this purpose (Staikou et al., 2003).

1.4.3.7 Heterogeneous nuclear ribonucleoproteins

In humans heterogeneous nuclear ribonucleoproteins (hnRNP) I and K only interact with Y1 and Y3 RNA and the ability to bind to Y4 and Y5 has not been demonstrated (Fabini et al., 2001; Fouraux et al., 2002). Both proteins have been linked to cellular processes, such as splicing, the regulation of RNA folding (Belisova et al., 2005) and translation (Ostareck et al., 1997). Like Nucleolin these proteins bind the large internal loop of hY1 and hY3. hnRNP I and K are unable to bind in the absence of La (Fabini et al., 2001; Fouraux et al., 2002). HnRNP K is a key player in cellular response to apoptosis, as it regulates the transcription of XIAP (Xiao et al., 2013). HnRNP I (also known as polypyrimidine tract binding protein) has recently been shown to regulate miRNA activity (Engels et al., 2012). It has also been found to colocalize with human Y RNAs in perinucleolar compartments (discussed later) (Matera et al., 1995)

1.4.3.8 Ro RNP binding protein I

Ro RNP binding protein I (RoBPI) is also known as Poly-U Binding Splicing Factor 60 KDa (PUF60). In human cells RoBPI mainly associates with Y5 (Bouffard et al., 2000), but has also been shown to bind human Y1 and Y3 (Hogg and Collins, 2007). Most of RoBPI is located in the nucleus (Page-McCaw et al., 1999). RoBPI, in association with the splicing protein p54, is involved in the splicing of U2 snRNA (Page-McCaw et al., 1999). RoBPI contains a PUMP domain, which enables protein-protein interactions but may also be important for RNA binding (Page-McCaw et al., 1999). RoBPI also has two RRM domains, which enable RNA binding (Page-McCaw et al., 1999).

1.4.3.9 Interferon-induced protein with tetratricopeptide repeats 5

Interferon-induced protein with tetratricopeptide repeats 5 (IFIT5) was found to interact with only Y5 of the Y RNA family, although it interacts with a number of other small ncRNAs (Hogg and Collins, 2007). Only this year this protein has been crystallized and fully characterized (Abbas et al., 2013; Katibah et al., 2013; Feng et al., 2013). It is able to bind single-stranded A or U rich RNA or double-stranded AT rich DNA (Feng et al., 2013). Interestingly, like the Y5 RNA, this

protein is not found in mouse cells, but is found in humans and *Xenopus* (Feng et al., 2013). It is thought that the main function of IFIT5 is in cellular response to viral infection. ITIF5 recognises the 5' triphosphates in a short 5' overhang of single-stranded viral RNAs (Abbas et al., 2013) and has also been shown to bind to cellular 5' phosphorylated RNAs, such as tRNAs (Katibah et al., 2013; Hogg and Collins, 2007).

1.4.3.10 Ribosomal protein L5

Ribosomal protein L5 is another protein that binds to human Y5 RNA specifically. L5 is involved in shuttling 5S rRNA around the cell (Steitz et al., 1988) and it is believed to be involved in recruiting variant 5S rRNA to hY5-containing Ro RNPs (Hogg and Collins, 2007), although *in vitro* 5S rRNA can interact with hY5 in the absence of the L5 protein (Hogg and Collins, 2007).

1.4.3.11 Zipcode binding protein I

Zipcode-binding protein I (ZBPI) is located in the cytoplasm under normal conditions, but under conditions such as UV damage, it moves into the nucleus of mouse cells (Sim et al., 2011). ZBPI interacts with the mouse Y3 (mY3) RNA and is involved in shuttling the mY3-Ro60 complex out of the nucleus (Sim et al., 2011). Interaction of this protein with the hY RNAs has yet to be examined, but it is likely that these RNAs are exported via several nuclear export pathways.

1.4.3.12 Polynucleotide phosphorylase

In *Deinococcus radiodurans* the polynucleotide phosphorylase (PNPase) enzyme degrades structured ncRNAs. It interacts with a Ro60 ortholog known as Ro sixty-related (Rsr) and bacterial Y RNA and forms a complex to carry out this function (Chen et al., 2013). The bacterial Y RNA is key for the protein interactions of this degradation complex (Chen et al., 2013).

1.4.3.13 Other Y RNA-binding proteins

There are many more Y RNA-interacting proteins that have recently been identified but the function of these associations have not been thoroughly elucidated. MOV10 interacts with Ro60 in an RNA-dependent manner, which is likely to involve Y RNAs (Sim et al., 2011). YB-1 is also believed to interact with Ro60 in a Y RNA-dependent manner (Sim et al., 2011), but again further characterization is required. Also many proteins involved in the initiation of chromosomal RNA replication have been found to directly bind to Y RNAs (as described later), but the exact role of Y RNAs in this process is still currently unknown (Zhang et al., 2011a).

Every human Y RNA was originally thought to stably interact with both Ro60 and La to form Ro RNPs (Hendrick et al., 1981; Wolin and Steitz, 1984). More recently it has been suggested that not all Y RNAs take on this form and that they may have Ro60 and La independent functions (Langley et al., 2010). As discussed above, an expanding array of proteins appear to bind to Y RNAs in a specific manner, rather than binding to all with equal efficiency. More and more groups are now proposing that each Y RNA may be incorporated into a distinct RNP complex to carry out a particular function (Gendron et al., 2001; Fabini et al., 2001; Fouraux et al., 2002; Langley et al., 2010). What is still controversial, however, is the idea that Ro60 and La may not be included in every Y RNP (Langley et al., 2010).

Despite the suggestion that so many specialized Y RNA-containing RNP complexes exist, very few clear functions for these have been identified. Perhaps Y RNAs simply act as a mechanism for storing proteins within the cell until they are needed, for example under stress conditions. Y RNAs could recruit proteins so that a large store is available when released from the Y RNP. Alternatively, these proteins may have some function that is blocked by Y RNA binding. There are numerous interactions between Y RNA-binding proteins and other proteins/nucleic acids in subcellular locations other than the cytoplasm (Cheng et al., 1996; Staikou et al., 2003; Itine et al., 2000). For example, only the unphosphorylated form CRT associates with Ro RNPs (Cheng et al., 1996), but its phosphorylated form is known to be active in viral RNA replication (Singh et al., 1994). Phosphorylated La interacts with Y RNAs, but the phosphate group must be removed before this protein is active in tRNA maturation (Fan et al., 1998). Y RNA binding may also block the modification of these proteins, altering their function. Each Y RNA is unlikely to bind the majority of each of its protein partners, but many of these proteins are involved in cell stress, so proteins bound to Y RNAs may be stored for this purpose.

1.4.3.14 Y RNP localization

Human Y RNAs, when injected into the nuclei of *Xenopus* oocytes, are bound in Ro RNPs and exported in a Ro60-dependent manner (Simons, 1994). The Ro60 binding site however, is not enough for export, as deletion of region 3 of hY1 blocks export into the cytoplasm (Rutjes et al., 2001). Rutjes *et. al.* speculate that a nuclear factor interacts with the RNA in this region and this factor is essential for nuclear export of hY1 (Rutjes et al., 2001). Testing of the other human RNAs in *Xenopus* oocytes yielded similar results, implicating this pathway in the export of multiple Ro RNPs. This nuclear factor is likely to be ZBP1, which was found to be involved in the export of Ro60-mY3 in mouse cells (Sim et al., 2011). ZBP1 associates more strongly with Ro60 in the nucleus after UV irradiation and this complex is then exported into the cytoplasm. The mY3-Ro60 complex blocks a Chromosome region

maintenance 1 protein (CRM1)-independent nuclear export signal in ZBP1, so that this complex relies on CRM1-dependent export only (Sim et al., 2011). Mouse Y1, however, was found to be exported via a different pathway in this study (Sim et al., 2011).

Ro60 is both nuclear and cytoplasmic (Xia et al., 1987; Peek et al., 1993). It shuttles into and out of the nucleus depending upon environmental conditions. When mouse Y RNAs are knocked down using siRNAs then Ro60 relocates to the nucleus (Sim et al., 2009). Sim *et al.* made various mouse/*D. radiodurans* fusion proteins and found that the mouse HEAT repeat domain is required for the stabilization of mouse Y RNAs and for the accumulation of Ro60 in the nucleus after UV irradiation (Sim et al., 2009). Mouse Ro60 was also found to relocate to the nucleus after oxidative stress resulting from hydrogen peroxide treatment (Sim et al., 2009). They determined that the nuclear localisation signal is likely to reside in helices 1-12 and that these helices, in the context of a correctly folded Ro60 protein, are required for stress-dependent relocation of Ro60 (Sim et al., 2009). Interestingly, helices 14 and 16 of mRo60 are required for Y RNA binding, and its interaction with these RNAs is thought to mask the nuclear localization signal (Sim et al., 2009). Y RNA association with Ro60 must therefore be altered after UV treatment, possibly in a similar manner to Rsr-drY RNA when the PNPase enzyme binds (Chen et al., 2013). It is unclear what proportion of mY RNAs enter the nucleus with mRo60 during UV irradiation (O'Brien et al., 1993; Peek et al., 1993; Chen et al., 2003; Sim et al., 2009), but it seems likely that their regulation of Ro60 localization is dependent upon the other proteins present in the complex (Sim et al., 2009)

La is mainly nuclear and contains a nuclear localization signal (Simons, 1994), but is detectable in the cytoplasm where it is present in Ro RNPs (Peek et al., 1993; Boire and Craft, 1990). The alterations necessary to allow La to exit the nucleus are not yet understood. Mutant human Y RNAs lacking the La binding site are more rapidly exported from the nucleus, suggesting that La is involved in the nuclear retention of these RNAs (Simons et al., 1996). The study of human Y RNA export in *Xenopus* oocytes is complicated by the fact that Y RNAs undergo processing in these cells, and both xY RNAs and hY RNAs in the cytoplasm of *Xenopus* oocytes are not associated with La and are missing their terminal oligo(U) sequences (O'Brien et al., 1993; Simons et al., 1996). Y RNAs in human cells are known to be present in the cytoplasm, where they are stably bound to La (Peek et al., 1993; Boire and Craft, 1990). In certain cell lines, hY1, hY3 and hY5 have also been detected in discrete nuclear locations known as perinucleolar compartments (PNCs) (Matera et al., 1995). The function of these PNCs is unknown, but it has been suggested that they represent sites of RNA polymerase III transcript sequestration prior to nuclear export or following nuclear import (Matera et al., 1995). This

subset of Y RNAs is Ro60-independent and may also interact with hnRNP I, which colocalizes to PNCs (Matera et al., 1995).

1.4.4 Y RNA function

Although there are numerous groups working on this non-coding RNA family, only two functions have been experimentally validated. The first function is in semi-conservative DNA replication. The second is in Ro60 quality control.

1.4.4.1 Y RNAs in DNA replication

Before a cell divides, it must duplicate its DNA in order for a copy to enter into each daughter cell. During semi-conservative DNA replication the DNA strands separate and each is used as a template to produce a new strand. The new DNA molecules consist of one strand from the existing DNA and one strand of nascent DNA (Watson and Crick, 1953; Meselson and Stahl, 1958). Using a cell-free system as a model, Krude and colleagues evaluated the requirements for chromosomal DNA replication (Christov et al., 2006). In this cell-free system, the nuclei are removed from cells at the G1 stage of the cell cycle (Krude et al., 1997). These cells are ready to undergo replication, but have not yet begun to replicate their DNA. Cytosolic extracts from replicating cells in S phase of the cell cycle can be added to these nuclei, resulting in the stimulation of chromosomal replication (Krude et al., 1997). The cytosolic extracts can be fractionated and different fractions added to the nuclei to determine which are necessary for replication. Using this method, Y RNAs were found to be crucial for the replication process (Christov et al., 2006). The exact role of these RNAs has not yet been determined, but they have recently been found to be essential for the initiation of replication and are not thought to be necessary for elongation of the nascent DNA strand (Gardiner et al., 2009; Krude et al., 2009). Human Y RNAs are recruited to chromatin by their association with protein complexes involved in DNA replication initiation (Zhang et al., 2011a).

Further studies by the Krude group identified that the human Y RNAs act redundantly in their role during replication, as addition of any one of the hY RNAs in the absence of any of the others still stimulates semi-conservative DNA replication (Christov et al., 2006). In fact, the block in replication resulting from targeted knock-down of a human Y RNA can also be rescued by the addition of any vertebrate Y RNA, suggesting that their function in replication is conserved across all vertebrate Y RNAs (Gardiner et al., 2009). Non-vertebrate Y RNAs cannot stimulate replication in human cells (Gardiner et al., 2009).

As already discussed, certain regions of Y RNAs are more conserved than others. Analysis of this conservation can give some indication of which regions are present

in the majority of vertebrate Y RNAs. Mutagenesis studies, combined with this analysis, were used to narrow the region essential for replication down to a short duplex within region 3 of each Y RNA (Gardiner et al., 2009). Addition of this short double-stranded duplex alone is sufficient to stimulate DNA replication in a human cell depleted of one of the hY RNAs. Intriguingly this duplex must be in an RNA form, as the addition of the same sequence constructed of DNA did not initiate replication (Gardiner et al., 2009). In the presence of this conserved duplex alone DNA replication can occur, but this duplex was found to associate at many locations on euchromatin and heterochromatin. Full length Y RNAs associate with regions of the DNA where replication has yet to occur and dissociate on the initiation of replication at these sites (Zhang et al., 2011a). The RNA region encompassing the Ro60 binding site, along with the large internal loop of each Y RNA, is required for specific euchromatin association, but is not sufficient for DNA replication to occur (Zhang et al., 2011a). A number of regions of the Y RNA must therefore be necessary for efficient DNA replication in mammalian cells. hY1 and hY3 associate with euchromatin in the nucleus at distinct, but overlapping sites. hY5 associates with euchromatin at different sites to hY1 and hY1 was found to cause hY5 dissociation. This suggests that, although all hY RNAs have a redundant function in chromosomal RNA replication, certain Y RNAs may do this more efficiently while others have become specialized for other functions.

The majority of Y RNAs are believed to be in a complex with both Ro60 and La (Hendrick et al., 1981). Experiments in mouse cells that do not express Ro60 indicate that chromosomal replication can still occur in the absence of Ro60 (Chen et al., 2003; Xue et al., 2003a). Y RNA mutants lacking La/Ro60 binding sites can still initiate chromosomal replication, providing further evidence that neither of these proteins are directly required (Langley et al., 2010). Along with the necessity of Y RNAs for cellular DNA replication, the Murine Leukaemia Virus incorporates immature hY RNA transcripts into new virus particles (Garcia et al., 2009), where they also presumably have a function in viral genome replication (Langley et al., 2010). Ro60 and La are not present in these particles, further indicating that Y RNAs have a function independent of either of these two proteins (Garcia et al., 2009; Langley et al., 2010). Proteins involved in the initiation of DNA replication are believed to bind only to Ro60/La/ Nucleolin-free Y RNAs (Langley et al., 2010). Y RNAs may have a level of redundancy in their ability to initiate DNA replication (Gardiner et al., 2009), but it is possible that, under normal circumstances, hY5 undertakes a unique role in the nucleolus, such as 5S rRNA quality control (Zhang et al., 2011a). So far, the role of Y RNAs in DNA replication does not appear to have been validated by any other group. Nonetheless their involvement in semi-conservative DNA replication is most scrutinized and the most convincing function

directly involving human Y RNAs to date.

1.4.4.2 Y RNAs in RNPs

The binding of Ro60 homologs to Y RNAs in nematodes and in mice increases the stability of these RNAs (Labbé et al., 2000; Chen et al., 2003). Therefore, by regulating RNA levels, Ro60 is likely to at least indirectly play a role in the functioning of Y RNA within the cell.

1.4.4.3 Y RNAs and Ro60 function

Green *et al.* found that under normal circumstances, Y RNAs compete with misfolded ncRNAs for access to Ro60 and have a higher binding affinity for Ro60 than these ncRNAs (Green et al., 1998). In 2006 partially overlapping binding sites for Y RNAs and ncRNAs were identified on the Ro60 protein (Fuchs et al., 2006). It therefore seems likely that these RNAs are acting as "gatekeepers" for Ro60 function, allowing access only when necessary (Hogg and Collins, 2007; Chen et al., 2013).

In *D. radiodurans*, Y RNAs can recruit the exonuclease PNPase to the Rsr RNP (Wurtmann and Wolin, 2010). Binding of this enzyme to the large internal loop in region 4 leads the Y RNA to undergo a conformational change in relation to Ro60, opening up the area around the central cavity of this protein to allow ncRNAs access. A misfolded ncRNA enters the Ro60 central cavity, where it is then passed on to PNPase and is targeted for degradation (Wurtmann and Wolin, 2010). When PNPase interacts with Rsr/drY RNA, its targets are altered as it has greater a affinity for structured RNAs, but a reduced preference for single stranded RNAs (Chen et al., 2013). This Rsr-Y RNA-PNPase interaction is conserved in other bacteria as it is also found in *Salmonella typhimurium* (Chen et al., 2013). Further evidence for a recruitment function can be seen in human cells, where hY5 interacts with misfolded 5S ribosomal RNA through a protein intermediate, the ribosomal protein L5 (Hogg and Collins, 2007). It seems likely that hY5 is recruiting this ribosomal RNA to Ro60 for quality control purposes, and that perhaps the other human Y RNAs have similar interactions with intermediate proteins, as well as with RNA processing/refolding proteins (Hogg and Collins, 2007). Prior to this work the community was torn as to whether Y RNAs helped or hindered the binding of Ro proteins to ncRNAs.

In *D. radiodurans* cells the Ro60 ortholog Rsr is known to be involved in the targeting and maturation of 23S rRNA and this processing involves the exonucleases RNasePH and RNaseII (Chen et al., 2007). The bacterial Y RNA (drY RNA) binds to Rsr at 30°C and prevents interaction with 23S rRNA. At higher temperatures Rsr becomes active, probably by a temperature-dependent drY RNA release mechanism

(Chen et al., 2007). It is also possible that, as with mammalian Y RNAs, drY RNA is not released from this Ro ortholog, but undergoes a structural change at higher temperatures. Bacterial cells lacking the drY RNA can still aid 23S rRNA maturation (Chen et al., 2007), but this could be an artifact and drY may normally function as a gatekeeper of Rsr function when present (Hogg and Collins, 2007; Chen et al., 2013).

D. radiodurans is a bacterium that is resistant to UV-damage. As well as playing a role in RNA processing, Rsr is also key for this resistance (Chen et al., 2000). Knocking out Rsr results in bacterial strains that are more sensitive to UV light. Rsr interacts with a number of small RNAs after UV damage, including drY RNA (Chen et al., 2000). In *D. radiodurans* Rsr levels increase after UV irradiation (Chen et al., 2000), whereas in mouse cells Ro60 levels remain the same, but Ro60 relocates to the nucleus (Chen et al., 2003). The role of Ro60 orthologs in bacterial or mouse cells in response to UV remains to be determined. As with mouse Ro60, mouse Y3 also relocates to the nucleus on UV irradiation, suggesting that it is involved in the UV function of Ro60 in these cells (Chen et al., 2003).

1.4.4.4 Y RNAs in splicing

La and the Cellular nucleic acid binding protein (CNBP) both target the 5' untranslated region (UTR) of mRNAs that contain 5' terminal oligopyrimidine sequences (5' TOPs) (Pellizzoni et al., 1998). In experiments where binding to the L4 ribosomal mRNA was tested, La and CNBP were found to bind to this mRNA mutually exclusively, despite the fact that they have different binding sites (Pellizzoni et al., 1998). La binds to 5' TOP mRNAs via a mechanism that involves both RRMs and possibly the La motif (Pellizzoni et al., 1998). The interaction of either La or CNBP requires Ro60 and this interaction is RNA-dependent (Pellizzoni et al., 1998). As Ro RNPs are known to be stable and can include La, it seems highly likely that this RNA component is a Y RNA. One caveat with this theory is that binding experiments were conducted using the *Xenopus* system and xY RNAs dissociate from Y RNAs in the nucleus under normal circumstances, so the conditions for La interaction with 5' TOP mRNAs must be further investigated. Studies in patients with decreased levels of CNBP indicate that this protein positively regulates the translation of 5' TOP mRNAs (Huichalaf et al., 2009) but the effect of La on 5' TOP mRNAs expression has not yet been categorically determined (Crosio et al., 2000; Zhu et al., 2001). Recent studies of stress-regulated translation of 5'TOP mRNAs indicate that it is extremely complex and that expression is likely to be cell-type and tissue specific (Damgaard and Lykke-Andersen, 2011; Ivanov et al., 2011b).

1.4.4.5 Y RNAs and Apoptosis

In 1999 Rujtes and coworkers discovered that during apoptosis human Y RNAs are cleaved to produce products of between 22 and 36 nucleotides (Rutjes et al., 1999b). These fragments are detectable as early as 1.5 hours after treatment and follow a similar pattern to the cleavage of the U1-70K protein, which is known to undergo caspase 3-dependent cleavage (Casciola-Rosen et al., 1994b). The hY RNAs appear to be cleaved in a size dependent manner, with the longer RNAs targeted more rapidly than the smaller ones (Rutjes et al., 1999b). Further investigation by this group using various inhibitors of specific caspases determined that this cleavage appears to be caspase-dependent. When inducing apoptosis using anti-Fas antibodies, they discovered that all of the caspase inhibitors blocking the death receptor apoptosis pathway delayed the accumulation of these Y RNA products (Rutjes et al., 1999b). The group concluded that, had the mitochondrial apoptosis pathway been activated, it is likely that fragment production would also be delayed by inhibitors of this pathway (Rutjes et al., 1999b). Similar cleavage patterns were seen with different inducers of apoptosis and in a number of human cell lines, as well as one mouse cell line. Many of the cleavage products were still able to bind to Ro60 and La (Rutjes et al., 1999b).

1.4.4.6 Y RNAs as microRNAs?

Since the publication of the Rutjes paper in 1999 there has been much speculation as to the function of these RNA fragments (Rutjes et al., 1999b). Records were created in miRBase for a number of Y RNA fragments, although a miRNA-like function has never been proved. Work by our group (Nicolas et al., 2012) indicates that these fragments are not Dicer-dependent, while the majority of mammalian miRNAs are Dicer-dependent. Immunoprecipitation experiments did not detect any interaction with Argonaute 2, suggesting that they do not associate with the RISC complex (Nicolas et al., 2012). In anion exchange separation experiments these fragments were detected in protein complexes that were distinct from miRNA-containing protein complexes (Nicolas et al., 2012). All of these lines of experimentation together provide strong evidence that these fragments are unlikely to act as miRNAs (Nicolas et al., 2012).

This PhD project will focus on the validation of other Poly(I:C) induced 30-35 nucleotide ncRNA-derived sRNAs, as well as examining the sequence and structural requirements for the generation of Y RNA-derived fragments (ydRNAs) from the hY5 RNA.

2 MATERIALS AND METHODS

2.1 CELL CULTURE

2.1.1 Cell lines and cell culture media

Cell lines were grown in culture incubators at 5% CO_2 and at 37°C, unless otherwise stated. Each line was cultured in the appropriate cell culture medium. Cell culture media and reagents were obtained from Life Technologies, unless otherwise stated.

NIH/3T3, SW 1353, HeLa and MCF7 cell lines were cultured in Dubeccos modified eagles medium (DMEM) supplemented with 10% (v/v) standard quality filter-sterilized and heat-inactivated fetal bovine serum (FBS) (GE Healthcare) and 2 mM L-glutamine.

The DLD-1 cell line was cultured in a 1:1 mixture of DMEM GlutaMAX I and Hams F12 GlutaMAX I, supplemented with 10% FBS and 2mM L-glutamine. The DF-1 cell line was cultured in DMEM GlutaMAX I, supplemented with 10% FBS and 2 mM L- glutamine.

2.1.2 Passaging of cells

When passaging, culture media was removed and cells were briefly washed in Dubeccos phosphate buffered saline without calcium or magnesium (DPBS) before 0.25% (v/v) trypsin-EDTA was added to cover the cells. The cells were incubated at 37°C until detached. The trypsin-EDTA was neutralized with an equal volume of the appropriate culture media. Cells were centrifuged at 250 x g for 4 minutes and the media removed, before re-suspension in fresh media and transfer to a new flask at the appropriate dilution.

2.1.3 Seeding of cells

For each cell line the required number of cells were seeded in the appropriate media 1 to 3 days prior to transfection. Cells were always seeded in the absence of antibiotics. An equal volume of 0.4% tryphan blue (Sigma-Aldrich) was added to the media containing the cells and a haemocytometer was used to calculate the cell number. Tryphan blue can only enter dead cells, so any cells stained with tryphan blue were excluded from the cell count.

2.1.4 Transfection of cells

2.1.4.1 Transient transfection of cells with Lipofectamine2000

Cells were transfected with Lipofectamine2000 (Life Technologies) according to manufacturer's instructions. Transfection mixes were prepared in OptimMEM I reduced

serum with GlutaMAX I medium (OptiMEM) and added drop-wise onto the cells. The OptiMEM was replaced with the appropriate FBS-containing growth medium 4-6 hours after transfection. For DF-1 cells the media was changed 3 hours after transfection.

5-20 μg of plasmid DNA was added to each T75cm^2 flask and Lipofectamine2000 was used at a ratio of 2.5:1, with 25 μL added for every 10 μg of plasmid DNA. For treatment of cells with Polyinosinic:polycytidylic acid potassium salt [Poly(I:C)] (Sigma-Aldrich) transfection mixes containing 30 μL of Lipofectamine2000 and 24 μL of Poly(I:C) were prepared for each flask, giving a final concentration of 10 μg/mL. A Lipofectamine2000 control-transfected flask was included in all experiments. Where appropriate, empty vector and untreated control transfections were also included.

2.1.4.2 Transient transfection of cells with Fugene6

Cells were transfected with Fugene6 (Promega) according to manufacturer's instructions. Transfection mixes were prepared in OptiMEM. The media in all flasks was replaced with the appropriate FBS-containing growth media containing FBS prior to transfection. Transfection mixes were diluted in the appropriate cell culture medium immediately prior to transfection.

5-20 μg of plasmid DNA was added to each T75cm^2 flask and Fugene6 was used at a ratio of 3:1, with 15 μL added for every 5 μg of plasmid DNA. A Fugene6 control transfected flask was included in all experiments. Where appropriate, empty vector and untreated control transfections were also included.

2.1.5 Stress treatments

2.1.5.1 Hygromycin treatment

Cells were treated with Hygromycin B solution (Melford) at 800 μg/mL diluted in the appropriate cell culture media. Cells were grown for between 8 hours and 3 days before harvesting. Hygromycin is an apoptosis inducer.

2.1.5.2 PBS treatment

Cells were treated with DPBS for 30-60 minutes before harvesting.

2.1.5.3 Poly(I:C) treatment

Poly(I:C) stocks were prepared by Dr. Hall at a concentration of 2.5 mg/mL. Cells were transfected with 10 μg/mL Poly(I:C) unless otherwise stated (see Section

2.1.4.1 for details of transfection). Cells were grown for 4-8 hours before harvesting. Poly(I:C) is a viral infection mimetic.

2.1.5.4 Staurosporine treatment

Staurosporine (Cell Signaling Technology) stocks were prepared at a concentration of 100 μM, Staurosporine was diluted from stocks in the appropriate cell culture medium. Cells were treated with a final concentration of 1 μM Staurosporine for 4 hours prior to harvesting.

2.1.5.5 Lipofectamine2000 control treatment

Cells were transfected with Lipofectamine2000 in the absence of Poly(I:C) for 4-8 hours before harvesting.

2.1.5.6 Untreated control conditions

Cells were grown in the absence of transfection mixes for 4-8 hours before harvesting.

2.1.6 Flow cytometry of cells treated with Poly(I:C) or Hygromycin

DLD-1 cells were plated at the appropriate number in 3 wells of a 6 well plate per condition. 24 hours after plating, the cells were treated with Hygromycin at 800 μg/mL or the media was replaced with fresh media. For Poly(I:C) treatment, cells were transfected with 10 μg/mL, Lipofectamine2000 mixes alone, or were left untreated 24 hours after plating. 55.5 hours after plating, cells were treated with DPBS. After a further 30 minutes, cells for all conditions were harvested by trypsinization, PBS washed and split into two tubes. For Annexin V/Propidium iodide (PI) flow cytometry analysis, cells were re-suspended in PBS with 200 mM calcium chloride. 1 μL of Annexin V (Invitrogen) was added and the samples were incubated at room temperature, in the dark, for 15 minutes. 1 μL of PI (Invitrogen) was added per tube and samples were loaded onto the Accuri C6 (BD Biosciences) flow cytometry machine. 2 wells of each treatment were harvested as normal for RNA extraction, 0.5 mL of TRIreagent was added to each well and both wells were combined to give 1 mL of TRIreagent per sample. During the extraction the amount of chloroform and isopropanol were scaled accordingly. As this was an initial test experiment, only one replicate of each condition was tested.

2.2 RNA EXTRACTION

2.2.1 RNA extraction from cell lines for northern blot analysis

Prior to extraction, all pipettes and work surfaces were rinsed with RNaseZAP (Sigma-Aldrich), wiped over and dried. All solutions except the TRIreagent (ThermoFisher Scientific) were sterilized by being passed through a 0.2 μm syringe needle (Life Technologies).

Each flask was processed individually. Media was completely removed, the flask placed on ice and TRIreagent was added drop-wise onto the flask surface. The TRIreagent was pipetted up and down several times and the flask surface was scraped using a disposable cell scraper (Corning) to ensure complete cell lysis. Where transfections/treatments resulted in cell lifting, the cell culture media was retained on harvesting and centrifuged at 250 x g for 4 minutes before media removal and re-suspension in TRIreagent. Each floating cell fraction was then recombined with the attached cell fraction. Samples were snap frozen on liquid nitrogen and stored at -80°C prior to downstream processing.

Samples were thawed on ice and centrifuged at 12,000 x g at 4°C for 5 minutes. Then the supernatant was transferred to a new 2 mL centrifuge tube. Samples were incubated at room temperature for 5 minutes, prior to the addition of 300 μL of chloroform (ThermoFisher Scientific). The samples were mixed by vortexing for 20 seconds, before being incubated at room temperature for a further 2 minutes, followed by centrifugation at 12000 x g for 15 minutes at 4°C. On ice, the aqueous phase of each sample was removed and transferred to a tube containing 750 μl isopropanol (ThermoFisher Scientific). To aid in identification of the RNA pellet, 0.5 μL of glycoblue (Life Technologies) was added to each tube, before mixing by inversion and incubation at -20°C overnight, or for at least 2 hours. Samples were then centrifuged at 12000 x g for 10 minutes at 4°C and the supernatant removed. Samples were washed in 1 mL of 80% ethanol twice, before being air dried for 5-10 minutes and re-suspended in Analytical Reagent Grade (ARG) water (ThermoFisher Scientific).

2.2.2 RNA extraction from cell lines for sequencing libraries

These samples were processed as in Section 2.2.1 but with an extra chloroform wash step prior to the isopropanol wash. Briefly, 300 μL of chloroform was added to a fresh tube at room temperature. The aqueous phase was removed and added to the tube containing chloroform. The samples were vortexed for 20 seconds and incubated at room temperature for 2 minutes prior to centrifugation at 12,000 x g for 15 minutes at 4°C. The aqueous phase was removed and added to the tube

containing isopropanol.

2.2.3 RNA quality control

The quality of the RNA was examined by agarose gel electrophoresis to confirm that the ribosomal RNA species were intact and that no degradation had occurred. Samples were diluted at a ratio of 1:1 in stop mix (95% (v/v) formamide, 5 mM EDTA, 1.9 mM xylene cyanol, 1.5 mM Bromophenol blue) and denatured by heating at 70°C for 2 minutes prior to gel electrophoresis. Samples were run on a 1.5% (w/v) agarose gel prepared in 0.5 x (v/v) Tris/Borate/EDTA (TBE) buffer (45 mM Tris, 45 mM Boric acid, 1 mM EDTA) and stained with 10 mg/mL ethidium bromide. Samples were imaged on a transilluminator (UPV) and analyzed using the Labworks Image Acquisition and Analysis software (UPV). The RNA concentration and optical density was determined by use of the Nanodrop 8000 spectrophotometer (ThermoFisher Scientific) and an optical density 260/280 ratio of 1.7 or above was accepted.

2.3 POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE) AND NORTHERN BLOTTING FOR sRNA

Unless otherwise stated, all reagents and apparatus were obtained from ThermoFisher Scientific. Prior to use, all pipettes, equipment and work surfaces were rinsed with RNaseZAP, wiped over and dried.

2.3.1 PAGE sample preparation

The required amount of RNA, 1-10 μg, was diluted in an equal volume of stop mix (95% (v/v) formamide, 5 mM EDTA, 1.9 mM xylene cyanol, 1.5 mM Bromophenol blue) and heated to 70°C for 2 minutes prior to loading.

2.3.2 PAGE separation of small RNA

Total RNA was size separated by gel electrophoresis using the mini PROTEAN III system (BIORAD). To prepare a 12% denaturing polyacrylamide gel, 6.3 g of urea was heated in 3 mL of ARG water in a microwave for 20 seconds, or until fully dissolved. The solution was allowed to cool and then the following were added; 1.5 mL 5 x TBE, 4.5 mL 40% (v/v) 19:1 Acrylamide/Bis solution (BIORAD). ARG water was added up to a volume of 15 mL and polymerization was induced by

the addition of 150 μL of 10% Ammonium persufate (APS) and 7.5 μL of tetramethyl-ethylenediamine (TEMED). Each gel was rapidly poured between two clean glass plates, a 1.0 mm 10 well comb was inserted, and left for 30 minutes to fully polymerize. Gels were run for 1.5 hours at 110 V in 0.5 x TBE. RNA quality and loading was confirmed by staining with 1 μL of 10 mg/mL ethidium bromide prior to transfer.

2.3.3 Semi-dry transfer

On the base of the semi-dry transfer apparatus (Scie-Plas) three BLT2 filter papers (Munktell), pre-soaked in 0.5 x TBE, were placed on top of each other and any air bubbles were removed. For each gel, a hybond-NX (GE Healthcare) membrane of the same size was soaked in ARG water and then 0.5 x TBE before being placed on top of the filter papers. Polyacrylamide gels were removed from the electrophoretic apparatus and, after ethidium bromide staining, were placed on top of the corresponding membranes. Finally, three 0.5 x TBE-soaked filter papers were placed on top of each polyacrylamide gel and air bubbles were removed. The top of the semi-dry apparatus was added and transfer of the RNA from the gel to the membrane was performed at 350 mA, at 4°C, for 30 minutes per polyacrylamide gel.

2.3.4 Chemical cross-linking

Chemical cross-linking was used for small RNA northern blotting as it leads to more effective hybridization than UV crosslinking. 122.5 μL of 12.5 M 1-methylimidazole was added to 10 mL of ARG water and mixed by inversion. The pH of the solution was adjusted to 8 by the addition of 6 M hydrochloric acid. 0.373 g of 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) was dissolved in this solution and the volume adjusted to 12 mL with ARG water. For each membrane, a filter paper slightly larger in size was soaked in cross-linking solution and then the membrane was placed, RNA side up, on top of this. Excess cross-linking solution was removed and then the filter paper and membrane were wrapped in cling film and incubated at 60°C for 1-2 hours.

2.3.5 Northern blotting

After cross-linking, membranes were briefly washed in ARG water before storage at -20°C or direct use in northern blotting.

2.3.5.1 Pre-hybridization and radioactive probe labeling

Prior to incubation with a radiolabeled probe, membranes were placed, RNA side facing inward, into hybridization tubes containing 10-15 mL of UltraHyb-Oligo hybridization buffer (Ambion) and rotated at 37°C in a HB-1000 hybridization oven (UVP) for 1.5 hours. During this time, for each membrane, an oligonucleotide complementary to the small RNA of interest was labeled with the γ^{-32}P from adenosine 5'-γ^{-32}P triphosphate triethylammonium salt (Perkin Elmer). Each reaction had a final concentration of 10 units of T4 polynucleotide kinase (PNK), 1 x kinase buffer, 0.5 mM oligonucleotide, and 1.1 MBq of γ^{-32} ATP. Immediately after the addition of γ^{-32}ATP, the reactions were incubated at 37°C for 1 hour.

2.3.5.2 Hybridization

After the pre-hybridization and radioactive probe labeling was complete, each probe reaction was diluted in a small amount of ARG water and transferred to the appropriate hybridization tube. Membranes were incubated with radiolabeled probe in a rotating HB-1000 hybridization oven, at 37°C, for 16 hours or overnight (see Table 2.3 for details of probes used for northern blot analysis of sRNA expression).

2.3.5.3 Removal of non-specific signal

To remove non-specifically bound radioactivity from the membranes, they were washed three times in wash solution [0.2 x sodium chloride/sodium citrate, 0.1% (w/v) sodium dodecyl sulphate (SDS)], at 37°C for 30 minutes with rotation.

2.3.5.4 Signal detection

On completion of the washes, each membrane was wrapped in cling film and placed in a radioactive cassette (Fujifilm). A phosporimaging screen (Fujifilm) was placed in each cassette on top of the membranes and was exposed for between 1 hour and 10 days. Screens were scanned on the Molecular imager FX pro plus (BIORAD) to visualize the radioactive signal.

2.3.5.5 Signal removal

Once the radioactive signal on the membrane had been detected, this signal was stripped off the membrane by washing twice in boiling stripping solution (0.1% SDS, 5 mM EDTA) for 20-30 minutes. Membranes were stored at -20°C.

Name	Sequence 5 to 3	Description
scaRNA2 32mer probe	GCCTCGTCTATCTGATCAAT	Targets a 31/32mer sRNA
scaRNA2 31mer probe	CAGAATCGCCTCGATAATCA	Targets a different 31mer closer to the 3' end
SNORD7 31mer probe	AGGTCAGGCTCTACTTCACTCATCATCGCAT	Targets most abundant sRNA
miR-191 probe	CAGCTGCTTTTGGGATTCCGTTG	Apoptosis loading control
miR-203 probe	CTAGTGGTCCTAAACATTTCAC	Alternative loading control
U6 snRNA probe	GCTAATCTTCTCTGTATCGTTCC	Standard loading control
ZMYM2 32mer probe	AGGTCAGGCTCTACTTCACTCATCATCGCAT	Targets most abundant sRNA
Met-tRNA half probe	AGCACGCTTCCGCTGCGCCACTCTGCT	Met probe from Fu paper
Cys-tRNAGCA half probe	AAATGCTCTACCACTGAGCTATACC	5' tRNA half
Gln-tRNATTG half probe	AGAGTGCTAACCATTACACCATGGG	5' tRNA half
Leu-tRNAAAG half probe	GCGCCTTAGACCGCTCGGCCACGCT	5' tRNA half
Ala-tRNAAGC half probe	ACGCGCTCTACCACTGAGCTACACC	5' tRNA half
hY5 3' end probe	AGCTAGTCAAGCGCGGTTGTGGGGG	Targets WT hY5 3' end
LNA10 probe	G*CTA*GTC*AAG	Last base is anchor and doesn't match Y5 sequence
LNA12 probe	A*CAG*CAA*GCT*AG	Detects majority of low-thoughput mutants
LNARo probe	A*AGC*GCG*GTT	Detects Ro60 site mutant only

Table 2.1: Probes used for northern blot analysis. *N denotes an LNA base.

2.4 GENERAL CLONING METHODS

All oligonucleotide primers were synthesized by Sigma-Aldrich and all reagents were from ThermoFisher Scientific unless otherwise stated.

2.5 DNA EXTRACTION

DNA extractions were carried out according to the DNeasy Blood & Tissue kit instructions (Qiagen). Human genomic DNA was extracted from HeLa cells by Dr. Wheeler or Dr. Nicolas.

2.6 DNA SEQUENCING

Plasmid and polymerase chain reaction (PCR) product sequences were determined by Sanger sequencing either by sending ready reactions to Genome Enterprise Ltd., or to Eurofins MWG. For ready reactions 1 μL of DNA was used in a PCR reaction to give a final volume of 10% (v/v) Half BigDye Terminator reagent, 10% (v/v) BigDye Terminator reagent, 1 x BigDye Sequencing Buffer and 1 μM of the appropriate sequencing primer. All sequencing reagents were obtained from Applied Biosystems. The PCR was run under the following conditions: 96°C for 2 minutes, followed by 25 cycles of 96°C for 30 seconds; 50°C for 15 seconds; and 60°C for 4 minutes. The entire reaction was sent to Genome Enterprise Ltd or an aliquot of the reaction was diluted to the requested concentration and sent to Eurofins MWG. Sequence analysis was done using BioEdit Sequence Alignment Software (Ibis Biosciences).

2.6.1 Amplification of sequences from genomic DNA

For amplification of sequences from genomic DNA, 100 ng of DNA was added to a PCR mix with a final concentration of 0.4 mM dNTPs, 0.5 mM of each of the primers, 1 x Phusion HF buffer and 0.4 units of Phusion High-fidelity DNA Polymerase (ThermoFisher Scientific). The PCR was run under the following conditions: 98°C for 3 minutes, followed by 30 cycles of (98°C for 1 minute; annealing temperature for 45 seconds; 72°C for 3 minutes); 72°C for 10 minutes; 4°C for ever.

2.6.2 Amplification of sequences from plasmid DNA

For amplification of sequences from plasmid DNA, 5-10 ng of DNA was added to a PCR mix with a final concentration of 0.4 mM dNTPs, 0.5 mM of each of the primers, 1 x Phusion HF buffer and 0.4 units of Phusion High-fidelity DNA Polymerase (ThermoFisher Scientific). Each reaction had a volume of 20 μL. The PCR

was run under the following conditions: 98°C for 3 minutes, followed by 30 cycles of (98°C for 1 minute, annealing temperature for 45 seconds, 72°C for 3 minutes); 72°C for 10 minutes; 4°C for ever.

2.6.3 Adenosine tailing

For standard cloning of PCR products (obtained using Phusion PCR) into the pGEM-T Easy plasmid, 1 μL of GoTaq DNA Polymerase (Promega) was added to each post-PCR reaction and these were then further incubated at 70°C for 15-30 minutes, prior to agarose gel electrophoresis.

2.6.4 Restriction digestion

Plasmid DNA was digested for 3-4 hours at 37°C and reactions were set up according to manufacturer's instructions. All plasmid DNA used for downstream cloning applications was heated to 65°C for 20 minutes to inactivate the restriction enzymes.

2.6.5 Agarose gel electrophoresis and DNA recovery

DNA samples were resolved using the appropriate (w/v) concentration of agarose dissolved in 0.5 x TBE. Either the 100 bp (New England Biolabs) or the 1 kb (New England Biolabs) double-stranded DNA ladder was run on each gel, depending upon the size of the product expected. Samples were mixed with 6 x loading dye (4M Urea, 40% (w/v) sucrose, 120 mM Tris-HCl pH7.5, 30 mM EDTA, 0.25% (w/v) orange G) before loading. Gels were visualized on a UV transilluminator and fragments resolving at the appropriate size were cut out using a sterile scalpel. Gel purification was carried out according to the Zymoclean Gel DNA Recovery kit instructions (Zymoresearch). Gel slices were dissolved at 55°C and DNA was eluted in 6-10 μL of ARG water.

2.6.6 Cloning into pGEM-T Easy

Ligations were performed according to manufacturer's recommendations (Promega). 50 ng of vector was routinely added to each ligation mix, with the appropriate amount of ligation fragment for a 1:1 and a 3:1 insert to vector ratio calculated depending on fragment size. The positive and negative ligation reactions recommended by the manufacturer were performed alongside all pGEM-T Easy transformations. Ligations were incubated at 4°C for 16 hours or overnight.

2.6.7 Transformation into DH5α cells

Unless otherwise stated, a 50 μL tube of home-made DH5α *E. coli* cells was thawed on ice for each transformation reaction. 1-2 μL of ligation mix was added to each tube of cells, mixed by swirling the end of the pipette tip in the liquid, and incubated on ice for 30 minutes. The samples were then placed at 42°C for 40 seconds to heat shock the cells and allow the DNA to enter. Samples were incubated on ice for a further two minutes before being diluted in 450 μL of room temperature luria broth (LB) (see Appendix 7) and placed in a 37°C shaker incubator for 1-1.5 hours. Following this step, cells were mixed by gently inverting and 100-200 μL was spread onto an LB selection plate containing selection antibiotic, 0.5 mM 5-bromo-4-chloro-3- indolyl--D-galactoside (X-GAL) and 80 μg/mL (Isopropyl -D-1-thiogalactopyranoside) IPTG and incubated at 37°C overnight.

2.6.8 'Home-made' super-competent DH5α cells

DH5α cells were pelleted from a 5 mL overnight culture in LB without antibiotics. The pellet was then streaked onto a bacterial M9 plate (Appendix 7) and then incubated at 37°C for 48-72 hours. A single colony was streaked onto an LB plate containing 10 mM $MgCl_2$ as well as an LB plate supplemented with 100 μg/ml Ampicillin. Plates were incubated at 37°C overnight. After confirming that there was no growth on the Ampicillin plate, a single colony from the LB/$MgCl_2$ plate was transferred to 5 mL pre-warmed TYM liquid culture (see Appendix 7) in a 50 mL centrifuge tube incubated at 37°C with shaking at 200 rpm overnight. 1 mL of this culture was then transferred to 100 mL of pre-warmed TYM broth in a 1 L flask and incubated at 37°C with shaking at 300 rpm for 1-3 hours, until the O.D.550 = 0.5. The cultures were transferred to 2 x 50 mL centrifuge tubes and spun at 1500 x g at 4°C for 12 minutes. The pellets were drained and re-suspended in 40 mL of cold TFB1 (see Appendix 7) solution. The cells were then incubated on ice for 10 minutes. Cells were centrifuged at 1500 x g for 8 minutes and gently re-suspended in 4 mL of TFB2 (see Appendix 7) before aliquoting and snap freezing on dry ice.

2.6.9 Determining transformation success

Cells containing plasmids with an insert appear white, while cells containing a plasmid without an insert appear blue on LB plates containing X-GAL/IPTG. White colonies were selected and transformation success determined by Sanger sequencing (see Section 2.6).

2.6.10 Plasmid DNA purification

Plasmid DNA was routinely purified from bacterial cells using the QIAprep Spin Miniprep kit (Qiagen) and re-suspended in ARG water. For plasmids used directly for transfection, plasmid DNA was purified using the HiSpeed Plasmid Midi kit (Qiagen) and re-suspended in Water for Injection (Life Technologies), unless otherwise stated. Plasmid DNA was resolved by agarose gel electrophoresis to check integrity, while concentration and quality were determined using the Nanodrop 8000.

2.6.11 De-phosphorylation

Plasmids were de-phosphorylated using the appropriate amount of shrimp alkaline phosphatase (SAP) according to the manufacturer's instructions (Roche). Plasmid DNA was incubated with the SAP enzyme for 60 minutes, before deactivation at 65°C for 20 minutes.

2.6.12 Primer annealing

To anneal two primers together they were mixed in equimolar amounts and added to a heating block at 95°C for 10 minutes. The hot block was turned off and allowed to cool down to room temperature, taking at least two hours. All annealed primer reactions were verified by agarose gel electrophoresis, to confirm that the majority of the DNA was resolving through the gel more slowly than either primer alone.

2.6.13 Cloning the human Y5 gene into an expression vector

2.6.13.1 pGEM-T Easy vector structure

The pGEM-T Easy vector system (Promega) is designed as an intermediate vector that can be amplified at high copy number in bacterial cells. In this project it was used as a destination vector, as the whole hY5 gene, including promoter and terminator sequences, was cloned into this plasmid. pGEM-T Easy is a small vector, designed for bacterial expression, so it should contain no elements that result in aberrant expression in eukaryotic cells.

The pGEM-T Easy vector is produced in a linear form, with a T overhang at the 3' end. For maximal ligation success, all fragments cloned into this vector must be 5' A tailed prior to ligation. This plasmid provides Ampicillin resistance to bacterial cells that contain it. The pGEM-T Easy vector also encodes the α-peptide sequence of the β-galactosidase gene which, when intact, results in blue colonies. When inserts are successfully cloned the reading frame of this peptide is disrupted, resulting in white colonies. This plasmid encodes M13 sequences on either side of the multiple

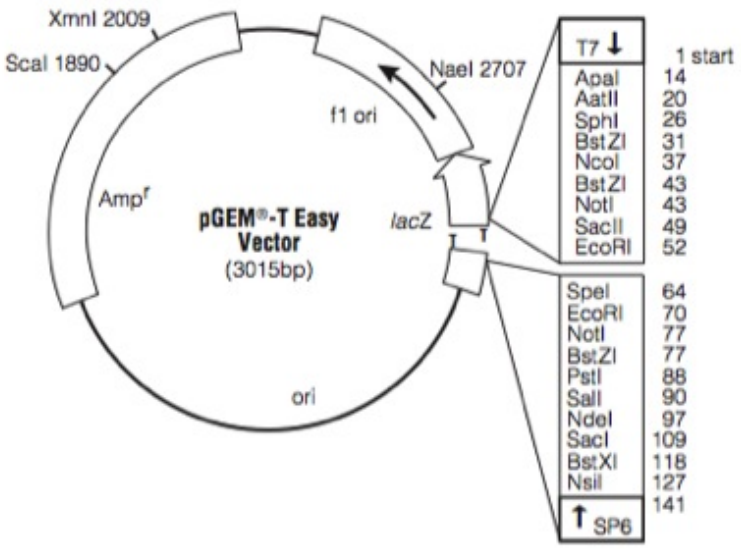


Figure 2.1: Plasmid map of the circularized version of the pGEM-T Easy plasmid. Showing the multiple cloning site and *in vitro* promoter sequences. Taken from Promega pGEMT-Easy Manual.

cloning site, so the sequence of the insert can be determined by sequencing with M13 forward and reverse primers.

2.6.13.2 Generating a human Y5 gene expression vector

The human Y5 gene was amplified by PCR from genomic DNA, derived from HeLa cells, using the standard high-fidelity genomic DNA PCR conditions and an annealing temperature of 64°C (See Section 2.6.1). The hY5 wild type forward and reverse primers were used in this reaction (Table 2.3). The PCR products were A-tailed and gel purified, before being ligated into the linear pGEM-T Easy plasmid. Ligated plasmids were transformed into DH5α cells and positive colonies were determined by restriction digestion using NdeI, for which there is a site within the plasmid and a site within the fragment, and gel electrophoresis. Out of the colonies that were sequenced using the M13 primers, one had the fragment inserted with no mismatches in the coding region or the sequence 500 bp upstream, so this was chosen as the expression construct. This plasmid was denoted pWT.

2.6.13.3 Expressing pWT in human, mouse and chicken cells

To determine the expression efficiency of the pWT plasmid in human cells, it was transfected into MCF7, NIH/3T3 and DF-1 cells. Briefly, cells were plated at the appropriate cell density in 3 wells of a 6 well plate and, 24-48 hours after plating cells were transfected with 2 μg of plasmid per well, using the Lipofectamine2000

transfection reagent. As controls for transfection, flasks were also treated with transfection mixes containing Lipofectamine2000 alone, or were grown in OptiMEM in the absence of transfection mixes. 24 hours after transfection, cells were harvested and total RNA was extracted. 1 μg of total RNA was loaded onto a PAGE gels and analyzed by northern blot. The full length wild type hY5 RNA was detected using a hY5-specific radiolabeled oligonucleotide primer.

2.6.14 Generating a human Y5 gene expression vector with restriction sites

Three primers were designed to insert restriction sites upstream and downstream of the ends of the hY5 RNA-encoding region. A restriction site for AgeI upstream of the start site, as well as a restriction site for AfeI downstream of the RNA-encoding region, was introduced using PCR. The primers were used to amplify a small fragment (the RNA-encoding region) and a large fragment (the promoter region) directly from the pWT plasmid. Each fragment was resolved by agarose gel electrophoresis and gel purified to remove excess primers and salts. The PCR mix was prepared without the primers and complementary regions within the 3' ends of the long and short fragments were annealed and extended for 5 cycles using the standard high-fidelity plasmid DNA PCR with an annealing temperature of 64°C (Figure 2.2). Primers were then added and the PCR continued for a further 25 cycles. The primers used to amplify the small fragment were wild type forward and overlap reverse and the primers used for the large fragment were overlap forward and alternative reverse (Table 2.3). Primers used to amplify the annealed and extended fragments were wild type forward and alternative reverse (see Figure 2.2).

The size of the small fragment was 100 bp and the size of the large fragment was 2900 bp, making it very difficult to determine the success of the overlap PCR by gel electrophoresis, as a size shift between the 2900 bp of the large fragment and the 3000 bp of the successful overlap PCR would be negligible. This PCR was monitored by the decrease in levels of the free small fragment. The resulting PCR product was cloned into the linear version of pGEM-T Easy. Overlap clones were tested by restriction digestion using NcoI and confirmed by sequencing using the M13 primers. A plasmid with no mismatches within the hY5 RNA-encoding region and with intact AfeI and AgeI restriction enzyme sites was selected. This plasmid was denoted pWT-2RE. A plasmid negative according to the IPTG/X-Gal colour selection was sequenced and confirmed to be a circularized version of the pGEM-T Easy plasmid, with no insert. This was used as a control for all hY5 transfection experiments and was denoted empty vector control (pEVC).

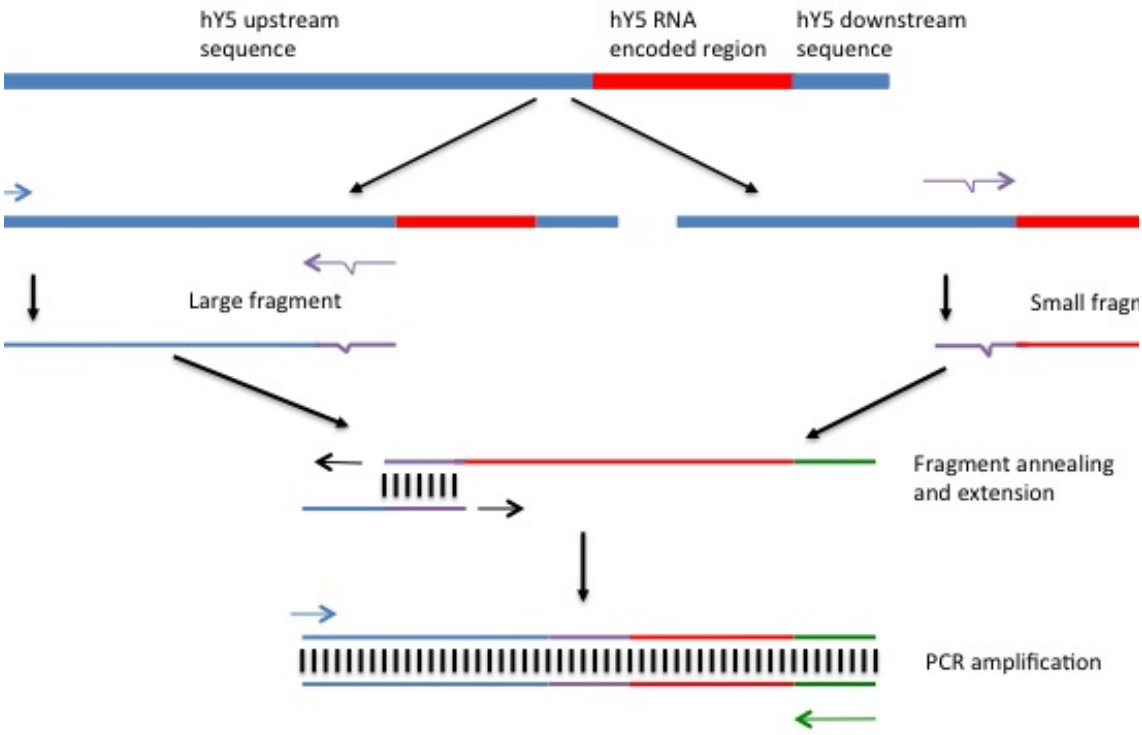


Figure 2.2: Schematic of the overlap PCR method.
Shows the method used to introduce restriction sites on either side of the RNA-encoding region of the hY5 gene.

2.6.15 Creating a stuffer plasmid for mutant plasmid generation

Restriction digestion of the pWT-2RE plasmid with AfeI and AgeI gave a wild type hY5 RNA-encoding fragment, along with a linearized plasmid (pWT-2RE-promoter). As this fragment was difficult to visualize on a gel and remove, there was a potential for it to re-ligate into the plasmid. To prevent this and to make it easier to determine the success of the restriction digest, a stuffer plasmid was created as an intermediate step in the cloning process. In this stuffer plasmid, the hY5 RNA-encoding region of the pWT-RE plasmid was removed by restriction digest and then replaced with an unrelated 1000 bp fragment.

This stuffer fragment was amplified from a plasmid containing an unrelated DNA sequence using the standard Phusion high-fidelity PCR conditions, an annealing temperature of 69°C, and the primers stuffer forward and stuffer reverse (Table 2.3). These primers were designed to introduce an Age1 site or an Afe1 site at either end of the stuffer fragment. After PCR amplification, the fragment was resolved by gel electrophoresis and gel purified to removed excess primers and salts. The fragment was then digested with restriction enzymes AfeI and AgeI to produce ends complementary to the linearized pWT-2RE-promoter plasmid.

The linear pWT-2RE-promoter plasmid was de-phosphorylated and gel extracted, to remove as much of the excised 100bp hY5 fragment as possible. The stuffer fragment was then ligated into the linear pWT-2RE-promoter plasmid according to the recommendations for pGEM-T Easy plasmids and transformed into DH5α cells. Positive colonies were selected and sequenced and one plasmid was selected for use in downstream mutant plasmid generation (pStuffer).

Name	Sequence 5' to 3'
hY5 wild type forward	AATACTAGTGAAGATCCATGGAGGTACATC
hY5 wild type reverse	GTAAACGTTGTCTACTACTGTTATTAGTGC
hY5 overlap forward	AATAACAAGAGACTCACCGGTAACACAGTT
hY5 overlap reverse	AACTGTGTTACCGGTGAGTCTCTTGTTATT
hY5 alternative reverse	TTATTAGCGCTAAAACAGCAAGCTAGTCAA
Stuffer forward	GAAGAGTGTCTCAAGTTCTGCGCCTCC
Stuffer reverse	CGTCCTGGCTGTCCTGGATGATCTGT
hY5 sequencing primer	GATGAAACAAAGCCGACA
M13 forward	CGCCAGGGTTTTCCCAGTCACGAC
M13 reverse	TCACACAGGAAACAGCTATGAC

Table 2.2: Primers used for PCR amplification

2.6.16 Generating hY5 RNA mutants

The pStuffer plasmid was digested with both AfeI and AgeI, gel extracted and then de-phosphorylated. Mutant primers were phosphorylated where necessary, annealed together and ligated into the digested pStuffer plasmid (See Appendix 7 for mutant primer details).

2.6.16.1 Mutant primer phosphorylation and annealing

350 pmol of each primer in a pair was diluted in 1 x PNK forward reaction buffer, before heating to 70°C for 10 minutes. 2 μL of 10 mM ATP and 10 units of T4 PNK were added and incubated at 37°C for 30 minutes to allow the ends to be phosphorylated. The PNK was then inactivated by heating to 65°C for 20 minutes. For those primers that were phosphorylated in house each of the two primer mixes were combined in equal volumes. For primers that arrived phosphorylated they were mixed together in 1 x PNK forward reaction buffer. Mixes were annealed as in Section 2.6.12

2.6.16.2 Ligation of annealed mutant primers into pStuffer

Double-stranded DNAs were ligated into the pStuffer plasmid at the ratios recommended for pGEM-T Easy. Ligation mixes were incubated at 16°C overnight, as recommended for blunt ended ligations. Ligation mixes were transformed into DH5α cells and plated on LB plates with Ampicillin selection. Positive colonies for each mutant were sent for Sanger sequencing. For each mutant, one colony was selected which had both the AfeI and AgeI the restriction sites intact, and no mismatches within the coding region, or 500 bp upstream.

2.6.17 Expression of the mutant plasmids in human MCF7 and mouse NIH/3T3 cells

For these experiments the treatments were scaled up to T75cm^2 flasks. Briefly, MCF7 cells were plated 48 hours, and NIH/3T3 cells were plated 24 hours, before transfection. Cells were transfected with 20 μg of plasmid DNA per flask, or were treated with either transfection mixes alone, or were untreated. Transfections were conducted using Lipofectamine2000, unless otherwise stated. Cells were transfected with mixes containing an unrelated plasmid, pGL3 or the empty vector control pEVC were included where appropriate. Cells were treated with Poly(I:C), Hygromycin or Staurosporine as detailed in Section 2.1.5

2.6.17.1 Removal of the AgeI restriction site and re-introduction of a longer downstream sequence

On transfecting the pWT-2RE plasmid into a number of different cell types, it was concluded that the resulting RNA transcripts were on the border of detection limits by northern blot analysis. Further investigation showed that the lower levels of transcription were a result of the introduction of the restriction sites. Due to time constraints, it was not possible to remove both restriction sites from each of the mutant plasmids, but a further amplification of each of the mutant hY5 RNA genes with a unique reverse primer removed the AgeI site and extended the 3' UTR sequence (see Table 2.3 for details of each reverse primer). Transfections in mouse cells indicated that this dramatically increased the transcription levels from each of the mutant plasmids. As the levels of the mutant RNAs in the transfected mouse cells were still lower than the levels of the hY5 RNA produced from the pWT plasmid, a wild type plasmid with 1 restriction site (pWT-1RE) plasmid was also created where the 3' restriction site was removed and a longer 3' UTR was re-instated. This plasmid was transfected alongside each mutant plasmid and was used for direct comparison in all experiments. These mutants were transfected using the Fugene6 transfection reagent, which is less toxic to cells. The amount of plasmid transfected ranged from 2.5-10 μg, as Fugene6 had a better transfection efficiency that Lipofectamine2000.

Name	Sequence 5' to 3'
ΔS mutants reverse	GTAAACGTTGTCTACTACTGTTATTAGTCAAAACAGCAAGCTAGTCAAGGGGGG
SubS mutants reverse	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGTCAAGATGGG
Δ ACC mutant	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGTCAAGCGCTG
Δ 3' end mutant reverse	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAAGCTAGTCAAGCGCGGTTG
ΔRo60 mutant	GTAAACGTTGTCTACTACTGTTAATTAGTGCAAAACAGCAAGCTCAAGCGCGGTTG

Table 2.3: Unique primers used for mutant correction PCR

2.6.18 High-throughput mutagenesis of the human Y5 RNA

For a more high-throughput method of hY5 mutagenesis, a different method of mutant generation was used. Long primers (IDT) were synthesised that contained equal proportions of each of the four bases at each of five positions per primer pool (Table 2.4). This equates to 1024 different sequences per pool. Three pools were synthesized and the position of the 5 nucleotides mutated was shifted sequentially along the hY5 sequence.

Initially, a pool of PCR products was to be generated from each of the three mutant primer pools and these were going to be directly transfected into NIH/3T3 cells and used to produce mutant RNAs. Due to a high level of primer-dimer adversely

affecting the success of the PCR, PCRs were initially carried out in the presence of either the forward or the reverse primer for 10 cycles, before one forward and one reverse reaction were combined and amplified for a further 25 cycles. The PCRs were run as previously described for plasmid DNA PCRs, except for the lower number of cycles. The annealing temperature was raised to 69°C, to overcome the primer-dimer problem. PCR products were purified using the QIAquick PCR purification kit (Qiagen) and transfected using the same method as for the single-mutant plasmids, except that half the amount of PCR product was added, as the PCR product was half the size of the pWT plasmid. Initial tests suggested that transfection of PCR products derived from a pGFP plasmid did not produce any detectable signal. Further northern blot experiments to detect wild type hY5 RNA transcribed from PCR products derived from pWT were also unsuccessful. For this reason, an alternative strategy was designed, where the PCR product pools were cloned into pGEM-T Easy and transfected into NIH/3T3 cells as plasmids.

2.6.18.1 PCR amplification, ligation and transformation

Using the wild type forward primer and either the 1, 2, or 3 reverse primer pools the hY5 gene was amplified and the degenerate sequences were incorporated, creating a pool of PCR products from each primer pool. PCR products were created for 3 separate replicates of each of primer pools 1, 2, and 3. PCR products were prepared as in Section 2.6.18. PCR products were A-tailed, gel extracted and ligated into the pGEM-T easy vector as already detailed. Transformations were carried out using super-competent 'home made' DH5α cells, or XL2 Blue *E. coli* cells for pool 3, and the ligations and transformations were optimized for each PCR reaction to ensure the best efficiency possible. *E. coli* cells were grown under blue/white colony selection and it was determined that a minimum of 4850 positive colonies were required for each replicate to give a 99% chance of selecting each of the 1024 possible sequences at least once. For each replicate of each pool, over 4850 positive colonies were harvested, for most pools the number was approximately double this.

2.6.18.2 Colony selection, harvesting and plasmid purification

For each pool, 10 positive colonies were selected from one of the replicates and the presence of an insert was confirmed using colony PCR. To harvest each plate, 3 mL of liquid LB was added and the cells were scraped using a disposable scraper until no colonies could be seen on the plate. The bacterial cell suspension was then transferred to a 50 mL centrifuge tube. For each replicate 6-8 plates were combined into one centrifuge tube. Each centrifuge tube was processed as a separate midi-prep. Four midi-prep samples were eluted in 1000 μL of water for injection (Invitrogen)

to maximize plasmid recovery.

2.6.18.3 Ethanol precipitation to concentrate DNA

To concentrate the plasmid DNA, each midi-prep was divided between two 1.5 mL centrifuge tubes and ethanol precipitated. Briefly, 50 μL of 3 M NaAc was added (pH 4.8), then 1250 μL of ice cold 100% ethanol was added and samples were placed in the -20°C freezer overnight. Samples were spun at 13,000 x rpm for 15 minutes at 4°C and the supernatant was removed. Samples were gently washed in 1 mL of 80% ethanol and air dried. For each replicate, all pellets were recombined and re-suspended in 100 μL of water for injection. Plasmid DNA integrity was determined by gel electrophoresis, while concentration and purity was determined using the Nanodrop 8000.

2.6.18.4 Transfection conditions

NIH/3T3 cells were plated at the appropriate number into T75cm^2 flasks. Transfection mixes were added to the cells 24 hours after plating. Each replicate plasmid preparation was transfected into a separate T75cm^2 flask, giving 3 biological replicates for each of the 3 pools. Transfection mixes were prepared as for the hY5 mutants, except that 5 μg was used instead of 10 μg (Section 2.6.16.2). 24 hours after transfection cells were treated with 1 μM STS. Cells were harvested 4 hours later, and total RNA was extracted.

2.6.18.5 cDNA library preparation for hY5 sRNAs

The RNA was extracted as detailed previously. 2 μg of total RNA was used in each of the small RNA libraries. Samples were processed according to the 'HD' small RNA library protocol written by Dr. Karim Sorefan, with a few minor adjustments. Firstly the adapters used were the FAM dye-labeled 3' 'HD' adapter (Dalmay Lab) and the RA5 5' RNA adapter (Illumina). The RP1 RNA PCR primer (Illumina) was replaced with a modified primer containing the RP1 sequence (Sigma), but with the extra 5 bases of the adapter along with three Cs at the 3' end. After the first PCR polyacrylamide gel, a 32 nt and a 24 nt band were separately cut from the gel and processed. cDNA libraries were prepared for cells treated with each replicate derived from plasmid pools 1, 2, and 3.

All reagents were from the Truseq small RNA Preparation kit unless otherwise stated. 2 μg of total RNA was used as the input for the kit. Each reaction was prepared using half of the volumes recommended in the kit instructions. The 2 μg of RNA was mixed with 0.5 μL of the FAM dye-labeled 3' 'HD' adapter (10 μM stock) and the volume was made up to a total of 2 μL with RNase free water. Samples

were heated at 70°C for 2 minutes and immediately placed on ice. In a new 0.2 ml tube, 3 μL of ligation mastermix was added. The mastermix was made up of 1.65 μL of T4 RNA Ligase I (New England Bioseciences), 3.3 μL of Ligase I buffer, 1.65 μL of RNase inhibitor and 1.65 μL of 10 mM ATP. Denatured RNA mixes were rapidly added to these fresh tubes and reactions were incubated at 28°C for 2 hours. Samples were then run using standard denaturing PAGE for RNA, although a single-stranded RNA ladder was included for size comparison (Ambion). The gels were stained with Sybrgold (Invitrogen) and scanned on the FX scanner. Bands of the appropriate size were cut from each lane, taking care to leave the majority of the adapter only band behind. A stock of NEB buffer 2 was diluted to 1 x and used to elute the RNA from the gel. To do this, a 0.2 mL centrifuge tube was placed inside a 2 mL centrifuge tube. two holes were made in the 0.2 mL centrifuge tube using a 25G 5/8 inch syringe needle. the gel fragments were placed inside the 0.2 mL centrifuge tube. 300 μL of buffer was added to each lower tube. samples were spun at maximum speed in a micro-centrifuge for 2 minutes. Samples were then incubated at 4°C, with shaking, overnight. Each gel mix was transferred to a Spin-X 45 μm filter column. Columns were centrifuged for 30 seconds at 600 x g. The eluate was transferred to a fresh 2 mL centrifuge tube. 30 μL of NaOAc and 975 μL of ice-cold 100% ethanol were added to each tube. Samples were placed in the -80°C for 30 minutes and the centrifuged at 20,000 x g for 20 minutes at 4°C. Samples were washed in 1 mL of ice-cold 80% ethanol and the dregs of the ethanol removed. Pellets were then air-dried for 3-4 minutes and re-suspended in 3.5 μL of RNase-free water.

The samples were transferred to a new 0.2 mL centrifuge tube and 0.5 μL of Illumina 5' adapter was added to each tube. Samples were incubated at 70°C for 2 minutes and immediately placed on ice. The ligation mastermix was prepared as before and samples were transferred to the tubes containing 3 μL of the mastermix, before being incubated at 28°C for 2 hours. On completion of the incubation, samples were moved to a fresh 0.2 mL centrifuge tube containing 1 μL of RTP RNA RT primer (Illumina) and incubated at 70°C for 2 minutes, before immediately placing on ice. 5.5 μL of RT mastermix was added to each tube. This was prepared by adding 1.65 μL of 12.5 mM dNTPs, 3.3 μL of 0.1 M DTT, 3.3 μL of RNase inhibitor and 3.3 μL of Superscript II. Samples were then incubated at 50°C for 1 hour.

Samples were PCR amplified using standard high fidelity PCR conditions, but with 0.8 μL of each primer and 2 μL of cDNA in a 20 μL reaction volume. The alternative RP1 forward primer (Table 2.5) was included in the mastermix and the Illumina index reverse primer was added separately to each tube. For each replicate, 3 PCR reactions were run for the appropriate number of cycles. On completion of the

first PCR, samples were run on an 8% non-denaturing polyacrylamide gel and both the 32 nt and 24 nt bands were cut out and eluted as detailed earlier. Samples from all 3 PCR products for each replicate were pooled and eluted together. Following elution, samples were subjected to a second gel extraction to remove contaminating miRNAs from the 32 nt band and to be sure that all adapter-adapter had been removed. Samples were sized using the 20bp dsDNA marker (Jena Bioscience). The full volume of each sample was sent to BaseClear for sequencing using the alternative sRNA sequencing primer (Table 2.5).

2.6.19 cDNA library preparation of new Poly(I:C)/control sRNA samples

New control and Poly(I:C) sRNA libraries were prepared as for the high-throughput hY5 libraries, except that both 'HD' adapters were used. Also the Illumina RP1 PCR primer and sequencing primers were used during library preparation. Three replicates were prepared for each condition and both SW 1353 and MCF7 samples were treated.

2.6.19.1 cDNA library preparation for full length hY5 RNAs

5 μg of total RNA was DNase treated using the Turbo DNase free kit (Ambion). 500 ng of this RNA was used in an RT-PCR performed according to Superscript II RT kit instructions (Invitrogen) for using a gene-specific primer, FL library RT (See table 2.6 for primer full length library primer details) . After first-strand cDNA synthesis a standard high-fidelity PCR was run for 15 cycles, using the FL round 1 forward and reverse primers and an annealing temperature of 63°C. This PCR product was then separated on a 6% PAGE gel, eluted overnight and ethanol precipitated as previously described. The samples were then diluted 1/100 and used in a second PCR reaction using the Illumina RP1 forward PCR primer and the appropriate Illumina index reverse primer. Each sample was PCR amplified for 10 cycles at an annealing temperature of 60°C. On completion of the PCR, samples were separated on a 6% PAGE gel and a band of the correct size excised and purified. cDNA libraries were prepared for cells treated with each of the three replicate samples for the three different plasmid preparations 1, 2, and 3.

2.6.19.2 Plasmid bias hY5 library preparation

5 ng of plasmid DNA was used in standard plasmid PCR amplification reactions using the PB library round 1 forward and reverse primers (See table 2.7 for primer plasmid bias library primer details). Samples for PCR were run for 15 cycles with

an annealing temperature of 57.5°C. Each PCR product was then separated on a 6% PAGE gel, eluted overnight and ethanol precipitated as previously described. The samples were then diluted 1/500 and used in a second PCR reaction using the Illumina RP1 forward primer and the appropriate Illumina index reverse primer. Each sample was PCR amplified for 10 cycles at an annealing temperature of 60°C. On completion of the PCR, samples were separated on a 6% PAGE gel and a band of the correct size excised and purified. DNA libraries were prepared for each replicate derived from the three different plasmid preparations A, B, and C.

Name	Sequence 5' to 3'
Pool 1	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGTCAAGCGCGGTNNNNNGGGG
	AGACAATGTTAAATCAACTTAACAATAACCC GAGACAATGTTAAATCAACTTAACAATAACCC
Pool 2	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACGCAAGCTAGTCAAGCNNNNNTGTGGGGGG
	AGACAATGTTAAATCAACTTAACAATAACCCGAGACAATGTTAAATCAACTTAACAATAACCC
Pool 3	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGTNNNNNGCGGTTGTGGGGGG
	AGACAATGTTAAATCAACTTAACAATAACCCGAGACAATGTTAAATCAACTTAACAATAACCC

Table 2.4: Reverse primers used for hY5 mutant pool generation. Ns denote degenerate bases

Name	Sequence 5' to 3'
Alternative RP1 forward	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATCCCC
Alternative sRNA sequencing primer	CTACACGTTCAGAGTTCTACAGTCCGACGATCCCC

Table 2.5: Custom primers used for hY5 mutant sRNA library preparation and sequencing

Name	Sequence 5' to 3'
FL library RT	AAACAGCAAGCTAGT
FL library round 1 F	GTTCAGAGTTCTACAGTCCGACGATCAGTTGGTCCGAGTGTTGTGGGT
FL library round 1 R	CCTTGGCACCCGAGAATTCCAAACAGCAAGCTAGTCAAGCG
FL/bias sequencing primer	GTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAAC

Table 2.6: Custom primers used for hY5 full length library preparation and sequencing

Name	Sequence 5' to 3'
PB library round 1 F	GTTCAGAGTTCTACAGTCCGACGATCTAACACAGTTGGTCCGAGTGTT
PB library round 1 R	CCTTGGCACCCGAGAATTCCATTAGTGCAAAACAGCAAGCT

Table 2.7: Custom primers used for hY5 plasmid bias library preparation

3 GENERATION OF hY5-DERIVED RNAS IS SECONDARY STRUCTURE DEPENDENT BUT SEQUENCE INDEPENDENT

3.1 INTRODUCTION

Human Y RNAs are between 84 and 112 nucleotides in length and are denoted hY1, hY3, hY4 and hY5 (Wolin and Steitz, 1983). The RNA originally named hY2 was later found to be a truncated version of hY1 (Wolin and Steitz, 1983), and so is no longer considered to be a separate Y RNA. This project focuses on the Y5 RNA, which appears to have arisen at some point in the tetrapod lineage and has a notably more divergent sequence in primates than any of the other Y RNAs (Perreault et al., 2007).

3.1.1 Human Y5 RNA

hY5 RNA is the smallest hY RNA and it arose the most recently during evolution (Kato et al., 1982; Wolin and Steitz, 1983; Perreault et al., 2007). With a smaller internal loop than the other hY RNAs (Kato et al., 1982; Wolin and Steitz, 1983) one might suppose that it has less functionality, but this RNA has taken on some surprising characteristics. For example, hY5 RNA has a 3' tail at least 5 nucleotides longer than the other hY RNAs (Wolin and Steitz, 1983). hY5 is thought to bind to the La protein more tightly than hY1-4 (Wolin and Steitz, 1983; Boire and Craft, 1990), possibly as a result of this extended tail. Alternatively, La could be interacting elsewhere on the molecule, as this protein has recently been found to associate with non uracil-containing sequences within RNAs (Fan et al., 1998), as well as recognising structures at the 5' ends of certain RNA molecules (Martino et al., 2012). Both Ro60 and La are known to undergo modifications under certain conditions, so it may be that the cytoplasmic hY1-4 and the mainly nuclear hY5 bind to a different fraction of these proteins (Gendron et al., 2001).

Another interesting characteristic of hY5 is that it uniquely interacts with a number of proteins that do not bind to the other hY RNAs (Hogg and Collins, 2007). There are few proteins that bind to only one of the other hY RNAs uniquely, more often binding to two or more (Fabini et al., 2001; Fouraux et al., 2002). hY5 is the only hY RNA that has been found to be directly involved in Ro60 QC functions through its recruitment of the ribosomal protein L5 (Hogg and Collins, 2007). Perhaps this is a specialized function that this RNA has evolved to perform. Evidence supporting a specialized and possibly vital function of this hY RNA is that it is found at the highest concentration of all of the Y RNAs within certain human cell types (Gendron et al., 2001), despite the fact that hY3 is the most conserved (Farris et al., 1995). All hY RNAs can function in chromosomal replication (Christov et al., 2006), but perhaps hY1 and hY2 are replication-focused and have limited alternative functions. One final unique characteristic of the hY5 RNA is that it acts as an autoantigen and can be targeted by antibodies in patients with SS or SLE

(Boulanger et al., 1995). Ro60 and La are also autoantigens for these disorders, but none of the other hY RNAs contain sequences that are recognised by autoantibodies (Boulanger et al., 1995). All of these lines of evidence together suggest that the hY5 RNA has a specialized function that has evolved separately from the other hY RNAs.

3.1.2 Preliminary work

As discussed in Chapter 1, Dr. Nicolas (from the Dalmay group) discovered that Y RNAs are cleaved specifically after treatment with Poly(I:C), a mimic of viral infection. This chapter will focus on the cleavage of hY5, as the cleavage of the other Y RNAs has been discussed in work by Dr. Hall (Hall, 2013). Prior to the start of this PhD project, Dr. Nicolas further characterised the conditions of hY5 cleavage.

Following on from the Poly(I:C) and control NGS libraries discussed in Chapter 1, Dr. Nicolas found that Poly(I:C)-dependent sRNAs were produced from both the 5' and 3' ends of the hY5 RNA: these sRNAs will be referred to as Y RNA-derived RNAs or ydRNAs. Northern blots of total RNA prepared in the same manner as for the NGS libraries led him to determine that the production of Y5 3' end ydRNAs was time-dependent and Poly(I:C) concentration-dependent, with the majority of the Y5 RNA cleaved after 24 hours (Figure 3.1). A 32 nucleotide ydRNA was detectable after 1 hour, appearing much earlier than the 24 nucleotide ydRNA, which was only detectable after 24 hours. Both ydRNAs increased in abundance at each time point over a period of 72 hours. We cannot say whether these sRNAs are derived from different Y5 molecules or whether the 24 nucleotide ydRNA is a processed version of the 32 nucleotide ydRNA, although the latter seems more likely. Northern blots of the same membranes using a different probe that detects cleavage products derived from the 5' end of hY5 identified another 32 nucleotide ydRNA, but a laddering was seen on the northern blot image, suggesting that this fragment might be a degradation product rather than a *bona fide* ydRNA (Figure 3.2).

Dr. Nicolas discovered that, after Poly(I:C) treatment, the hY5 3' ydRNAs entered apoptotic vesicles (data not shown). This is of interest because both Ro60 and La have been found to enter apoptotic bodies in keratinocytes (Casciola-Rosen et al., 1994a). Rutjes *et. al.* showed that hY RNAs are fragmented in apoptotic cells and that the smaller fragments seem to have lost the La binding site (Rutjes et al., 1999b). After Poly(I:C) treatment, Dr. Nicolas noted that MCF7 cells clearly looked like they were undergoing apoptosis, as they could be seen to be rounding up and undergoing chromatin condensation, although this was not confirmed experimentally.

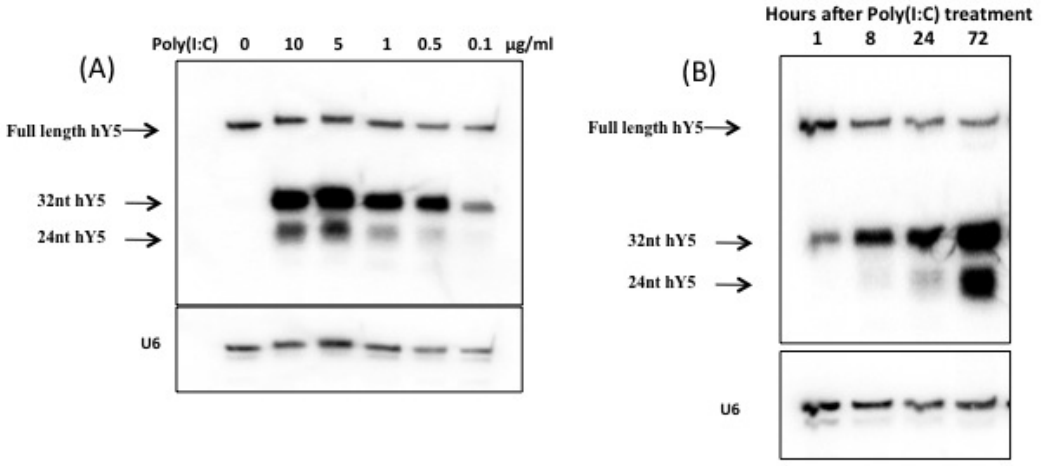


Figure 3.1: Poly(I:C) concentration- and treatment time- dependent cleavage of the hY5 RNA.
(A) Dr. Nicolas examined the effect of treating cells with different concentrations of Poly(I:C) on the production of Y5-derived 3' sRNAs. (B) Dr. Nicolas investigated the cleavage pattern of human Y5 RNA into Y5-derived 3' sRNAs after Poly(I:C) treatment over a period of 72 hours.

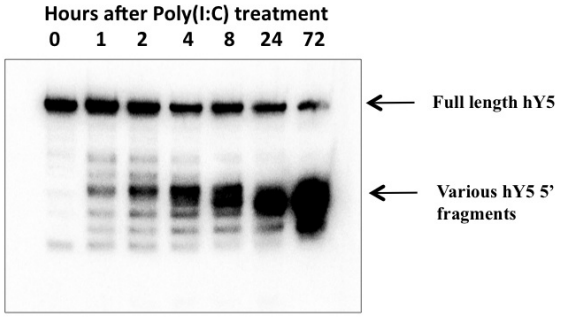


Figure 3.2: Poly(I:C) time-dependent cleavage of the hY5 RNA.
Dr. Nicolas examined the effect of Poly(I:C) treatment over a period of 72 hours on hY5 5' end ydRNA production.

When human cells detect foreign nucleic acids they mount an immune response. One of the proteins involved in this response is Protein Kinase RNA-activated (PKR) (Laurent et al., 1984). PKR detects and associates with dsDNA (Galabru and Hovanessian, 1985). Using a chemical inhibitor of PKR, Dr. Nicolas found that the cleavage at the Y5 3' end was PKR-independent (data not shown). Recognition of double-stranded RNA within the cell is complex and proteins other than PKR can be involved, so it is possible that the Y RNA cleavage products detected were dependent upon apoptosis rather than an immune response, or it could be an immune response activated by another cellular protein. In 2009 the Wolin group showed that, after infection with Murine Leukemia Virus, mouse Y RNAs were taken up by newly made virus particles independently of the Ro60 and La proteins (Garcia et al., 2009). It is therefore possible that this Y RNA cleavage, which we know occurs when we mimic viral infection, is to prevent the viral hijacking of cellular RNAs. Y RNAs play a key role in chromosomal replication in mammalian cells (Christov et al., 2006), but this virus has no need for DNA replication, so it seems likely that these RNAs are required by the virus for a different function.

The aim of this chapter was to mutate the sequence of the hY5 gene to produce mutant versions of the hY5 RNA. These mutants would be used to determine whether it is the sequence of the RNA that is recognised by proteins involved in cleavage or whether it is features of the secondary structure that are important.

3.2 RESULTS

3.2.1 Confirming apoptosis induction by Poly(I:C) treatment

The DLD-1 cell line was utilized for assaying apoptosis. DLD-1 cells were grown under various conditions, including Poly(I:C) treatment, and harvested. At the point of harvesting, samples were split into two and the first replicate was processed for flow cytometry, while total RNA was extracted from the second replicate and used for northern blotting. For flow cytometry experiments, cells were stained with Annexin V and PI before analysis on the flow cytometer. For each sample, the cells of interest were manually gated to exclude clumps of cells and cell debris (Figure 3.3). Flow cytometry clearly showed that over 50% the population of Poly(I:C) treated cells were undergoing apoptosis at the 8 hour time-point, while around 10% of cells treated with Hygromycin for 8 hours were undergoing apoptosis (Figure 3.4).

For northern blot analysis, the membrane was probed for the hY5 3' end-specific ydRNA. This ydRNA was detectable at low levels in the Hygromycin sample and higher levels in the Poly(I:C)-treated sample (Figure 3.5). The level of this ydRNA in the PBS-treated sample was similar to that in the control samples, indicating

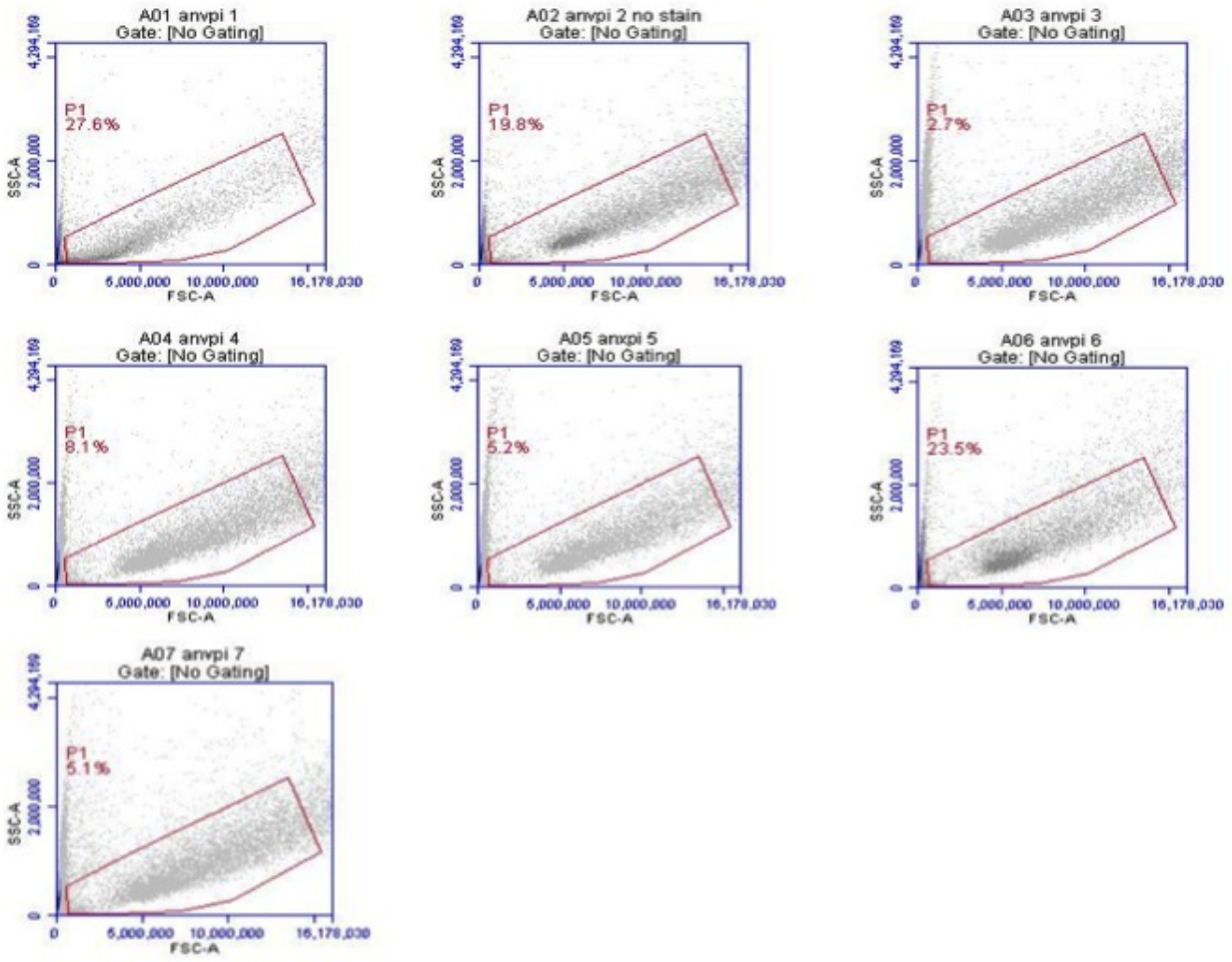


Figure 3.3: Manually gated images of DLD-1 cells analysed by flow cytometry.
Those cells within the gates shown were included in the analysis. FSC - forward scatter, SSC - side scatter.

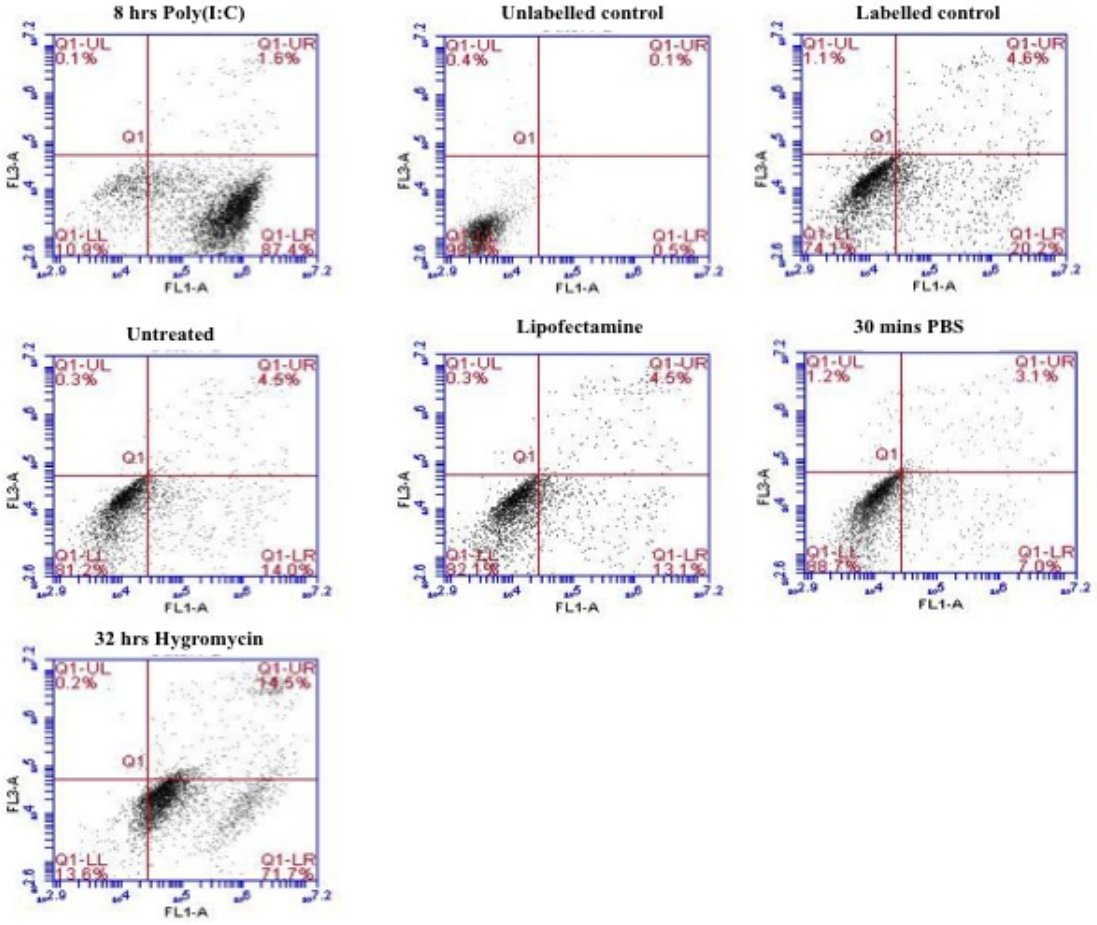


Figure 3.4: Flow cytometry analysis of apotosis in DLD-1 cells.
PI vs Annexin V fluorescence intensity plots for cells grown under various conditions. Cells were transfected with Poly(I:C), Lipofectamine2000 alone, or were treated with Hygromycin, PBS or left untreated (FL-1 - separation by Annexin V staining, FL-2 - separation by PI staining).

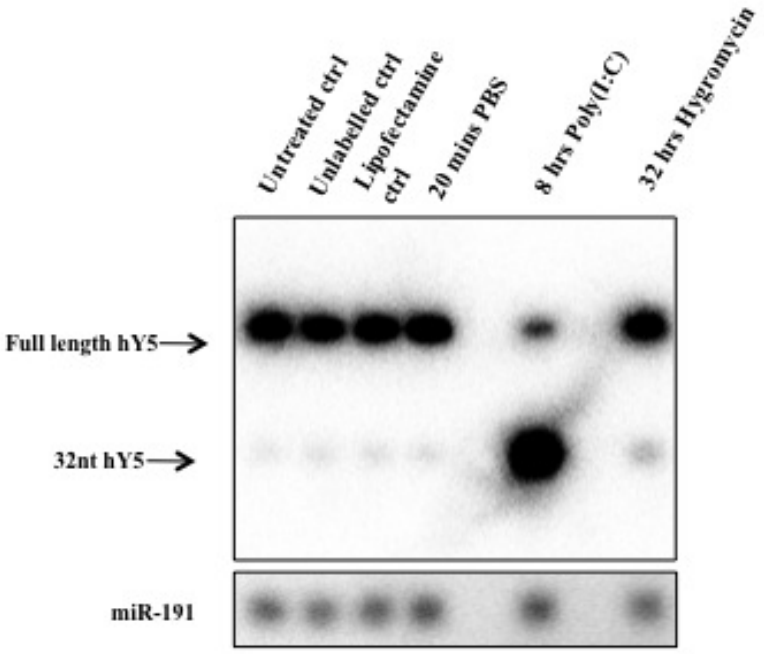


Figure 3.5: Detection of hY5 cleavage in DLD-1 cells grown under different conditions.
DLD-1 cells were plated in 6 well plates. After 24 hours they were transiently transfected with Poly(I:C) transfection mixes, Lipofectamine2000 alone, or were treated with PBS, Hygromycin or left untreated. After the required incubation, cells were harvested and either used for flow cytometry or for northern blotting. 3 μg of total RNA was separated by PAGE. The membrane was probed with a hY5 3' end-specific radiolabelled probe.

that PBS does not induce cleavage.

3.2.2 Cloning of the wild type hY5 RNA gene into the pGEM-T Easy plasmid

The hY5 RNA gene, including the RNA Polymerase III promoter and terminator sequences, was PCR amplified from genomic DNA and then cloned into the pGEM-T Easy plasmid (Figure /refy5 gene).

Maraia *et al.* were the first to fully characterize the hY5 RNA gene, and the sequence features important for trancription of this gene (Maraia et al., 1996). This gene contains a TATA box at positions -36 to -26 and a proximal sequence element at positions -64 to -47. Possible octamer sequences are also present at positions -242 to -234 and -216 to -209. In this paper Maraia noted that, in mouse cells, when a plasmid containing the hY5 gene was co-transfected with the same amount

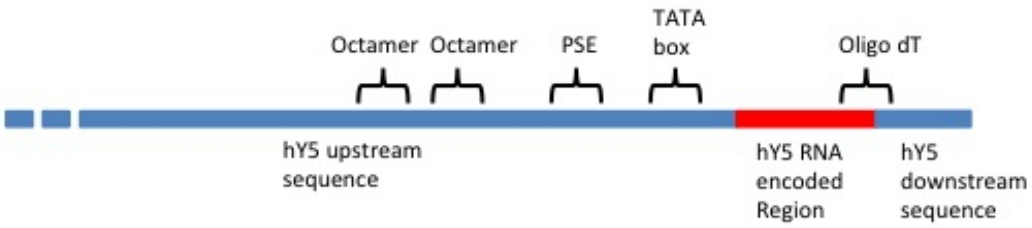


Figure 3.6: Structure of the hY5 gene.
2900 bp upstream of the hY5 RNA encoding region was cloned into the pGEM-T Easy plasmid. This includes two octamer sequence elements (-242 to -234 and -216 to -204), a proximal sequence element (-64 to -47), and a TATA box (-36 to -26). The transcription termination signal is an oligo dT stretch at the end of the RNA encoded region. 30 bp downstream of the oligo dT element was included in the sequence cloned.

of a plasmid containing the hY4 gene, despite the fact that hY5 is the most highly expressed Y RNA in numerous human cells line, hY4 was expressed at notably higher levels (Maraia et al., 1996). To take into account the possibility that important sequence elements might have been missed during the characterization of this gene, a 2.9 kb region upstream of the transcription start site was cloned into pGEM-T Easy, along with the RNA-encoding region. A 30 bp region downstream of the termination signal was also included: no important elements had been annotated beyond this point. This plasmid will be referred to as the pWT plasmid.

Y5 RNA is expressed in humans, but not in chickens or in mice. In order to confirm that the hY5 RNA can be produced from this plasmid, it was transfected into the human MCF7 cell line, as well as the mouse NIH/3T3 and the chicken DF-1 cell lines. These initial transfection experiments were conducted to determine whether transcription of hY5 from the plasmid would be sufficient to give a detectable difference between the levels of full length hY5 in untransfected versus pWT-transfected cells i.e. whether the exogenous hY5 could be distinguished from endogenous hY5. In transfecting the plasmid into mouse and chicken cells, the aim was to determine whether any detectable full length hY5 and any cleavage products could be detected in the transfected cells, which would indicate that the cleavage machinery is conserved between these organisms and humans.

A secondary aim was to determine the specificity of a northern probe designed to be hY5 3' end-specific, which should not hybridize to any of the other human Y RNAs or any of the mouse/chicken Y RNAs, which are not homologs of hY5. This probe should detect the full length hY5 RNA, as well as any fragments which contain the 3' end derived sequence that Dr. Nicolas identified.

Initial experiments involved transfection with 2 μg of plasmid DNA per well of a 6-

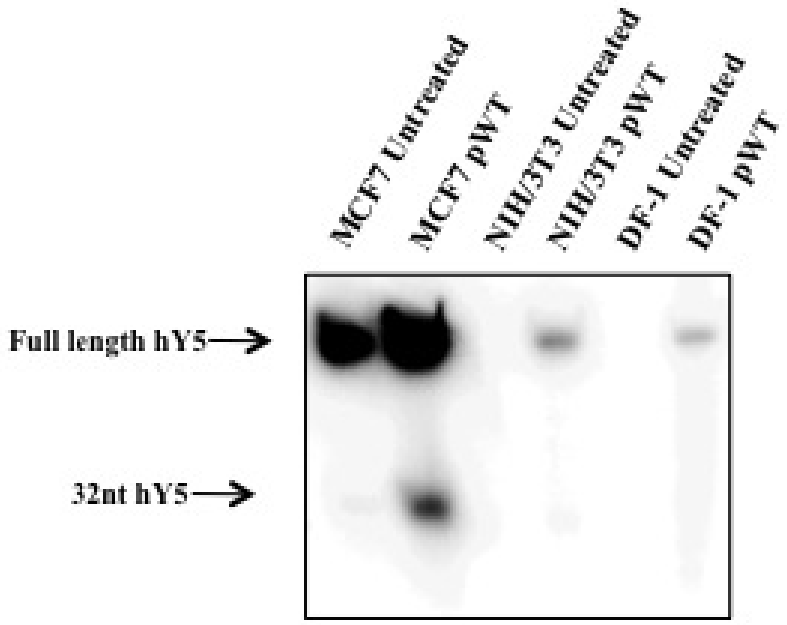


Figure 3.7: Detection of hY5 in cell lines derived from different organisms. Human (MCF7), mouse (NIH/3T3) and chicken (DF-1) cells were plated in 6 well plates. After 24 hours they were transiently transfected with 2 μg of plasmid DNA using the Lipofectamine2000 transfection reagent (lanes 2, 4, 6) or were left untreated (lanes 1, 3, 5). After a further 24 hours cells were harvested and total RNA was extracted. 1 μg of total RNA from each sample was used for northern blot analysis. Membranes were probed with a hY5 3' end-specific radiolabelled probe.

well plate. RNA was extracted and separated by denaturing PAGE, before northern blotting and radioactive probing using a probe specific for the hY5 3' end ydRNAs (Figure 3.7). A band corresponding to the full length hY5 RNA was seen in both the MCF7 untreated and the pWT-transfected cells. This band was noticably more intense in the transfected cells, indicating strong expression from the plasmid. A lower band was clearly visible in the transfected cells, whereas a band of this size was not visible in the untreated cells. This band represents the 32 nucleotide Y5 3' ydRNA. We know from previous experiments that the 24 nucleotide ydRNA is only detectable after longer incubations with Poly(I:C). There seemed to be no cross-hybridization of the hY5 3' end-specific probe with other hY RNAs in the human MCF7 cells, which would be present above the full length hY5 band as the other hY RNAs are longer. The specificity of this probe has since been more closely examined by Dr. Hall (Hall, 2013).

Transfection of pWT resulted in detectable levels of full length Y5 in both the chicken and mouse cell lines. Cleavage products were not detectable in these cell lines. There appeared to be no cross-hybridization of the hY5 3' end-specific probe with endogenous Y RNAs in either the mouse or chicken cells: these cells contain Y RNAs of a different size and only a single full length Y RNA band is visible in these samples. This band is also present only in the transfected cells, giving us confidence that it is the hY5 RNA produced from the plasmid.

3.2.3 Expression of hY5 from plasmids in MCF7 cells

The region targeted for mutagenesis was at the 3' end of the hY5 RNA gene. Mutations would be inserted by replacing a small region of the gene with a mutant double-stranded DNA. No suitable restriction enzyme sites were available close to this region in the pWT plasmids, so they were artificially inserted using a series of PCRs. The pWT plasmid was used as a template to amplify the hY5 gene in two separate sections, using primers to introduce an AgeI restriction enzyme site upstream and an AfeI restriction enzyme site downstream of the RNA-encoding region. Overlap PCR was used to combine these sections and the resulting PCR product was ligated into the pGEM-T Easy vector to give pWT-2RE (Figure 2.2). In order to determine the effect of the restriction sites on transcription of the RNA, the pWT and pWT-2RE plasmids were transfected into human MCF7 cells. As controls, cells transfected with an empty vector were included in the experiment, alongside untreated cells and cells grown in the presence of the Lipofectamine2000 transfection reagent alone. As an extra control, MCF7 cells were also transfected with an unrelated empty vector (pGL3).

The levels of hY5 full length RNA detected in MCF7 cells transfected with the

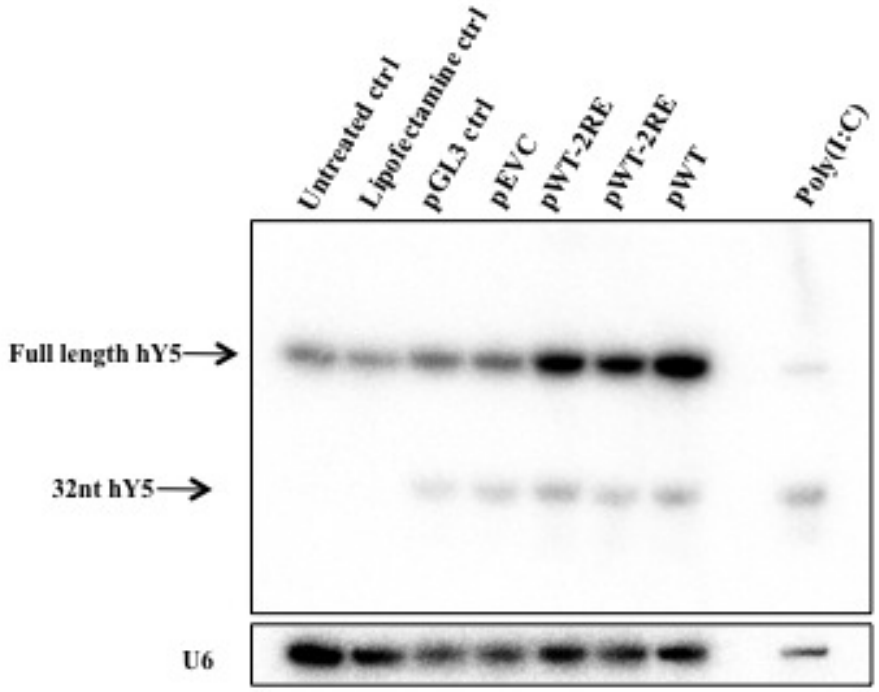


Figure 3.8: Detection of hY5 levels after transfection of MCF7 cells.
MCF7 cells were plated in T75cm^2 flasks and 48 hours later were transiently transfected with 20 μg of plasmid DNA using the Lipofectamine2000 transfection reagent. After a further 24 hours cells were harvested. 1 μg of total RNA was separated by denaturing PAGE and probed with the hY5 3' end-specific probe. As controls, untreated cells were grown alongside cells treated with Lipofectamine2000 alone, cells transfected with a pGL3 control plasmid or transfected with an empty vector. 250 ng of Poly(I:C)-treated MCF7 cell total RNA was loaded in lane 10.

pWT or pWT 2RE plasmids looked similar, and both had markedly higher levels of full length hY5 RNA than the controls (Figure 3.8). However, the levels of cleavage products detected in neither the pWT nor the pWT-2RE transfected lanes were noticably higher than endogenous hY5 levels (as determined by lanes 1-4), especially in those cells transfected with control plasmids. 250 ng of a Poly(I:C)-treated MCF7 total RNA sample was run in lane 10. This control confirms that both the full length hY5 RNA produced from these plasmids and the resulting cleavage products migrate at the same size as the endogenous full length RNA and cleavage products detected after Poly(I:C) treatment, suggesting that the exogenous RNA is being recognised and processed in the same way as the endogenous RNA.

3.2.4 Cleavage of exogenous hY5 in DLD-1 cells after Poly(I:C) treatment

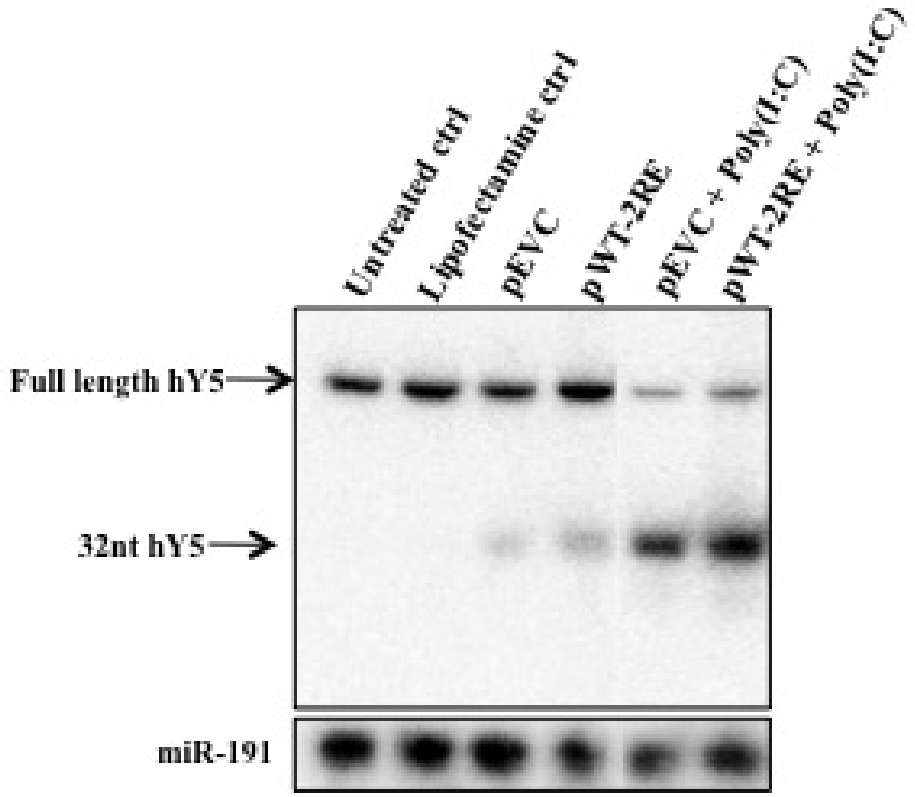


Figure 3.9: Detection of hY5 RNA levels after transfection of DLD-1 cells and Poly(I:C) treatment.
Cells were plated in T75cm^2 flasks and grown for 24 hours before transient transfection with 20 μg of plasmid DNA using the Lipofectamine2000 transfection reagent. 24 hours after transfection, cells were treated with 10 μg/ml Poly(I:C) or fresh media was added. After 8 hours of treatment cells were harvested and 10 μg of RNA was used for northern blot analysis. Samples were probed with the 12 nt hY5 3' end-specific LNA probe. As controls, untreated cells were grown alongside cells treated with Lipofectamine2000 alone or cells transfected with an empty vector control.

Variability in the levels of endogenous of hY5 has been reported in different tissues (O'Brien and Harley, 1990). Transient transfection experiments were therefore conducted in a different human cell line, DLD-1, to see whether transcription and cleavage rates might be more favourable (Figure 3.9). However, very high levels of endogenous hY5 were evident in DLD-1 cells, making interpretation of cleavage of exogenous hY5 RNA difficult. In these cells there was a barely detectable increase

in full length hY5 after transfection. Unlike the previous northern blot, this membrane was probed with a 12 nt locked nucleic acid (LNA) probe specific for hY5. This probe will be used to detect potential ydRNAs from mutant hY5 RNAs, as well as wild type ydRNAs. The original hY5-specific probe cannot be used to detect mutants, as the sequence it maps to overlaps the region targeted for mutagenesis.

Based on these results, we decided to move to expressing hY5 wild type and mutant plasmids in a system without endogenous hY5, so that the expression of wild type and mutant hY5 RNA expressed from plasmids could be directly compared and we could be confident that any conclusion drawn on the effect of the mutation would not be masked by endogenous hY5.

3.2.5 Analysis of the NIH/3T3 cell lines for use in hY5 mutant cleavage assays

Initial transient transfection experiments in mouse NIH/3T3 cells highlighted that there was a difference in transcription efficiency between the pWT and the pWT-2RE plasmids, with the pWT-2RE-transfected lanes showing much lower levels of full length hY5 than the pWT-transfected lanes (Figure 3.10). This dramatic difference was not as clear when these plasmids were expressed in MCF7 cells (See Figure 3.8). This difference was most likely masked by the endogenous hY5 RNA in human cells, although it could also be the result of inherent differences between the two organisms.

It was clear that this difference would make detection of mutant RNAs difficult in NIH/3T3 cells, so the pWT-2RE and the mutant plasmids were corrected by a second PCR using the hY5 wild type forward primer and a reverse primer designed to correct the most 3' restriction enzyme site by returning it to the wild type sequence. This new wild type plasmid was denoted pWT-1RE. The expression of hY5 from pWT, pWT-1RE, and pWT-2RE was compared by transfecting them into NIH/3T3 cells (Figure 3.11). Although the levels of full length RNA produced from the plasmid with one restriction enzyme site was lower than the plasmid with purely wild type sequence, the levels were vastly improved when compared to the plasmid with two sites, confirming that both of these restriction enzyme sites have an effect on the levels of transcription of this gene. Correction of the 5' restriction enzyme site would have required numerous overlap PCRs, so to take account of the differences in expression of hY5 from pWT and pWT-1RE, the levels of mutant hY5 RNA were compared to the levels of wild type hY5 RNA produced from the pWT-1RE plasmid in all subsequent experiments.

The levels of ydRNA detected in prior experiments were higher in cells treated with Lipofectamine2000/empty vector DNA mixes than untreated cells grown only in OptiMEM, suggesting some toxicity of these mixes to the cells. Fugene6 is a

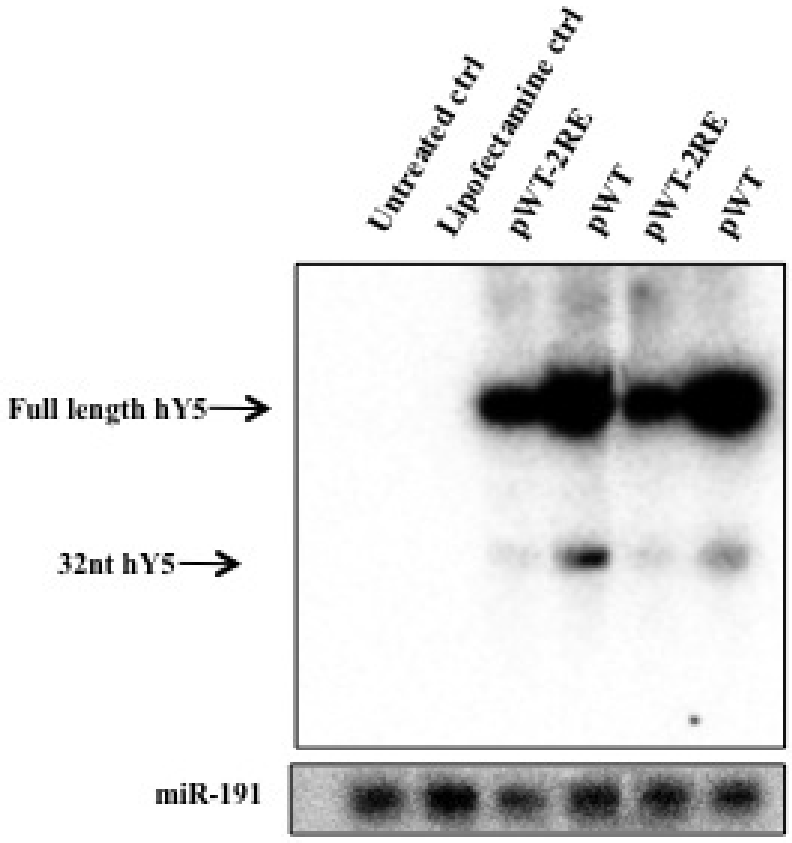


Figure 3.10: Detection of hY5 RNA levels from pWT and p2RE plasmids in NIH/3T3 cells.
Cells were plated in T75cm^2 flasks and 24 hours later were transfected with plasmid DNA using the Lipofectamine2000 transfection reagent. After a further 24 hours, cells were harvested and 3 μg total RNA was used for northern blot analysis. Samples were probed with the hY5 3' end-specific probe.

transfection reagent known to be less toxic to cells than Lipofectamine2000. Comparison of the growth of cells treated with Lipofectamine2000 alone and Fugene6 alone also indicated a greater toxicity, as cells grew faster and were more confluent after treatment with the latter transfection reagent than the former (data not shown). In lane 6 of Figure 3.12 cells were transfected in the same manner as for lane 2, but with Fugene6 rather than Lipofectamine2000. The higher level of full length hY5 in the Fugene6-transfected compared to the Lipofectamine2000-transfected lane shows that Fugene6 gave a higher transfection efficiency in NIH/3T3 cells, as well as a lower toxicity. For this reason, Fugene6 transfection reagent was used in all subsequent mutant hY5 transfection experiments.

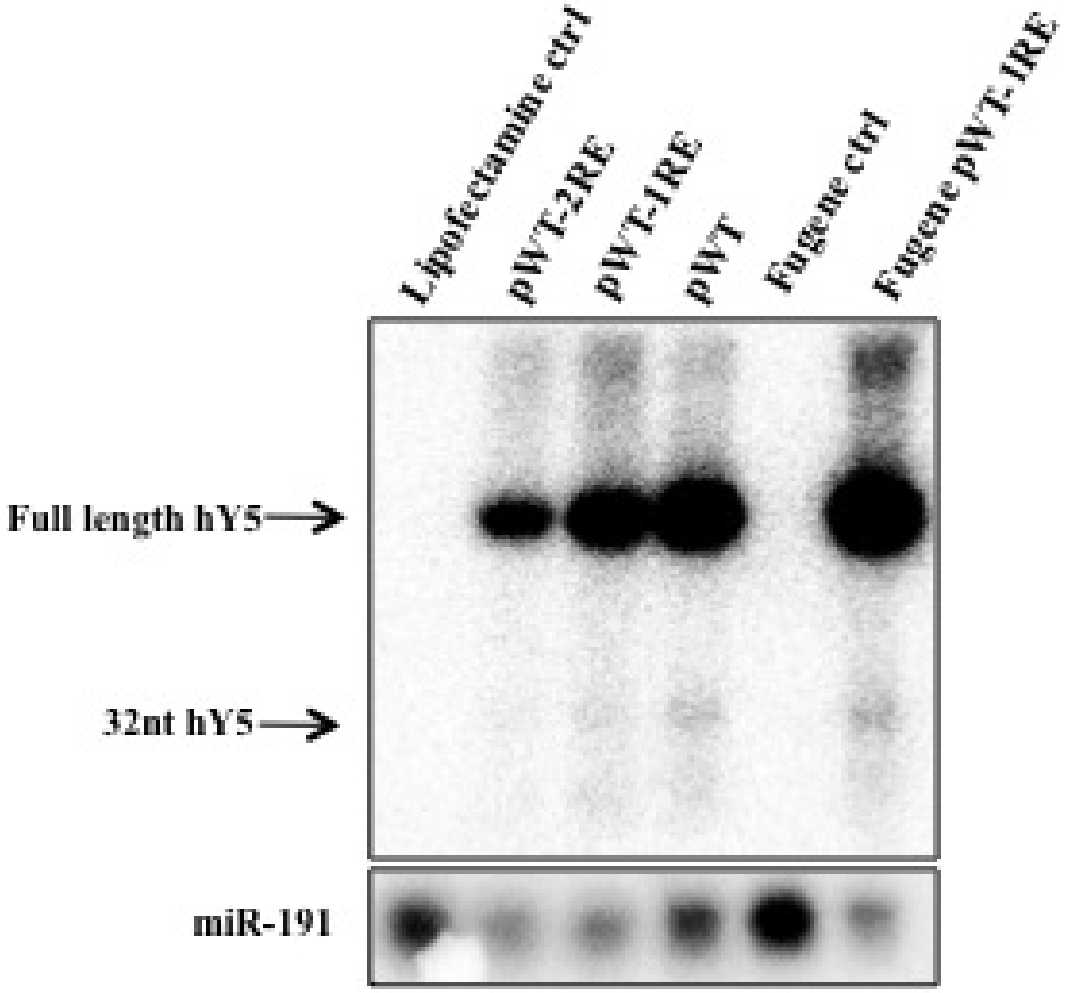


Figure 3.11: Detection of RNA expression from pWT, p1RE and p2RE in NIH/3T3 cells.
NIH/3T3 cells were plated in T75cm^2 flasks and 24 hours later were transfected with plasmid DNA using the Lipofectamine2000 transfection reagent, except for lanes 5 and 6, where the Fugene6 transfection reagent was used. After a further 24 hours, cells were harvested and 10 μg total RNA was used for northern blot analysis. Samples were probed with the hY5 3' end-specific probe.

3.2.6 Optimization of apoptosis induction and ydRNA detection in NIH/3T3 cells

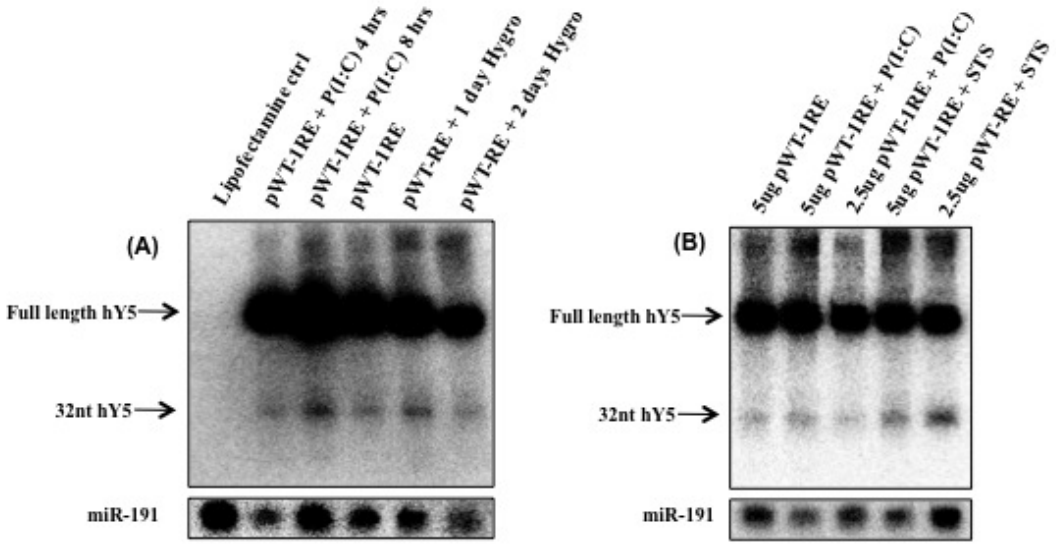


Figure 3.12: Detection of hY5 ydRNAs after apoptosis induction.
NIH/3T3 cells were plated in T75cm^2 flasks and 24 hours later were transfected with pWT-1RE using Lipofectamine2000 transfection reagent. After a further 24 hours, cells were treated with Poly(I:C), (A) Hygromycin or (B) Staurosporine (STS) for the stated time before harvesting and northern blotting. Control cells were treated with Lipofectamine2000 alone. Samples were probed with the hY5 3' end-specific probe. P(I:C) - Poly(I:C), Hygro - Hygromycin.

ydRNAs were at the limits of detection in unstressed NIH/3T3 cells and these cells were also much less sensitive to Poly(I:C) than MCF7 cells. Treatment with Poly(I:C) requires direct transfection into cells and is not ideal for testing the cleavage of exogenous hY5 RNA, as it requires a second transfection of cells already transfected with a plasmid. To minimize the transfection toxicity, along with switching to Fugene6 for transfections, drug-induced apoptosis was also tested. These drugs could be added directly to the cell culture media, thereby reducing stress from sources other than the drug itself. Addition of drugs diluted in cell culture media directly to the flasks should also increase the consistency between samples, as media from the same drug batch could be added to each flask. Test experiments were conducted using Hygromycin, but again the response to this drug was slow and resulted only in low levels of detectable hY5 ydRNA after three days of treatment (Figure 3.12 (A)). NIH/3T3 cells have been shown to be very sensitive to Staurosporine

(STS) treatment with visible signs of apoptosis occurring after just 4 hours (Liu et al., 2001). Experiments were therefore conducted to mirror these conditions and a 4 hour treatment with a final concentration of 1 μM of Staurosporine 24 hours after plasmid transfection gave detectable levels of hY5 ydRNA. Cells could clearly be seen to be undergoing apoptosis, showing signs of rounding under the microscope.

3.2.7 Analysis of the levels of mutant hY5 RNAs and cleavage products in NIH/3T3 cells

Sense and antisense primers covering the entire RNA-encoding region of the hY5 RNA gene and containing mutations (specificed later) were annealed to each other to give blunt and sticky ends complementary to the ends produced when DNA is digested with the AfeI (blunt) and AgeI (sticky) restriction enzymes. A stuffer plasmid was created where the 100 bp hY5-encoding region was replaced with a 1 kb unrelated sequence. This plasmid was then digested with AfeI and AgeI and ligated to the appropriate mutant DNA. To correct for any effects of the second restriction enzyme site on transcription efficiency, a final PCR amplification step was carried out to remove the terminal restriction enzyme site and restore the full 30 bp sequence downstream of the termination signal in all of the mutant plasmids. The PCR products containing the hY5 gene with wild type or mutant sequences were then re-cloned into new pGEM-T Easy plasmids. All constructs were sequenced at each stage to confirm that no mismatches had been introduced and to verify their identity.

3.2.7.1 The deletion mutants

Figure 3.13 shows the predicted secondary structures of the wild type hY5 RNA, the mutant with one strand deleted (Δ1S), and the mutant with both strands deleted (Δ2S). These structures were predicted using the RNAfold default prediction algorithm. (See Appendix 7 for a fuller prediction of the secondary structure of each mutant using all RNA folding algorithms in RNAfold and in Mfold). The Δ1S mutant was designed so that a sequence on the right-hand side of the stem (contained in the red boxed region on the wild type prediction) was deleted. This mutation was predicted to have a dramatic effect on the secondary structure of the resulting RNA molecule (Figure 3.13 shows that this mutation affects the larger bulge in region 2 and the stem in region 3). With the secondary structure predicted to be so dramatically altered, the hypothesis was that this mutant would produce no detectable cleavage products as it would not be recognised by the cleavage machinery.

The Δ2S mutant was designed so that the sequences on both sides of the stem within the red boxed region were deleted. According to predictions, this second

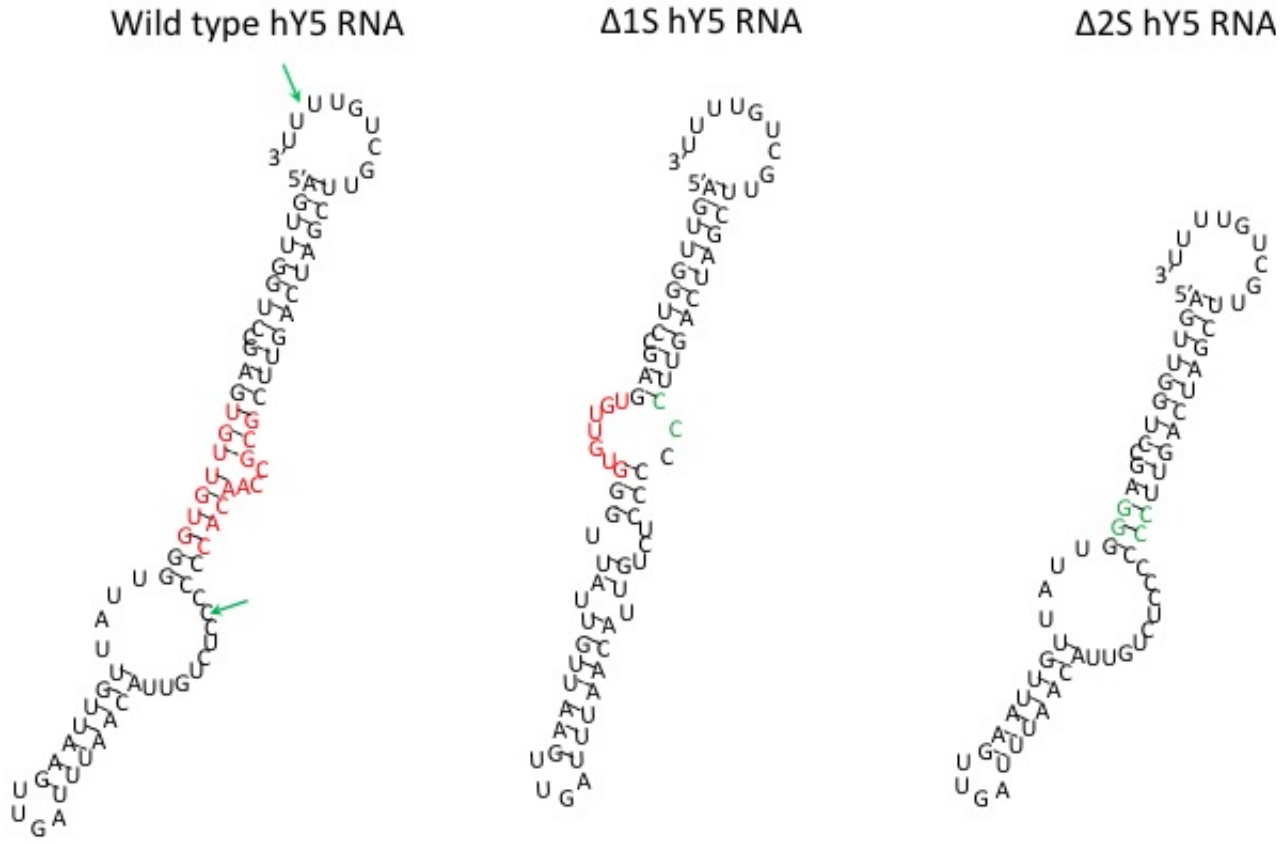


Figure 3.13: Secondary structure predictions of the deletion mutants compared to the wild type hY5 RNA.
Predictions were run using the Turner 2004 RNAfold algorithm, assuming a temperature of 37°C, and were then redrawn manually. The red nucleotides represent the remaining wild type sequence and the green nucleotides represent the nucleotides on the borders of the deleted regions. The green arrows represent the cleavage positions on the wild type structure that produce the 32 nucleotide fragment.

mutation should rescue the first as, although a portion of the RNA molecule has been deleted, the rest of the structure closely resembles the wild type structure. Determining whether this mutant was cleaved would provide us with information as to whether the region targeted for deletion is important for cleavage, or whether it is dispensable in this process.

The full length Δ1S mutant RNA appeared to be consistently lower than either the wild type or Δ2S mutant (Figure 3.14). Neither the Δ1S nor Δ2S mutants produced fragments that were detectable with the 12 nt Y5 3' end-specific LNA probe, even when the membrane was exposed for a much longer period of time (data not shown). For Δ1S it is possible that there was a fragment produced with the same efficiency as the wild type, but that it was below the levels of detection as a result of the levels of the Δ1S mutant full length RNA being so much lower than the levels of the wild type full length RNA. This is less likely to be the case for Δ2S which, in some of the replicates, was produced at levels comparable to the wild type but still no fragments were detected.

3.2.7.2 The substitution mutants

Figure 3.15 shows the predicted secondary structures of the wild type, sub1S, and sub2S mutant RNAs using the default prediction algorithm in RNAfold. Sub1S was designed so that the sequence on the right-hand side of the stem, here highlighted in the blue box on the wild type prediction, was substituted for different nucleotides. The sequence was chosen to deliberately disrupt the secondary structure of the RNA molecule as much as possible (Figure 3.15 shows that this mutation affects regions 2, 3, and 4, introducing an extra stem loop compared to the wild type prediction). With the secondary structure predicted to be so dramatically altered, the hypothesis was that this mutant would produce no detectable cleavage products.

The sub2S mutant was designed so that the sequences on both sides of the stem in the blue boxed region were substituted for different nucleotides. According to predictions, this second mutation rescues the first as, although a portion of the RNA molecule has a different sequence to the wild type, the secondary structure resembles the wild type RNA. Determining whether this mutant was cleaved would provide us with information as to whether it is the sequence within this region or whether it is the secondary structure that is important for cleavage to occur. The blue region highlighted on the sub1S and sub 2S mutant predictions shows the nucleotides that replaced those on the wild type prediction.

All repeat transfections of these mutants clearly showed that there was no 32 nt fragment produced from sub1S, but there was from sub2S (Figure 3.16). This indicates that a change in the structure abrogated cleavage, but where the structure was the same and the sequence was different, a clear 32 nt apoptosis-dependent

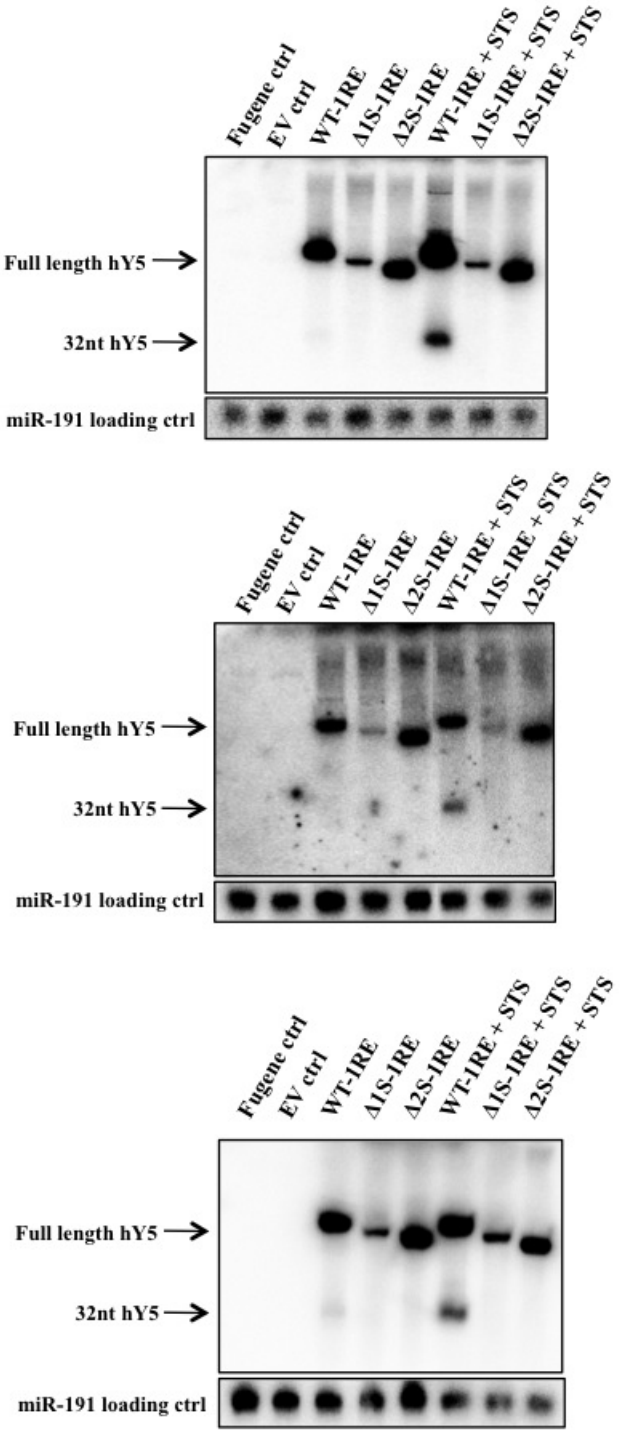


Figure 3.14: Detection of hY5 deletion mutant cleavage after STS treatment. NIH/3T3 cells were plated in T75cm^2 flasks and 24 hours later were transfected with plasmid DNA using Fugene6 transfection reagent. Then 24 hours after transfection, cells were treated with 1 μM Staurosporine for 4 hours before harvesting and 10 μg of total RNA was used in northern blotting. Samples were probed with a 10 nt hY5 LNA probe. In experiment 1 cells were transfected with 2.5 μg (top) of plasmid DNA, in experiment 2 cells were transfected with 5 μg (middle) of plasmid DNA, and in experiment 3 cells were transfected with 10 μg (bottom) of plasmid DNA.

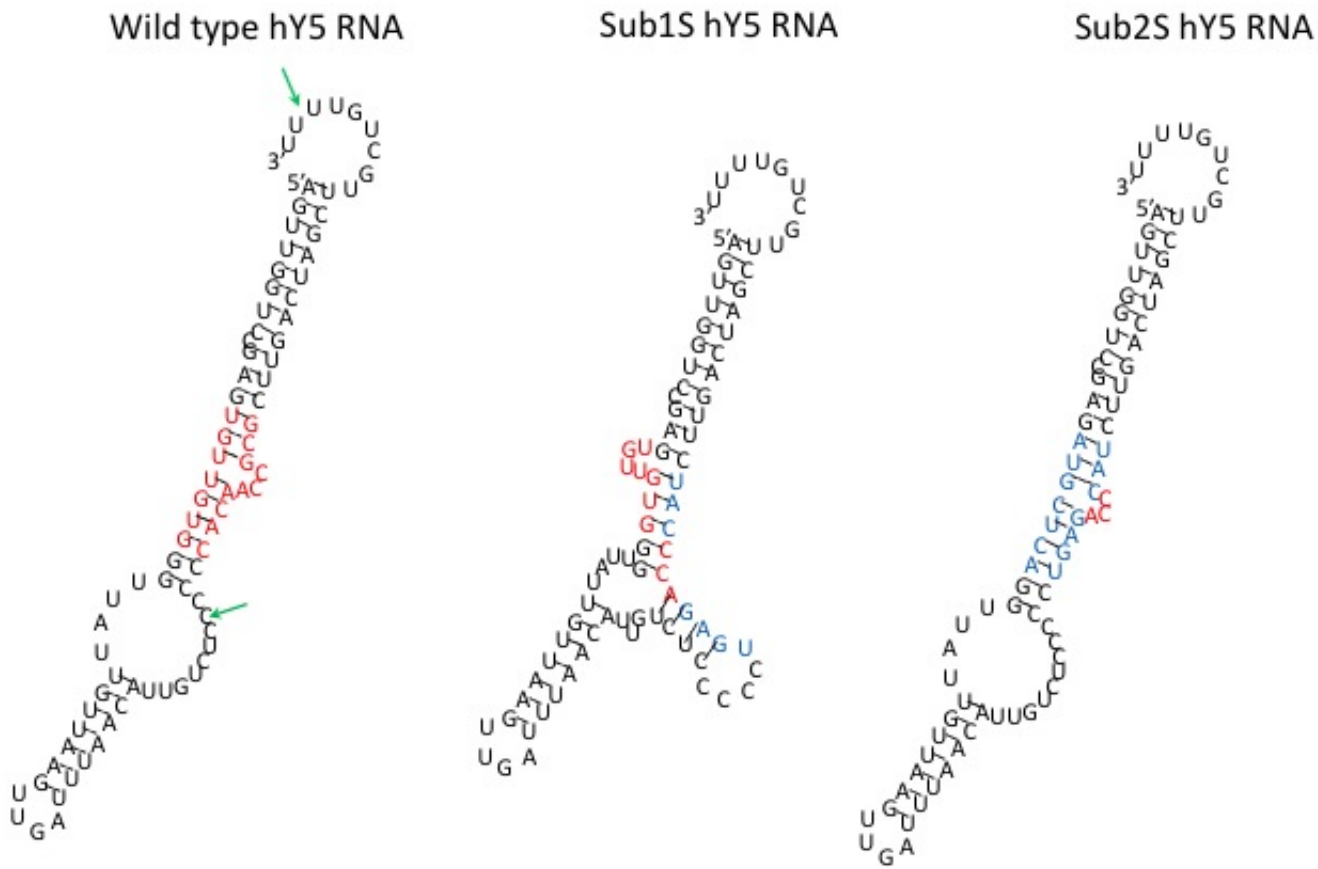


Figure 3.15: Secondary structure predictions of the substitution mutants compared to the wild type hY5 RNA.
Predictions were run using the Turner 2004 RNAfold algorithm, assuming a temperature of 37°C, and were then redrawn manually. The red nucleotides represent the remaining wild type sequence and the blue nucleotides represent the nucleotides that have been substituted. The green arrows represent the cleavage positions on the wild type structure that produce the 32 nucleotide fragment.

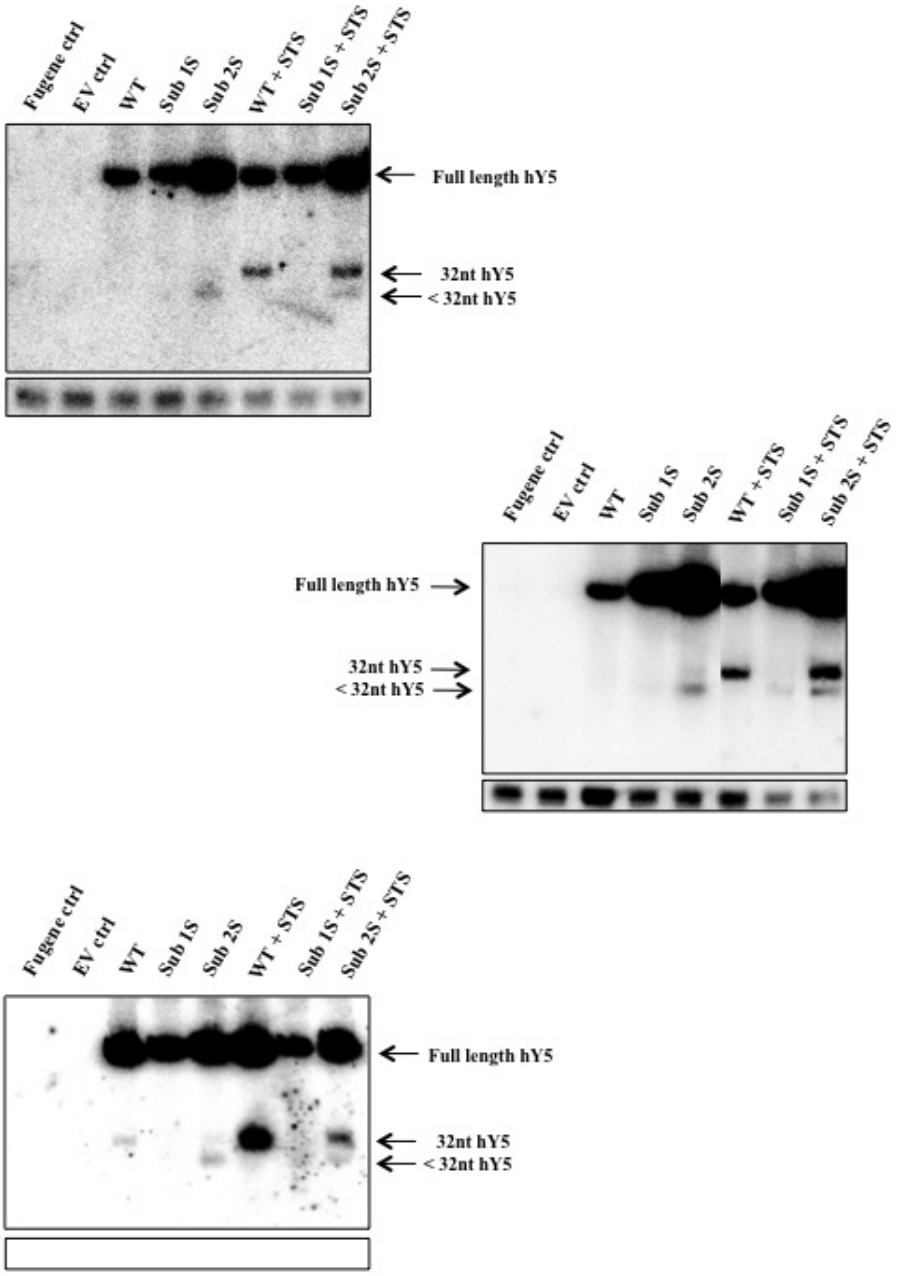


Figure 3.16: Detection of hY5 substitution mutant cleavage after STS treatment. NIH/3T3 cells were plated in T75cm^2 flasks and 24 hours later were transfected with 5 μg of plasmid DNA using Fugene6 transfection reagent. Then 24 hours after transfection, cells were treated with 1 μM Staurosporine for 4 hours before harvesting. Membranes were probed with the 12 nt hY5 LNA probe.

cleavage product was detectable. Interestingly, there was a smaller cleavage product detectable in the sub2S lane that was not detectable in the wild type lane. This cleavage product was not apoposis-dependent as it was present at similar levels in both the unstressed cells and the cells undergoing apoptosis. It is possible that this cleavage product was also present in the sub1S STS-treated lane, but was too faint to detect with any level of confidence.

On different northern membranes, a Δ3' mutant was run alongside sub1S and sub2S (Figure 3.17). This mutant was missing three nucleotides between the Ro60- and La-binding sites towards the 3' end of the molecule. The region where the three nucleotides were deleted is shown in orange on the predicted structure (Figure 3.18). The Δ3' mutant could not be detected using the 12 nt LNA probe, as some of the nucleotides required for hybridization were deleted in this mutant (data not shown). The membranes that included this mutant were reprobed with a 10 nt LNA probe designed further away from the 3' end of the full length RNA, and therefore not overlapping the deleted region in Δ3'. The new 10 nt LNA probe detected a similar band pattern to the 12 nt LNA probe, but this time it also detected a full length RNA and a cleavage product in the Δ3' samples. In repeat 1, the full length Δ3' mutant RNA was present at even higher levels than the full length sub2S mutant, which explains why an sRNA was present at low levels in the Δ3' untreated lane, and an sRNA was present in the Δ3' STS-treated lane at higher levels than in the WT STS-treated lane. This Δ3' ydRNA band was not clearly detectable in the untreated sample in the other two repeats. This Δ3' ydRNA was noticeably smaller than the WT ydRNA, suggesting that the cleavage happened in this mutant at the same position as in the WT, resulting in a smaller product.

3.2.7.3 Further mutants

Two proteins known to consistently interact with all types of hY RNA are Ro60 and La; interestingly they both interact with the region of the hY5 molecule that is present in the 5' and 3' apoptosis-dependent ydRNAs. The ΔRo mutant was designed to further investigate the role of Ro60 in the cleavage of hY5 (Figure 3.18). The orange line highlights where 4 nt from the right and 3 nt from the left of the terminal stem were deleted. This mutation includes a deletion of the C bulge.The secondary structure is not predicted to be dramatically altered by these deletions. The region deleted is, however, one of the most conserved regions across Y RNAs and across organisms, as it is the region that interacts with Ro60. Similar mutants have been made by a number of groups and have been shown to lack the ability to associate with Ro60. The ΔRo mutant has no detectable ydRNAs 3.19, even though the full length mutant is present at only slightly lower levels than the wild type RNA.

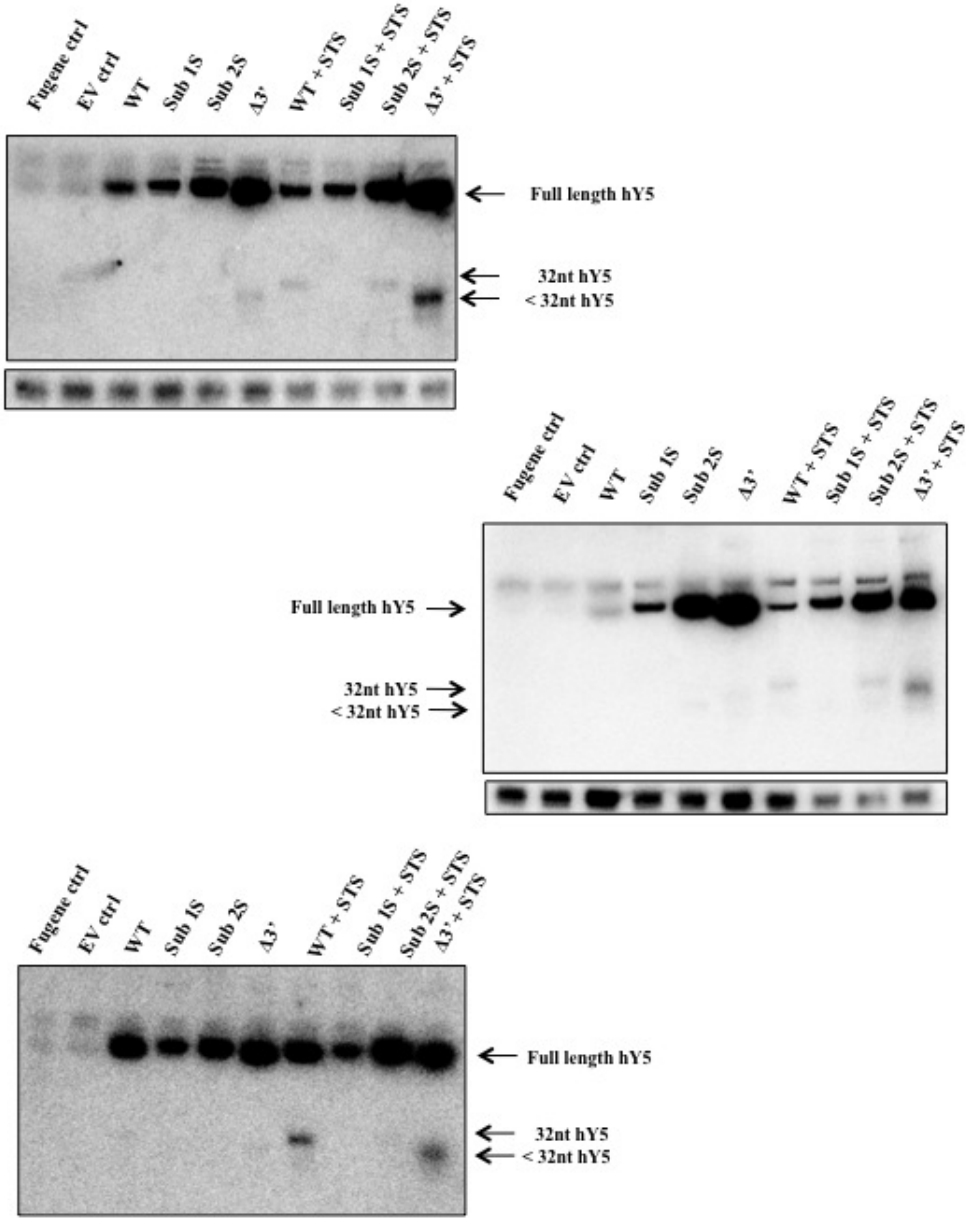


Figure 3.17: Alternative detection of hY5 substitution mutant cleavage after STS treatment.
NIH/3T3 cells were plated in T75cm^2 flasks and 24 hours later were transfected with 5 μg of plasmid DNA using Fugene6 transfection reagent. Then 24 hours after transfection, cells were treated with 1 μM Staurosporine for 4 hours before harvesting. Northern blots were probed with the 10 nt hY5 LNA probe.

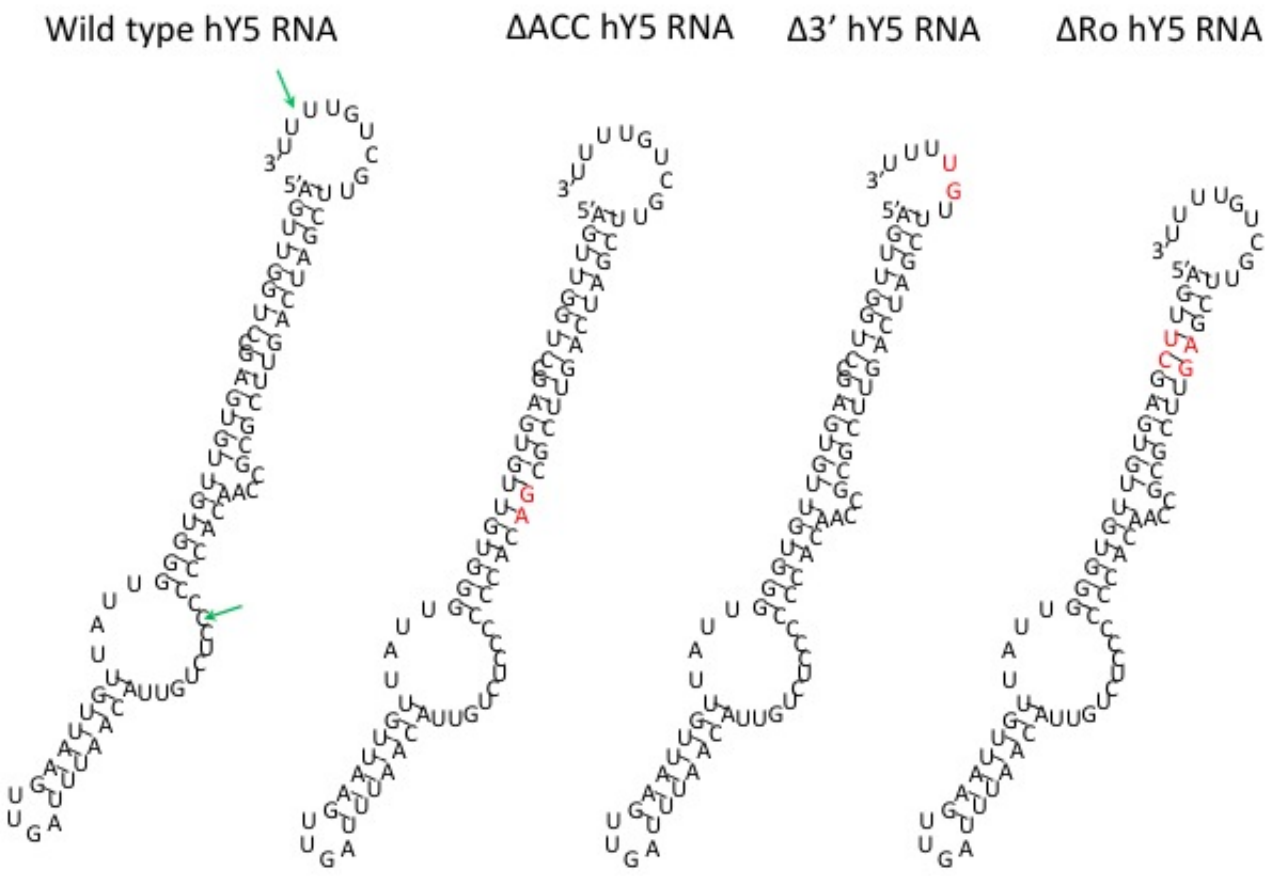


Figure 3.18: Secondary structure predictions of further mutants compared to the wild type hY5 RNA.
Predictions were run using the Turner 2004 RNAfold algorithm, assuming a temperature of 37°C, and were then redrawn manually. The red nucleotides represent the nucleotides on the borders of the deleted regions. The green arrows represent the cleavage positions on the wild type structure that produce the 32 nucleotide fragment.

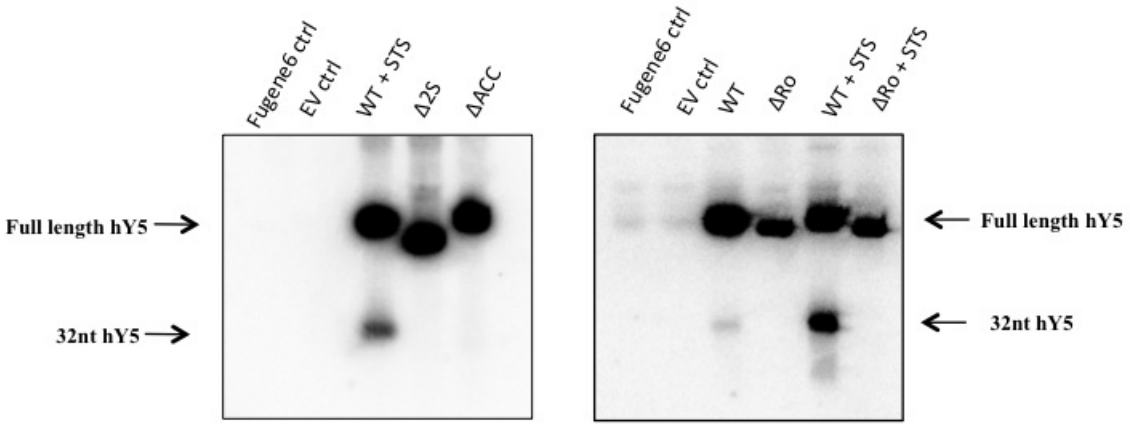


Figure 3.19: Detection of hY5 further mutants cleavage after STS treatment. NIH/3T3 cells were plated in T75cm^2 flasks and 24 hours later were transfected with 5 μg of plasmid DNA using Fugene6 transfection reagent. Then 24 hours after transfection cells were treated with 1 μM Staurosporine for 4 hours before harvest. Northern blots were probed with the 12 nt hY5 LNA probe.

To examine the effect of smaller changes on cleavage, a further mutant was designed (Figure 3.18). In the ΔACC mutant, 3 nts in the predicted large bulge were deleted. The regions where the nucleotides were deleted is marked on the structure prediction in orange. The aim of this mutant was to determine whether this bulge is necessary for cleavage. The large bulge in experimentally validated Y RNA structures is thought to be nearer the terminal end than the predicted structure, but the stem produced by the prediction program when the ACC nucleotides are deleted has a low predicted free energy of -32.31 kcal/mol, so this deletion is likely to disrupt any downstream bulge which may be present in the wild type. The Wolin group propose that this bulge is important for the recognition of this RNA by Ro60 (Green et al., 1998), so this mutant was also designed to be used in immunoprecipitation experiments, to determine whether the large bulge really is necessary for Ro60 binding.

The ΔACC mutant, like the Δ2S mutant, which was run here as a control, is not cleaved(Figure 3.19). No ydRNAs are seen in either of these lanes, while a clear ydRNA band is visible in the wild type lane.

3.3 DISCUSSION

Poly(I:C) has been shown by others to induce apoptosis (Larson et al., 1969), but has not been tested at the concentrations and incubation times used in the preliminary work and, although the cells look to be undergoing apoptosis when inspected visually, one of the first aims of this project was to prove that apoptosis was occurring under these conditions. Initial transfection experiments were conducted in MCF7 and DLD-1 cells, both of which visibly showed signs of undergoing apoptosis after 8 hours of treatment. Both MCF7 and DLD-1 cells were derived from human cancers and both have characteristics different from normal human cells. MCF7 has been documented to lack the Caspase 3 protein (Kirsch et al., 1999), a key effector protein in one of the major apoptosis pathways. For this reason, the DLD-1 cell line was instead utilized for assaying apoptosis. DLD-1 cells were grown under various conditions, including Poly(I:C) treatment and treatment with the apoptosis inducer Hygromycin, and subsequently harvested. Samples were split into two and processed either for northern blot analysis or for flow cytometry analysis. Flow cytometry samples were further split and stained with Annexin V or PI. On obtaining the flow cytometry data, each sample was gated to select cells and remove debris. Phosphatidyl serine residues (PS) are located on the inner surface of the cell membrane in healthy cells. Early in apoptosis these PS residues are translocated to the outside of the membrane. Annexin-V targets these PS residues and stains them. PI only enters dead cells and indicates early or late necrosis. The Annexin V/PI

plots (Figure 3.4) show that 87.4% of cells in the Poly(I:C) treated sample are undergoing apoptosis and 10.9% are healthy, whereas in the Hygromycin treated cells 71.7% of cells are undergoing apoptosis and 13.6% of cells are healthy. Only a small proportion of cells under either of these two treatment conditions have necrosed. From the Annexin V/PI plots the location of the majority of cells is different for the Poly(I:C) and Hygromycin conditions. In the Poly(I:C)-treated sample the majority of cells are in the centre of the lower right quadrant, while in the Hygromycin-treated sample a small proportion are in this position, but the majority of the cells in this quadrant are on the border of the lower left and the lower right quadrants, suggesting an earlier stage of apoptosis. This is reflected in the median FL1-A value (the position along the X axis, representing the level of Annexin-V staining). In all other conditions tested, no detectable population of cells was seen undergoing apoptosis or necrosis (See Appendix 7 for Flow cytometry data table and SSC/FSC gating plots).

Figure 3.5 shows that the level of 32 nt hY5-derived ydRNA in the Hygromycin lane is slightly above the levels in the control lanes and that the level of ydRNA is higher again in the Poly(I:C)-treated sample. This northern blot correlates with the flow cytometry results and from these data we can deduce that the 32 nt Y5 3' ydRNA is likely to be apoptosis-induced. Alongside these flow cytometry experiments, Dr. Nicolas also treated MCF7 cells with varying concentrations of Hygromycin. The northern blot image in Figure 3.20 shows that Hygromycin, like Poly(I:C), induces hY5 cleavage in this cell line, providing further evidence that this ydRNA is apoptosis-induced. From this we can conclude that the specific Y5-derived RNAs in our experiments, as with those detected in (Rutjes et al., 1999b), are apoptosis-dependent rather than viral infection-dependent, especially as we know that this ydRNA is PKR-independent.

Early optimization experiments in mouse cells highlighted the fact that the wild type hY5 RNA was still cleaved to produce a 32 nt 3' ydRNA, despite the fact that this RNA is not endogenous to this organism. These results indicate that the cleavage protein or cleavage complex that produced this sRNA is conserved between mice and humans. This is supported by data from Dr. Hall, which showed that endogenous mouse Y RNAs were also cleaved after Poly(I:C) treatment (Hall, 2013). Further experiments to quantify the levels of these ydRNAs in treated and untreated cells found that they were present in untreated human cells, albeit at much lower levels (Nicolas et al., 2012). So far we have not investigated whether all cells in a normal untreated population produce ydRNAs at low levels or whether these RNAs are derived from a small number of cells within the population which are currently undergoing apoptosis during their natural life-cycle. Fluorescent activated flow cytometry could be used to answer this question. If a large sample of untreated

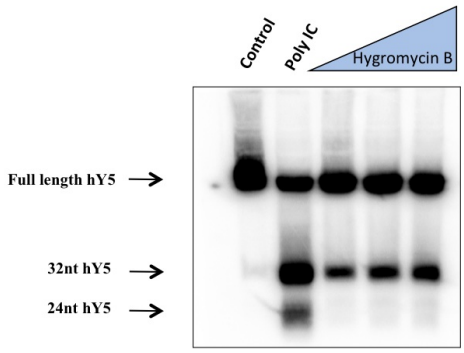


Figure 3.20: Detection of hY5-derived small RNAs after treatment of MCF7 cells with different apoptosis inducers.
Dr. Nicolas showed that, when MCF7 cells are treated with increasing concentrations of Hygromycin, there is an increase in the levels of hY5-derived small RNAs detected by nothern blot analysis. These levels are comparable to those detected after 8 hours of Poly(I:C) treatment.

human cells was sorted by an early apoptosis marker and then hY5 RNA cleavage determined in the apoptotic and non-apoptotic populations, then we could compare the cleavage of hY5 in these populations and determine whether it occurs in non-apoptotic cels.

The addition of a plasmid encoding hY5 RNA led to an increase in the levels of hY5-derived cleavage products in human cells (see Figure 3.7), even in the absence of an apoptosis inducer. This is likely due to the toxic effects of the transfection agent increasing the levels of apoptosis in these cells compared with untransfected cells.

In order to create each of the mutants, their sequences were manually altered and then a combination of Mfold and RNAfold algorithms were used to predict the resultant structural alterations. Both the deletion and substitution mutants were designed to alter regions 2 and 3 of the RNA. The Ro60 and La binding sites were conserved as much as possible when designing these mutants so as not to affect the stability of the full length RNA and complicate interpretation of the results.

In all three repeat experiments of the deletion mutants, the full length Δ1S mutant was present at consistently lower levels than either the wild type or the Δ2S mutant. This could have occurred for a number of reasons. Firstly, the Δ1S RNA could be transcribed at lower levels than the wild type or the Δ2S mutant RNA. The human Y RNA genes have been reported to contain elements that regulate their own transcription and it is possible that one of these was disrupted in this mutant. The fact that this did not occur for the full length Δ2S mutant, which had even more sequence deleted, suggests that this was not the case. A second possibility is that the mutation disrupted the structure of this RNA in such a way that it was recognized

as misfolded and was more rapidly targeted for degradation than the Δ2S mutant. Both deletion mutants were designed to maintain the integrity of the Ro60 binding site and keep the C bulge intact. Green *et al.* have published data to show that the large bulge in the wild type structure is also necessary for Ro60 binding and allows the RNA structure to open up enough to allow Ro60 access (Green et al., 1998). This requirement for the large bulge has not been verified by other groups. If this bulge is necessary for Ro60 binding, then it is unlikely that either of these mutants would bind to Ro60. Y RNAs bind to the outer surface of the Ro60 ring (Stein et al., 2005), but this protein can also interact with misfolded RNAs (Green et al., 1998). It is therefore possible that, not only did it no longer recognize these mutants as Y RNAs, but it was also actively involved in the degradation of Δ1S. Another possibility is that this mutant RNA interacted with other proteins, resulting in its degradation.

It is not possible to make conclusions about the targeting of the Δ1S mutant for cleavage, as the full length levels were so much lower than the levels of full length wild type RNA. No cleavage products were visible in the Δ1S lane, even after STS treatment, but it may be that this RNA was fragmented at a similar rate to the wild type and the cleavage products were not detectable because they were below the levels of detection. Although the full length Δ2S mutant was present at similar levels to the full length wild type Y5 RNA, no detectable cleavage products could be seen. This mutant has a similar secondary structure to the wild type, but is missing the large bulge in region 2 as well as some of the stem in regions 2 and 3. We can conclude that region 2 must be necessary for the production of detectable 32 nt 3' ydRNA.

Along with the binding of Ro60 in this region, the Prujin group has shown that the binding of a nuclear factor to each Y RNA is required for the export of Ro RNPs into the cytoplasm (Rutjes et al., 2001). This nuclear factor is probably ZPB1 (Sim and Wolin, 2011). The exact binding site of this protein has not been identified, but it requires a 10 nt terminal stem for Y RNA interaction (Sim and Wolin, 2011). The Δ1S mutant has an altered secondary structure resulting in a shorter terminal stem, while the Δ2S mutant has a terminal stem one base pair shorter than the wild type RNA. It is therefore interesting that the Δ1S mutant is less stable than the Δ2S mutant. Simons *et. al.* discovered that both Ro60 and this nuclear factor must interact with Y RNAs for them to be relocated into the cytoplasm, but that shorter mutant Y RNAs can be exported when they contain only one of these proteins (Simons et al., 1996). Therefore the Δ1S mutant may not be able to efficiently bind either Ro60 or the nuclear factor, while Δ2S may be able to associate with the nuclear protein more strongly, and this association is enough for transport into the cytoplasm. Simons *et. al.* also hypothesized that degradation occurs in the

nucleus and that, if Y RNA mutants are successfully transported to the cytoplasm, they evade degradation (Simons et al., 1996), giving a plausible explanation for the increased stability of Δ2S over Δ1S. We know that hY5 fragments are found in apoptotic bodies after apoptosis-induction, but presumably this can only occur if they can be exported from the nucleus. Only immunoprecipitation and cellular localisation studies will categorically determine whether this hypothesis is correct, although the ΔRo mutant, which is missing the Ro60 binding site, appears to be as stable, if not more stable, than the wild type RNA, suggesting that stability is not reliant upon Ro60 binding. Perhaps it is only the stability of mouse-derived Y RNAs that is reliant upon mRo60 binding and, as these mutants are derived from a Y RNA which is not naturally produced in these cells, the degradation machinery that targets Ro60-free mouse Y RNAs may not recognise these mutants.

From Dr. Hall's experiments in MCF7 cells (Hall, 2013) it is not possible to make a clear conclusion about the importance of Ro60 in ydRNA production from hY5 RNA, as siRNA-mediated transient Ro60 knockdown and subsequent Poly(I:C) treatment in these cells resulted in a reduced level of full length hY5 and therefore also those ydRNAs derived from it. Using calculations of the ratios of full length to ydRNA in each lane and comparing the siRNA-treated lanes with control lanes, Dr. Hall concluded that Ro60 was required for cleavage (Hall, 2013). To further explore the involvement of this protein, he went on to transiently transfect pWT into mouse embryonic stem cells (mESCs) that were Ro60 -/- and then induced apoptosis. Interestingly, the mouse Y RNAs levels were drastically reduced in the Ro60 -/- cells, while the transfected hY5 levels were actually higher in the Ro60 -/- cells compared to the Ro60 wild type cells. Comparison of the Poly(I:C) treated Ro60 wild type and Ro60 -/- cells in these experiments highlighted that Ro60 is certainly required for cleavage as there were no cleavage products present in the Ro60 -/- cells. It seems likely that the increase in full length levels of hY5 in the Ro60 -/- Poly(I:C)-treated cells compared to the Ro60 wild type Poly(I:C)-treated cells arose because the full length RNA was not processed into ydRNAs in these cells. It is somewhat puzzling that this exogenous Y RNA is unaffected by Ro60 absence, while endogenous Y RNAs are degraded. It would be interesting to create a plasmid containing one of the other hY RNAs and determine whether the same phenomenon is seen on transfection into mouse cells. Perhaps this is an artifact resulting from the production of an artificial RNA or perhaps hY5 RNA cannot function correctly in a cell that does not naturally express it. During their investigations into Ro60 function, the Wolin group designed a number of Ro60 mutants to test its function in Y RNA binding and ncRNA quality control (Stein et al., 2005). To further confirm the involvement of this protein in cleavage, a number of these mutants could be expressed in Ro60 -/- mESCs cells and analysed for cleavage ability.

Unlike the deletion mutants, none of the nucleotide substitutions in the sub1S or sub2S mutants adversely affected RNA stability. In fact the full length sub2S mutant RNA was present at consistently higher levels than either the wild type or the sub1S mutant RNA. The sub1S mutant, like the deletion mutants, did not produce detectable 32 nt 3' ydRNAs on apoptosis induction. Some sub2S molecules however, appeared to be processed in a similar manner to the wild type, resulting in a 32 nt ydRNA which was undetectable in the untreated lane and was present at similar levels to the wild type-derived 32 nt ydRNA in the STS-treated lane. What was unexpected was that this mutant RNA also appeared to be processed in a novel manner to yield an apoptosis-independent ydRNA, as shown by the <32 nucleotide band which was detectable at similar levels in the untreated and STS-treated sub2S lane. It is conceivable that this <32nt small RNA was produced in the cells transfected with the wild type hY5 plasmid but was not at detectable levels as a result of the lower level of full length wild type hY5, but this seems unlikely as this ydRNA was not detected in any previous experiments and was not present in the wild type lane of this experiment, even if the exposure was increased (data not shown). It is possible that the full length sub2S RNA has been altered in such a way that it was recognised as a primary microRNA, although this also seems unlikely as the mutant RNA is not predicted to form the expected stem-loop structure of a primary microRNA. However, this ydRNA is likely to be produced by a different enzyme to the one we are investigating, as it is produced under different conditions.

The sub2S mutant was designed so that the ACC bulge was maintained and the sequences on either side were substituted; Ro60 binding should therefore not be affected. The fact that neither of the deletion mutants nor the sub1S mutant produce cleavage products, while the Sub2S mutant does, suggests that it is the secondary structure rather than the nucleotide sequence, within the region targeted for mutation that is important for cleavage. As the 12 nt LNA probe did not detect the full length Δ3' end mutant RNA or any cleavage products, the three nucleotides deleted in this mutant must be essential for the hybridization of this probe. These nucleotides must therefore be present in all of the substitution mutant ydRNAs detected with the 12 nt LNA probe, in order for this probe to hybridize and produce a signal.

As the large bulge in region 2 is thought to be vital for Ro60 binding (Green et al., 1998), a mutant (ΔACC) was designed where the ACC in the predicted bulge was deleted. When this mutant was compared to the wild type no cleavage products were detected. This is likely because Ro60 binding is crucial for cleavage to occur and this mutation prevents this interaction. In experiments to validate the secondary structure of Y RNAs, the large bulge was found to be nearer the terminal end of the molecule than predicted here (van Gelder et al., 1994). We therefore cannot

with complete confidence conclude that this mutation removes the large bulge, but the striking effect of deleting these nucleotides suggests this to be the case. This conclusion is supported by data from Dr. Hall, where the cleavage products are absent when Ro60 is targeted for siRNA-knockdown (Hall, 2013). It is also possible that another protein, such as the nuclear protein already mentioned, cannot interact with the ΔACC mutant, resulting in a lack of cleavage products. Either way we can conclude that these three nucleotides are essential for cleavage.

The Δ3' mutant could not be detected using the 12 nt LNA probe. The northern blots that included this mutant were re-probed with a 10 nt LNA probe designed further away from the 3' end of the full length RNA, which therefore did not overlap the deleted region. The new 10 nt LNA probe appeared to detect a similar band pattern to the 12 nt LNA probe, although it was less sensitive. This time, however, it also detected a full length RNA and a cleavage product in the Δ3' STS-treated lane. The <32 nt apoptosis-independent band was not visible using this probe, while it was clearly visible when the same membranes are probed with the 12 nt LNA probe (data not shown). As this probe is shorter than the previous LNA probe, this may be due to a lower sensitivity under the same hybridization conditions, or this small RNA may not be picked up by the new probe as it is missing nucleotides required for hybridization. The 10 nt LNA probe was designed against a sequence which is more conserved between Y RNAs than the sequence targeted by the 12 nt LNA probe, which explains why there are a number of bands visible above the full length hY5 RNA band; these are mouse Y RNAs. As there are no small RNA bands present in any of the control lanes, we can be sure that all ydRNAs detected on these northerns are Y5 RNA-derived. The full length Δ3' mutant RNA is present at even higher levels than the full length sub2S mutant, which explains why a ydRNA is present at low levels in the Δ3' untreated lane, and a ydRNA is present in the Δ3' STS-treated lane at higher levels than in the wild type STS-treated lane. This mutant ydRNA is noticeably smaller than the wild type 32 nt ydRNA and this could be for two possible reasons. Firstly, the enzymes involved in cleavage could cut the RNA at a particular site and therefore a 3 nt deletion close to the end of the RNA resulted in a shorter product. The second possibility is that proteins were protecting the RNA and, despite the fact that three nucleotides were missing between where Ro60 and La bind, the rest of the RNA was protected in the same way and the RNA was still cut at the same position. Immunoprecipitation experiments are required to be sure that both Ro60 and La still interact with this mutant and to determine whether it is shorter because protein binding has been lost.

There are two possible hypotheses to explain how apoptosis-dependent Y RNA cleavage occurs. In the first hypothesis, RNA binding proteins interact with each Y RNA in regions 1, 2 and 3 and protect the RNA from degradation. An endonuclease

nicks the RNA somewhere at an internal nucleotide, and then exonucleases degrade the regions of the RNA that are not protected, including the 3' part of the RNA. These proteins include, but are not exclusively, Ro60 and La. The smaller hY5 3' ydRNA is likely to result from the release of La from the 32 nt ydRNA, as this lower band is up-regulated when La is targeted for siRNA-knockdown (Hall, 2013). Numerous groups have used RNase protection assays and electromobility shift assays to determine the regions bound by Ro60, and these regions do not fully encompass the RNA sequences present in the detected apoptosis-dependent ydRNAs (van Gelder et al., 1994; Teunissen et al., 2000). It is possible that Ro60 interacts with these RNAs differently *in vivo* compared to *in vitro* and this could explain these discrepancies, but it seems more likely that other proteins are involved. These other proteins may bind to Y RNAs directly, or they may protect Y RNAs by interacting with Ro60. These questions could be answered by conducting RNA immunoprecipitation experiments and determining the proteins present in complexes with Y RNAs, both in untreated cells and in apoptosis-induced cells. If the protein protection theory is correct, then this raises another possible senario to explain the accumulation of cleavage products. The cleavage rate may be unaltered between apoptotic and untreated cells, but more proteins may bind to Y RNAs during apoptosis than under normal circumstances. The ydRNAs may therefore accumulate as more are protected from exnouclease digestion than under normal circumstances.

In the second hypothesis a cleavage complex binds to the Y RNA and this complex includes an endonuclease. This endonuclease cuts the RNA in a particular position to produce a cleavage fragment. The fragment is stable, while the rest of the RNA is targeted for degradation. Perhaps the fragment is rapidly incorporated into a functional protein complex and is therefore protected from further degradation. Data obtained by Dr. Hall support the second hypothesis, as the hY5 3' end fragment is present at three times higher levels than the hY5 5' end fragment (unpublished data). If both fragments exist as they are protected from degradation by Y RNA binding proteins, then it seems unlikely that one fragment would be present at levels so much higher than the other. Mapping of the predominant 3' and 5' fragments onto the hY5 RNA secondary structure also indicates that these fragments are unlikely to be the result of Y RNA binding protein protection, as the cleavage positions of these two fragments are on very different regions of the RNA. It seems likely that the RNA secondary structure would be partially unwound after the initial cut, and this is how the cleavage sites occur at such different locations on each of these hY5 ends. It is unlikely that these cleavages could occur without some reorganisation of the proteins that are bound.

Follow up experiments to these inital studies must involve Ro60 binding assays in order to more closely determine the role of Ro60 in these cleavage events. When

these binding assays are analyzed it will be important to take into account that Ro60 has two binding surfaces for RNAs. In X-ray crystallography experiments Y RNAs were found to bind on the outer surface of this protein (Stein et al., 2005), but Y RNA fragments were also found to bind inside the Ro60 ring. A positive result in a Ro60 binding assay does not therefore tell us where these mutant RNAs and mutant RNA fragments are binding and how they are being recognised. It is possible that the 3' end fragment is more abundant than the 5' fragment as it interacts with Ro60 inside the ring, while the 5' end is released. So far, whether any of the 3' and 5' fragments remain bound to each other has not been investigated.

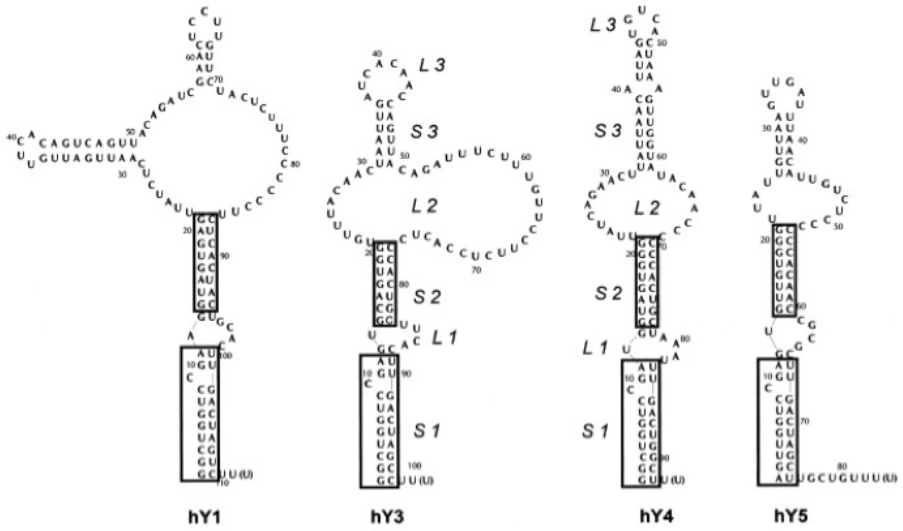


Figure 3.21: Secondary structure of human Y RNAs determined by a combination of prediction and experimental validation.
Reproduced from Teunissen *et al.* 2000.

Initial NGS experiments comparing untreated MCF7 cells to cells treated with Poly(I:C) for 8 hours, showed that after this treatment all hY RNAs were cleaved to produce 3' and 5' fragments. Findings by Rutjes *et al.* in 1999 corroborate our NGS data, as they found that all hY RNAs were processed to give cleavage products after the induction of apoptosis. We have shown that, after 8 hours of Poly(I:C) treatment, DLD-1 cells were undergoing apoptosis, and that the hY5 RNA was cleaved. From our initial sequencing data in MCF7 cells, we determined that the predominant sizes of the hY5 fragments were 32 and 24 nt and the fragment sizes of the other Y RNAs were 18-32 nucleotides (Hall, 2013). The method Rutjes *et. al.* used for examining Y RNA fragments was high resolution PAGE whereupon fragments that

differ by 1 nt in size can be separated (Rutjes et al., 1999b). Their method involved probing for all of the hY RNAs at once and used four oligonucleotides antisense to the entire length of each of the hY RNAs as probes, so each gel represents all fragments for each of the four hY RNAs combined (Rutjes et al., 1999b). Dr. Hall's data determined that a number of the hY RNAs produce fragments predominantly of a size between 30-32nt and 27nt (Hall, 2013), although like hY5 it is possible that these are further processed into smaller fragments that are less abundant or are present at higher levels during different treatment conditions. Comparison of his data with that of Rutjes suggests that the larger band they see on their sizing gels was made up mainly of the 5' fragments of hY1, hY3 and hY5, which were all around 31-32nt (Rutjes et al., 1999b). We have reached the same conclusion, but due to the length of the sequence reads in the initial NGS dataset and the adapter removal method, we cannot be sure of the accuracy of this sizing and are awaiting sequencing data with longer reads to confirm this (See Chapter 5).

RNA immunoprecipitation experiments carried out by this group support our data as, without sequence information and solely going on size, they determined that the longer fragments bind to both Ro60 and La and that the shorter fragments do not bind to La (Rutjes et al., 1999b). They suggest that this might be because the shorter fragments have lost the terminal uracils (Rutjes et al., 1999b). This theory is supported by our hY5 data, as the 32 nt fragment contains the La binding site, but the shorter fragment that is produced later does not. In the other hY RNAs the run of Us recognised by La is closer to the end than in hY5, which makes it more likely that a mixed population of 3' fragments exist that do or do not contain the terminal uracils and therefore some can and some cannot bind to La. La siRNA experiments conducted by Dr. Hall, as previously discussed, resulted in an increase in the ratio of the 24 nt ydRNA compared to the 32 nt ydRNA (Hall, 2013), suggesting that the shorter one is derived from the larger one after La protein release.

One discrepancy between our data and that of Rutjes is that they found that a 31nt fragment did not interact with La (Rutjes et al., 1999b). It is possible that this fragment had lost enough of the La binding site that it could no longer stably interact with La. This fragment is unlikely to be derived from hY5, as a hY5-derived 31 nt fragment would still contain three uracils at the 3' end. It is possible that the 31 nt fragment they investigated was the predominant hY5 RNA 5' fragment and it did not bind to La as it was no longer bound to its complementary 3' fragment. More experiments need to be done to examine the protein binding of each of these fragments individually to determine whether they are likely to be functional.

Figure 3.21 shows the secondary structures of human Y RNAs deduced by combining secondary structure predictions and experimental data. When designing the deletion and substitution mutants tested here, the lower or terminal stem was

maintained where possible (this terminal stem is shown at the top of the secondary structures predicted by both RNAfold and Mfold, as the structure has been inverted by the prediction program). The sequence of the larger bulge does not seem to be conserved, but the bulge itself is present in the majority of mammalian Y RNAs, suggesting that it is required for the function of these Y RNAs (Maraia et al., 1996). Determining whether the deletion mutants, along with sub1S and ΔACC, bind to Ro60 in immunoprecipitation experiments should provide more evidence as to the importance of this bulge in Ro60 binding, and therefore potentially strengthen our conclusion that Ro60 is essential for apoptosis-induced cleavage of this RNA.

When considering the presence or absence of cleavage products from the mutants compared to the wild type we must bear in mind that there may be other nucleases involved. We know that these hY5 ydRNAs are present at low levels in untreated cells, but that their levels are dramatically up-regulated after Poly(I:C) induction. We cannot assume that this increase in levels is the result of the increased targeting of the full length Y RNAs by a processing enzyme. It is also equally feasible that, in untreated cells, these small RNAs are produced and are then immediately targeted and rapidly degraded by an exonuclease. If this scenario is correct, then this exonuclease could be down-regulated or re-distributed during apoptosis, resulting in the accumulation of these ydRNAs. It does however, seem far more likely that the majority of the mutants tested are not cleaved because of a block in Ro60 binding, rather than ydRNAs being produced and degraded by a different nuclease to that which processes the wild type RNA.

So far we have intensely speculated on what is required for the processing of Y RNAs, hY5 in particular. We have discussed numerous possible models for how cleavage occurs, but we have not discussed the possible reasons for this cleavage. Y RNAs interact with many proteins in human cells, both blocking and assisting with the interaction of these proteins with RNAs and with other proteins. These Y RNA-protein interactions can alter function and/or localisation within the cell. As a defining function for these Y RNAs alone has not been identified, their potential roles during apoptosis appear endless. Work by our group indicates that these ydRNAs do not act as microRNAs as they do not associate with any of the microRNA processing machinery or the RISC complex (Nicolas et al., 2012), but further analysis of the function of these ydRNAs has been hampered by the fact that we cannot separate them from the full length RNAs they are derived from. Another group of small RNA researchers has potentially found a way of artificially processing full length Y RNAs to leave only the ydRNAs (Meister *et al.* unpublished data), if this is the case then this should aid the investigation of the function of these fragments.

4 HIGH-THROUGHPUT hY5 RNA MUTAGENESIS

4.1 INTRODUCTION

In the last chapter a number of different hY5 mutant plasmids were generated and tested. The expression of each mutant RNA provided different information on the requirements for cleavage to produce ydRNAs. Although each mutant was informative, the process of designing, synthesizing and testing the mutants was lengthy and this limits the usefulness of this method. Northern blotting is a low-throughput method that can take several weeks to produce results and only provides information on a few mutants at a time. NGS is a technique that can be utilized to examine millions of sRNAs at a time and has the potential to be much more informative than northern blotting.

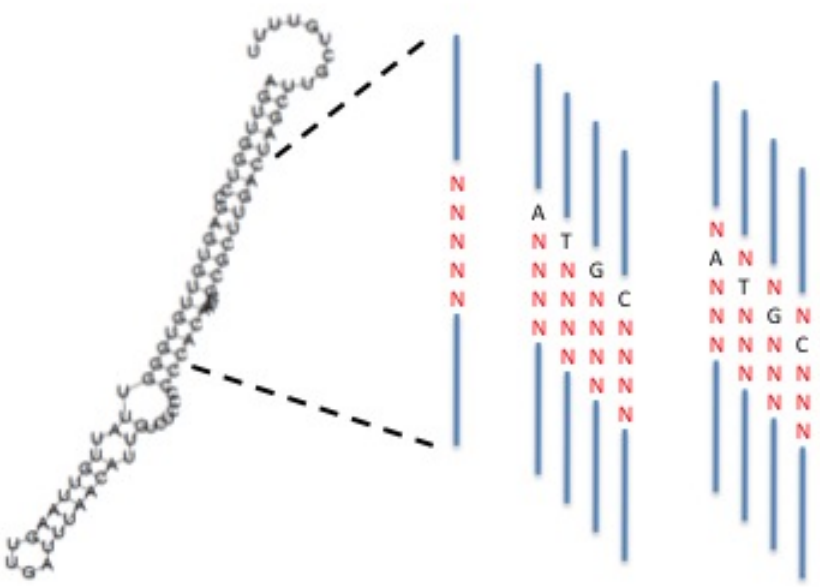


Figure 4.1: Schematic diagram of an individual pool of mutant oligonucleotides. Oligonucleotide primers were synthesized by IDT and incorporated into each mutant plasmid pool.

4.1.1 Methodology

A high-throughput method of mutant generation was designed where a five-nucleotide region of the hY5 RNA was mutated. Instead of testing a few mutants at a time, pools were prepared that contained every possible nucleotide combination at each

of these five positions (Figure 4.1). Three different pools of mutants were produced. Between one pool and the next, the position of the mutated nucleotides was moved sequentially along the DNA molecule (Figure 4.2). This was achieved by the synthesis of long primers containing degenerate nucleotides at each of the five positions. The degenerate nucleotides were hand-mixed by the company undertaking the synthesis. This ensured, as accurately as possible, that each different nucleotide would be present at each of the five positions 25% of the time. Each pool was represented by three different biological replicates, prepared separately right from the PCR stage.

The primer mixes were used as reverse primers and were combined with the hY5 wild type forward primer to PCR amplify the hY5 region, including promoter and terminator sequences (Figure 4.2). Initially, the PCR products were purified and transfected directly into mouse NIH/3T3 cells, but no hY5 RNA could be detected in these cells. Using an alternative method, the PCR products were ligated into the pGEM-T Easy plasmid and transformed into bacteria. If all four possible nucleotides were present in each of the five positions in each pool, this would give 1024 unique plasmid sequences per pool. Dr. Irina Mohorianu (a statistician linked to the Dalmay laboratory) calculated that, in order for each of the 1024 sequences to be represented within a pool at a probability of 99%, after transformation we would need more than 4850 positive bacterial colonies to be harvested per pool. The number of LB agar plates prepared for each replicate ensured that over this number of positive colonies were harvested, often the number of colonies harvested was close to double the number required. The plasmids were then purified and transfected into NIH/3T3 cells according to the conditions used for single mutant plasmid transfection. The cells were treated with STS for 4 hours prior to harvesting to increase the rate of Y RNA cleavage. Upon harvesting, total RNA was extracted and used to create a number of different types of library (Figure 4.3).

A novel method of generating cDNA libraries from sRNAs for NGS has been developed, in the Dalmay laboratory, by Dr. Karim Sorefan (Sorefan et al., 2012). This method has the advantage over the current Illumina method, as it overcomes the inherent preference of the ligase enzyme for particular adapter-RNA-adapter sequences (Sorefan et al., 2012). If the abundance of different sequences in a library are to be accurately compared to each other, overcoming this preference is vital. This novel sequencing method, here referred to as the 'HD' method, was integral in the sRNA libraries prepared in this chapter. sRNA libraries were prepared to determine which of the full length mutant hY5 sequences expressed in mutant cells underwent apoptosis-dependent cleavage to produce ydRNAs. In the low-level mutagenesis experiments in Chapter 3, the full length mutant RNAs were not always expressed at the same levels as the wild type full length hY5 RNA (see Figures 3.16 & 3.14).

Full length Δ1S was consistently lower in concentration than the wild type full length hY5, even in different experiments using different plasmid preparations. It seems likely that the mutation within the Δ1S plasmid sequence adversely affected levels of the full length Δ1S RNA, resulting in a lower rate of transcription or a lower stability after transcription. For the mutant pools, if a mutant sequence was not detected in the sRNA population, this could be as a result of a lack of cleavage, but it may also be explained by low levels of the full-length RNA that mutant ydRNA was derived from. Creating full-length RNA libraries from the same total RNA samples used for the sRNA libraries should enable us to account for any differences in the concentrations of each full-length mutant RNA, providing a potential explanation for some of the sRNA sequences differentially expressed when compared to the wild type sequence. For this reason, total RNA was also used to prepare libraries where the full length hY5 mutants were amplified.

By transfecting cells with PCR products, the levels of one mutant construct being higher than the other would have been minimized. As this method of RNA expression was unsuccessful, cloning of the PCR products into plasmids and transfection of the resulting plasmids was the only option. Due to the efficiency of ligation and the differences in the grown of the bacterial cells transformed with the resulting positive plasmids, inherent increases in certain sequences would be introduced during the cloning process. To take into account these differences and correct for them, a small amount of each plasmid pool was retained and plasmid bias libraries were prepared from these samples. This would enable us to determine which mutant sequences were over- or under-represented in the plasmid pools.

4.1.2 Library preparation

Briefly, 2 μg of total RNA was used in 3' 'HD' adapter ligations (Sorefan et al., 2012) (Figure 4.4). These samples were separated using denaturing PAGE and the region containing the 30mers, including the ydRNAs, was excised and purified. The resulting samples were then ligated to the 5' Illumina small RNA adapter and used as a template for cDNA synthesis. cDNA was then amplified using PCR and separated by PAGE. The 30mer inset-containing region was excised and purified from the gel. The same sized band was gel purified for a second time to eliminate adapter-adapter sequences, and sent to BaseClear for sequencing.

STS treatment, while enhancing the rate of Y RNA cleavage, also increases the cleavage rates of many other ncRNA, if this occurred during these experiments it would dilute the hY5 mutant sequences in the NGS dataset. To overcome this problem, an enrichment was included in the PCR amplification stage of the library preparation. The 5' end of the wild type ydRNA of interest contains 5 C residues.

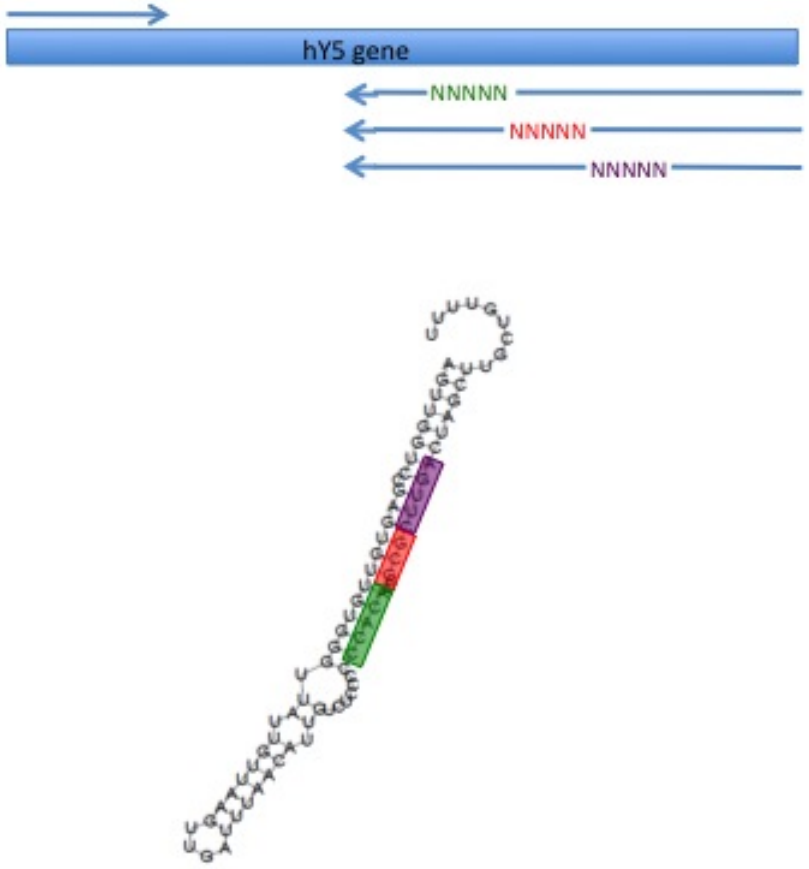


Figure 4.2: Schematic diagram detailing the initial PCR of the hY5 sequence. The wild type hY5 forward primer and the pool of reverse primers were used in the amplification.

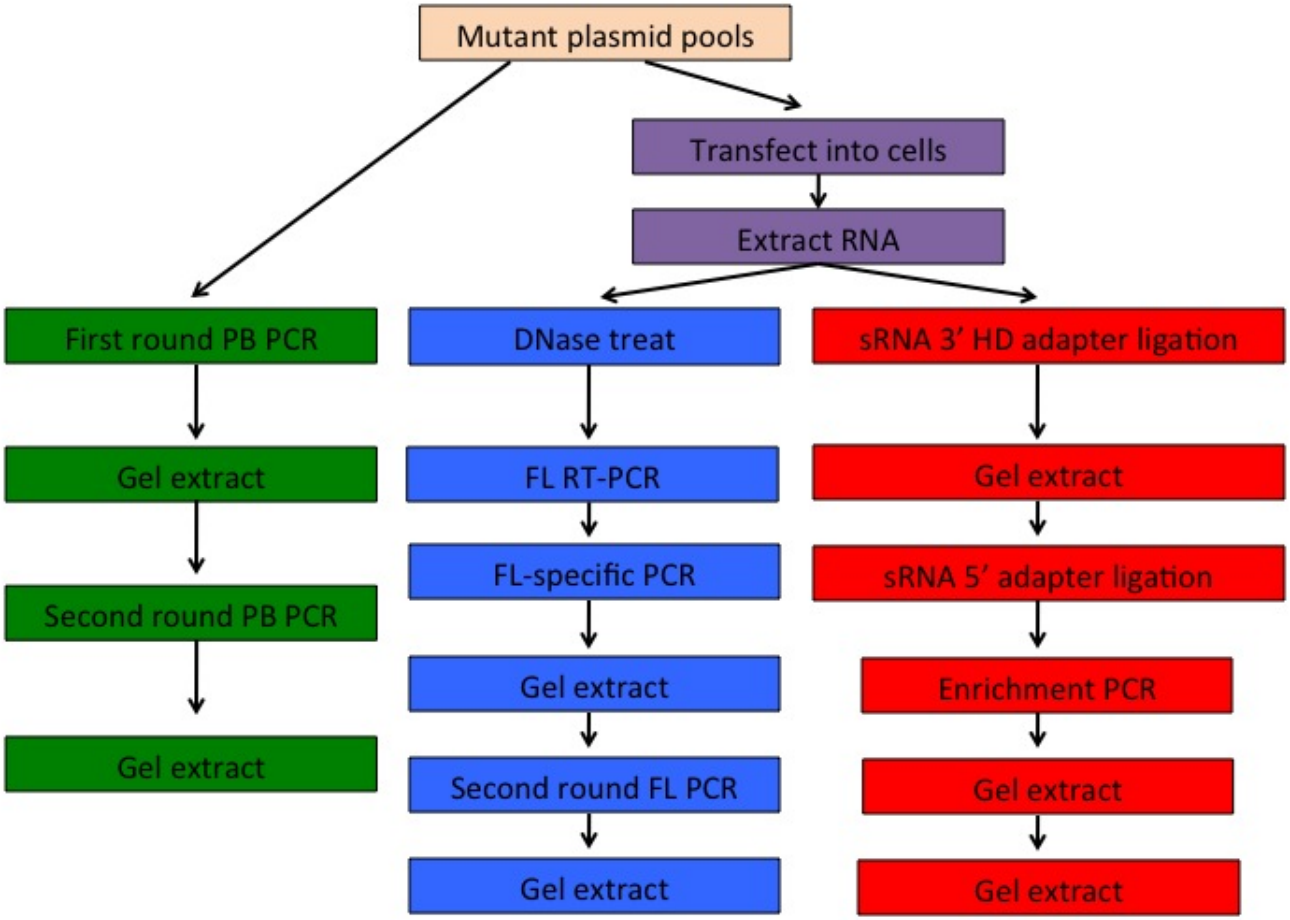


Figure 4.3: Overview of the hY5 high-throughput mutagenesis workflow. FL - full length, PB - plasmid bias.

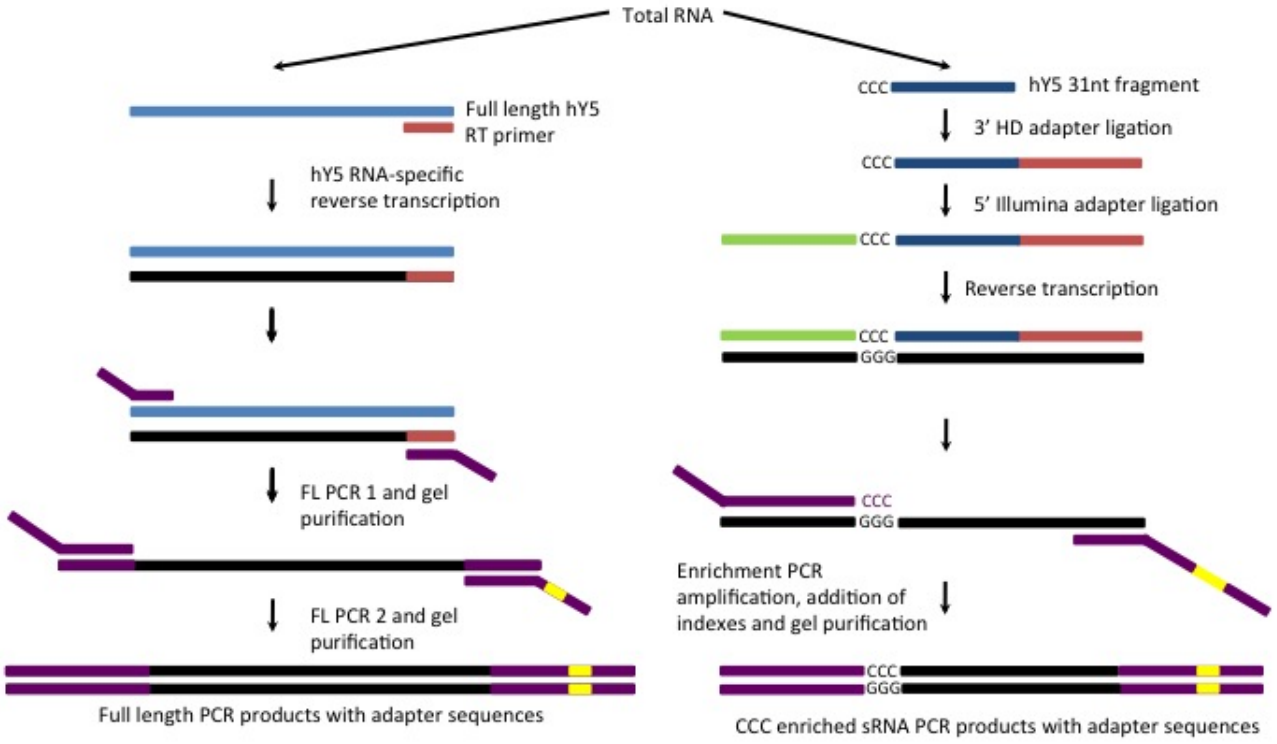


Figure 4.4: Schematic diagram of sRNA and full length Y RNA cDNA library generation.

The positions of the mutated nucleotides in each of the three different pools were chosen so that the four 5'-most Cs were kept intact. To increase the amplification of the ydRNAs, the Illumina RP1 forward primer, used at the PCR stage, was extended to include five extra nucleotides of the 5' adapter sequence and the 3 Cs from the 5' end of the ydRNAs. The amplification of other sequences with 3 Cs at the 5' end of a similar size was expected, and also necessary to introduce enough heterogeneity into the samples for the products to spread adequately across the glass plate during sequencing.

For full length library preparation, the total RNA sample was DNase-treated to remove contaminating DNA (Figure 4.4). An RNA-specific primer was used to reverse-transcribe only hY5-related sequences. These cDNAs were then PCR amplified using the FL round 1 primers, which were designed only to target the full length sequences and not any Y5 degradation products. In this first round PCR some of the sequences required for NGS were added to both ends of the cDNA. Products of the correct size were gel extracted and, in a final round of PCR, the standard Illumina Truseq sRNA kit primers were used to add the rest of the sequences required for NGS onto the ends of each of the products. The samples were then gel extracted for a final time.

As for the full length libraries, plasmid bias library preparation entailed two rounds of PCR amplification to add sequences required for NGS sequencing to the PCR products. Custom primers were used for the first round PCR and the primers from the Illumina Truseq sRNA kit were used for the second round PCR. Samples were gel extracted between PCRs and following the final PCR reaction to purify and concentrate the products.

4.1.3 Aims

The first aim of this chapter was to see if the 'HD' NGS technique could be utilized to generate a more high-throughput method of hY5 mutant generation and analysis. The second and ultimate aim being to gain more information on hY5 cleavage requirements. Our hypothesis is that the wild type RNA has the optimal level of cleavage and that the mutants where no cleavage occurs have sequences which more dramatically alter the structure of the RNA within a region important for the binding of proteins or nucleic acids important in the cleavage process. It will only be a combined analysis of all of the mutants together that will allow us to narrow down the region important for cleavage and to give us some indication of RNA structure or particular sequence motifs which may be important for this cleavage to occur.

4.2 Results

For amplification of the hY5 sequence to produce the plasmid pools, the PCR was initially tested on a truncated version of the pWT plasmid, which was missing the last 30 nucleotides of the hY5 RNA-encoding region and the downstream regulatory sequence. Unfortunately, using this truncated version of the hY5 plasmid in the PCR amplification did not produce PCR products at the concentration required for ligation, presumably as a result of limited annealing of the primer to this truncated plasmid. For this reason the pWT plasmid was used as a PCR template.

For each plasmid pool, three replicates were prepared and the appropriate number of plates were used to produce at least 4850 different white colonies of transformed DH5α or XL2 Blue *E. Coli* cells for each replicate. Instead of selecting positive colonies from each bacterial plate, the positive colonies on each plate were counted and all colonies on a plate were harvested together, including the negative colonies. A number of plates were pooled together to produce each replicate. Several plates were processed in one midiprep reaction, before concentrating by ethanol precipitation, further pooling of samples, and transfection into NIH/3T3 cells. 50 ng of each plasmid pool was set aside and used to assess plasmid bias. For each different plasmid pool, 5 μg of each replicate plasmid sample was transfected into a different T75cm^2 flask. The amount of plasmid transfected was reduced in relation to the single mutagenesis experiments as the plasmid pools appeared to have a higher level of toxicity to the cells than when a single plasmid alone was transfected, possibly as a result of the mixture of plasmids present within each pool.

24 hours after transfection, cells were treated with STS for 4 hours and harvested in TRIreagent. After RNA extraction each of the samples was split into two tubes and the RNA in one of the tubes was DNase treated to remove contaminating DNA. 3 μg of this RNA was used for northern blot analysis, to gauge the success of the transfections and to determine the level of the full length Y5 mutants and ydRNAs in each sample, compared to controls of known concentration. RNA from the cells transfected with the first replicate of each pool was analyzed (Figure 4.5). All plasmid pool samples showed clear expression of full length Y RNA and ydRNA products. The sample from the first replicate of pool 3 had the highest expression levels of full length Y5 RNA. Interestingly, the sizes of the ydRNAs produced from each of the different pools appears to vary, with pool 1 producing the longest ydRNAs. The ydRNAs from samples transfected with pools 2 and 3 may also be slightly different in size, with pool 3 producing slightly shorter ydRNAs than pool 2.

All of the full length RNAs are as abundant, if not more abundant, than the weakest control oligonucleotide, which was at 5 femtomoles per well. As 3 μg of total RNA from each of the plasmid pool-transfected samples was loaded into each

well, the full length plasmids are present at approximately 1.67 femtomoles per μg, which is over 1 million molecules per μg.

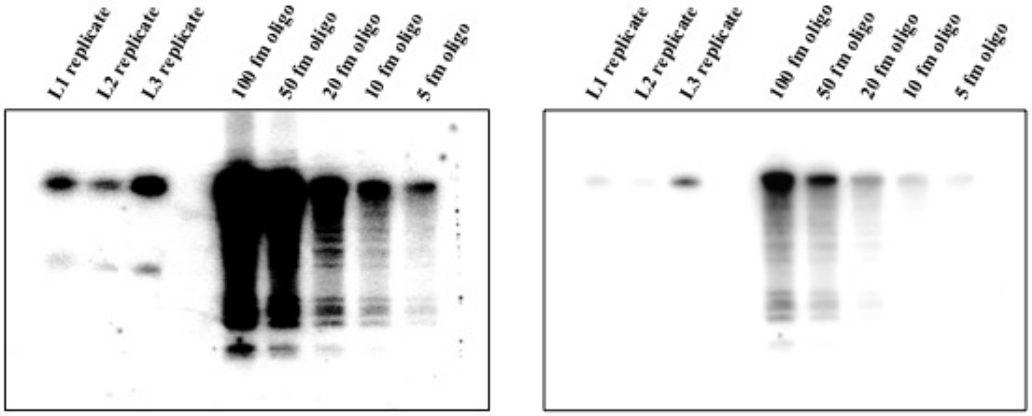


Figure 4.5: Mutant pool expression in NIH/3T3 pools.
(A) 3 μg of total RNA prepared from cells transfected with the first replicate of each of the mutant plasmid pools was used for northern blot analysis. In lanes 5-9 the Δ2S R oligo was loaded at different concentrations, to act as a control for size and concentration. The membrane was probed with the 12 nt hY5-specific LNA probe, which should detect all mutant Y5 RNAs.

The first tube of each total RNA sample was used for full length library preparation. PCR amplification was achieved by two rounds of PCR and PAGE purification (Figure 4.4). A clear band of the correct size was seen after each PCR reaction and was intense enough to be easily cut from the gel and precipitated (see Figure 4.6). After the final gel extraction, one tenth of the volume of each sample was quantified by PAGE and all were deemed to be of high enough concentration to be sequenced by NGS.

The second tubes were used to prepare sRNA libraries. Samples were ligated to a 3' and then a 5' adapter, before cDNA synthesis and PCR amplification. A clear band was visible on the gel, below the 100 nucleotide marker. This band corresponds to the 30-35 nucleotide sRNAs known to be present after Poly(I:C) treatment [Figure 4.7 (A)]. A band was also visible at 24 nucleotides, as was a band lower down on the gel, which is most likely made up of adapter-adapter sequences. The 30mer band was cut from the gel and purified according to the Illumina protocol. A second gel purification showed a stronger 32 nucleotide product and weaker 24 nucleotide product [Figure 4.7 (B)]. It also showed that the adapter-adapter products were dramatically reduced in these samples compared to those after the first gel extraction. In a final test, and to enable calculation of the approximate volume of each sample to pool with the other sRNA library samples, a third PAGE gel was run containing one tenth of the total volume of sample [Figure 4.7 (C)]. On completing

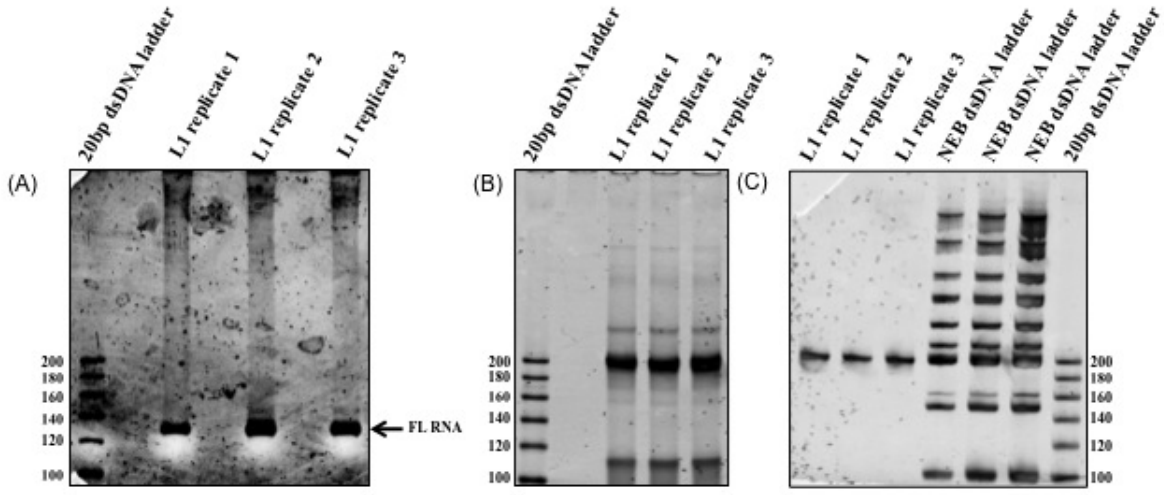


Figure 4.6: cDNA library preparation of full length hY5 mutant RNAs. Libraries were prepared as in [Figure 4.4 (A)]. (A) Gel separation after first round PCR amplification (B) Gel separation after second round PCR amplification (C) Final size confirmation gel prior to sample processing. FL - full length.

this analysis, the sRNA library samples were deemed to be of high enough quality and quantity to be analyzed using the Illumina truseq NGS method.

Plasmid bias sample libraries were prepared using 50 ng of each replicate plasmid pool. Samples were separated using PAGE after each round of PCR and a final PAGE gel was used to confirm that they were the correct size and concentration for sequencing (Figure 4.8). A strong band of the correct size was seen on all of the PAGE gels and the samples were deemed suitable for NGS. Due to a complication of the requirements for heterogenity in sequencing samples, there was a delay in the sequencing of the plasmid bias and full length control libraries, so these have not yet been analyzed.

4.2.1 sRNA library results

The 'HD' adapter sequence was removed from the 3' end of each sequence read by removing the 8 nucleotide standard adapter sequence and the preceeding by 4 degenerate nucleotides. Correlation and MA plots were used to compare the 3 replicate samples transfected with each plasmid pool (Figures 4.9, 4.10, & 4.11) and the outlier replicate was removed from each analysis. Figure 4.12 shows the correlation plots comparing the two remaining replicates for each mutant pool. The sequences from the remaining replicates were mapped to the mouse genome and to

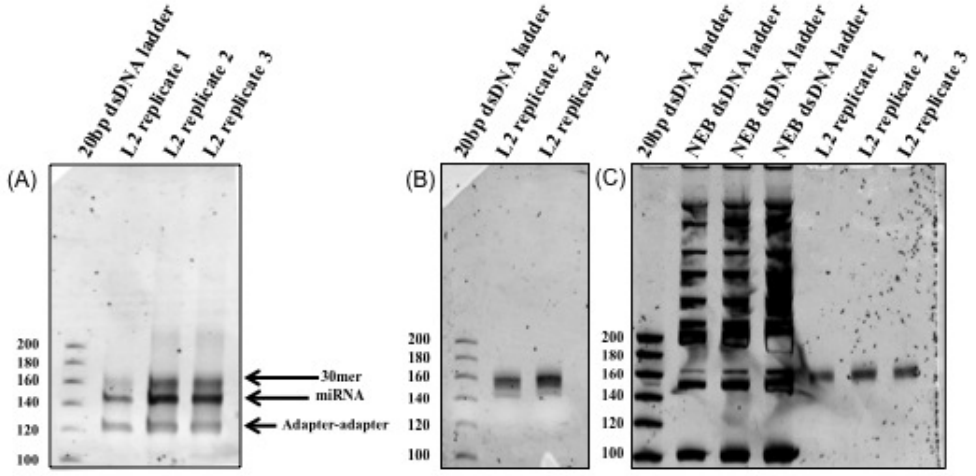


Figure 4.7: cDNA library preparation of mutant ydRNAs.
Libraries were prepared as in [Figure 4.4 (B)]. (A) Gel separation after enrichment PCR amplification (B) Gel separation after first gel extraction (C) Final size confirmation gel prior to sample processing.

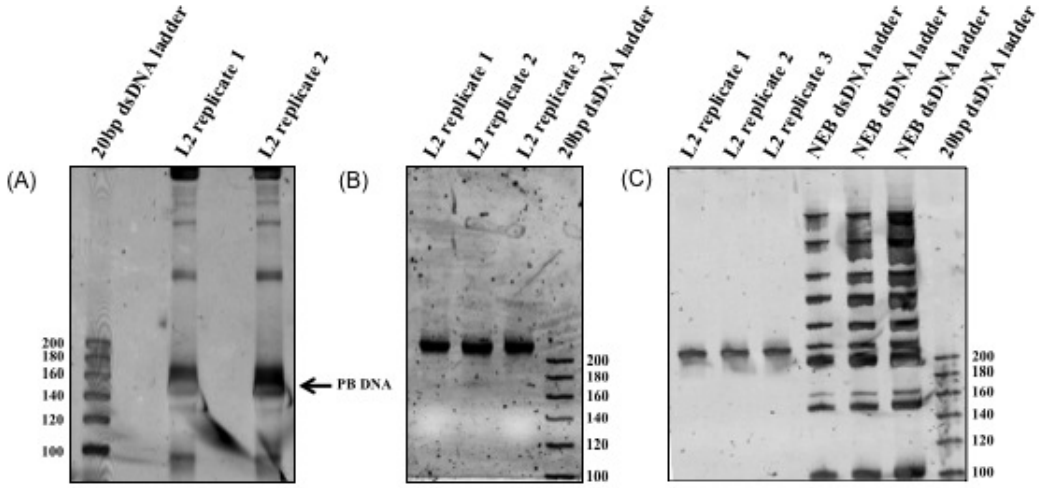


Figure 4.8: Plasmid bias library preparation of mutant plasmid replicates.
50 ng of plasmid DNA was amplified in two rounds of PCR and gel extraction. A) Gel separation after first round PCR amplification (B) Gel separation after second round PCR amplification (C) Final size confirmation gel prior to sample processing. PB - plasmid bias.

a set of hY5 mutant reference sequences. These mutant reference sequences were created by removing the region targeted for mutagenesis from the wild type hY5 sequence, and replacing it with all of the possible mutations within that region. A set of reference sequences was created for each separate mutant pool.

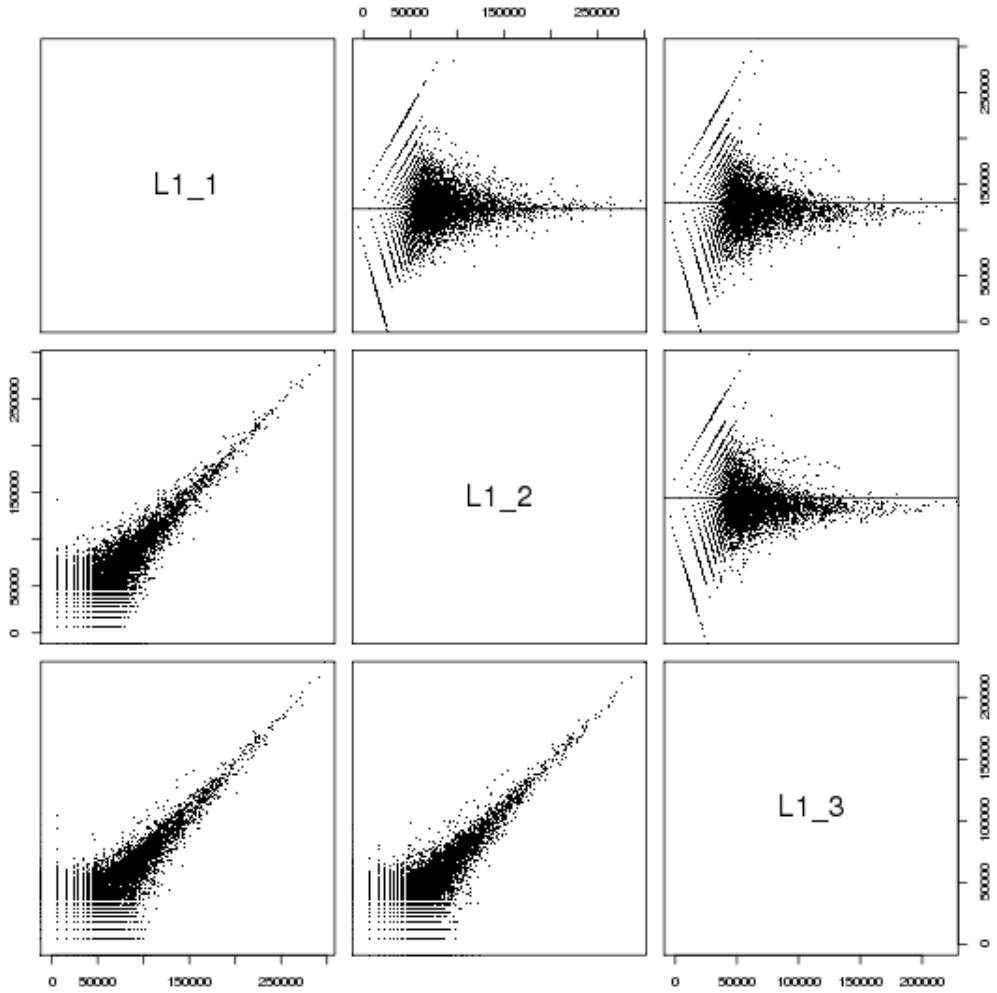


Figure 4.9: Correlation and MA plots for plasmid pool 1 replicates.
These plots show that the third replicate is the least similar of the three, so this replicate was removed from the analysis (Plots prepared by Matthew Beckers).

Table 4.1 details the redundant reads of each replicate mapping to the mouse genome and those mapping to the reference hY5 sequence set. Table 4.2 shows the non-redundant reads of each replicate mapping to the mouse genome and to the hY5 sequences. 12-25% of the reads map to the mouse genome and between 3-6% map to hY5. The removed replicates are MP1 R3, MP2 R1, and MP3 R2. Tables 4.1 and 4.2 show that these replicates are the least similar to the other two replicates for that library.

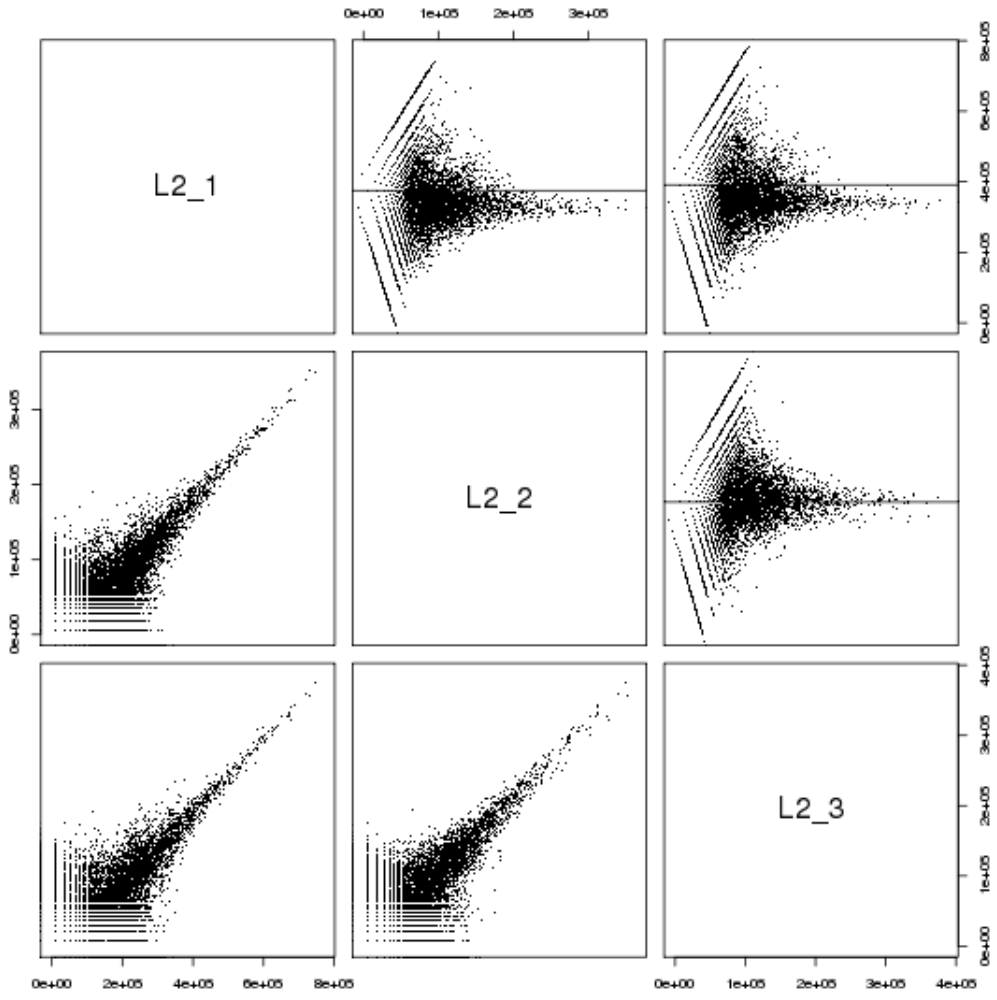


Figure 4.10: Correlation and MA plots for plasmid pool 2 replicates.
These plots show that the first replicate is the least similar of the three, so this replicate was removed from the analysis (Plots prepared by Matthew Beckers).

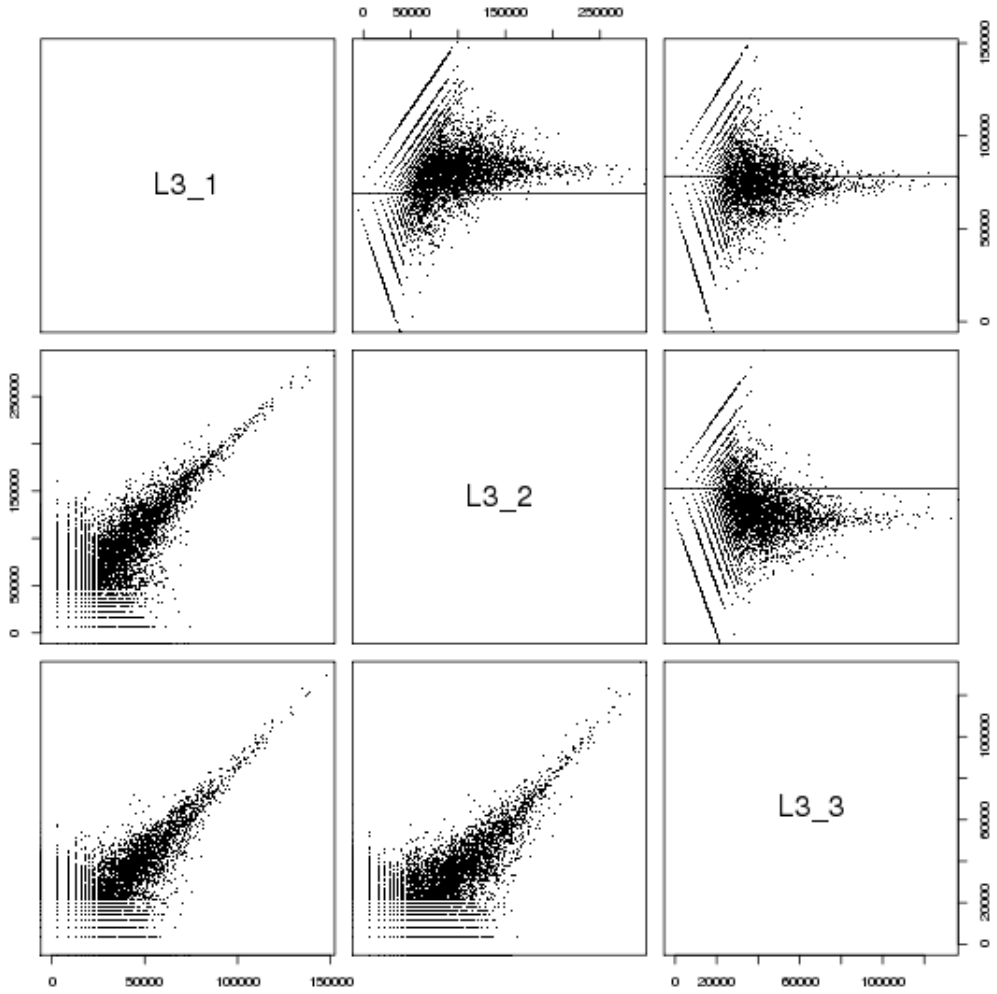


Figure 4.11: Correlation and MA plots for plasmid pool 3 replicates.
These plots show that the second replicate is the least similar of the three, so this replicate was removed from the analysis (Plots prepared by Matthew Beckers).

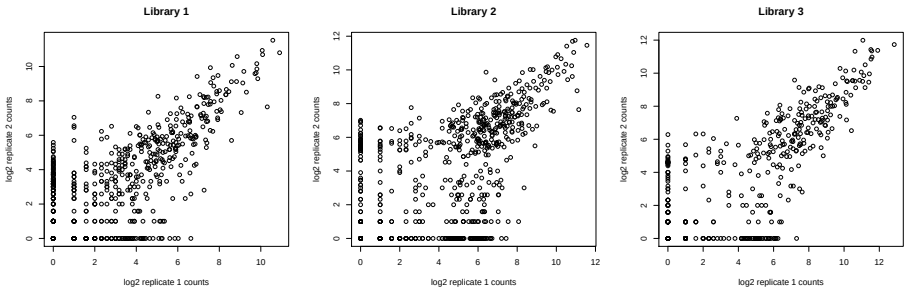


Figure 4.12: Correlation plots of included replicates.
Each of the two replicates for each library was compared against the other to determine similarity (Plots prepared by Matthew Beckers).

MP and rep	Total Reads	hY5	% hY5	Mouse genome	% Mouse genome
MP1 R1	2823676	43169	1.53	1629839	57.72
MP1 R2	3002246	52761	1.76	1764092	58.76
MP1 R3	2182559	32941	1.51	1289863	59.1
MP2 R1	4884128	73091	1.5	2936298	60.12
MP2 R2	2525008	84626	3.35	1513646	59.95
MP2 R3	24444252	83252	3.41	1380416	56.48
MP3 R1	1367354	108133	7.91	749813	54.84
MP3 R2	3546754	398831	11.24	1782200	50.25
MP3 R3	1061010	65785	6.2	575316	54.22

Table 4.1: Redundant reads mapping to hY5 and the mouse genome.
This table summarizes the redundant reads of each replicate (R) and each mutant pool (MP) mapping to the mouse genome and mapping to the hY5 mutant reference sequences (Data provided by Matthew Beckers).

MP and rep	Total Reads	hY5	% hY5	Mouse genome	% Mouse genome
MP1 R1	78087	2702	3.46	19768	25.32
MP1 R2	77896	2655	3.41	20021	25.7
MP1 R3	60648	2103	3.47	15149	24.98
MP2 R1	57217	1450	2.53	10271	17.95
MP2 R2	50032	2294	4.59	10230	20.45
MP2 R3	39765	1738	4.37	7025	17.67
MP3 R1	31809	2036	6.4	4997	15.71
MP3 R2	54325	2824	5.2	6996	12.88
MP3 R3	28082	1698	6.05	4754	16.93

Table 4.2: Non-redundant reads mapping to hY5 and the mouse genome.
This table summarizes the non-redundant reads of each replicate (R) and each mutant pool (MP) mapping to the mouse genome and mapping to the hY5 mutant reference sequences (Data provided by Matthew Beckers).

Figure 4.13 shows both the size and abundance of the reads mapping to the mouse genome. Notably there is a peak at 19 nucleotides, along with the peak of around 31 nucleotides. This suggests that mouse cells may also have an abundant 30mer class of sRNAs after apoptosis treatment. On investigation of the mouse-derived 19 nucleotide sequences present in this dataset, this peak was found to comprise mainly of two highly abundant sequences mapping to three places in the genome. Both sequences map to microRNAs, one of which is encoded in two different locations in the genome. The significance of these sequences in dataset is currently unclear.

It is surprising that such a high 19 nucleotide peak would be present in the data, as only a region of approximately 32 nucleotides was cut from the gel during library preparation. These sequences are likely to have a much higher read count in the total RNA samples than that seen in this dataset, where the majority of sRNAs of this size have been removed. Mapping of the 19 nucleotide reads to both of these miRNAs shows that the boundaries of the sequencing reads are distinct and are the same in all of the libraries (Figure 4.14).

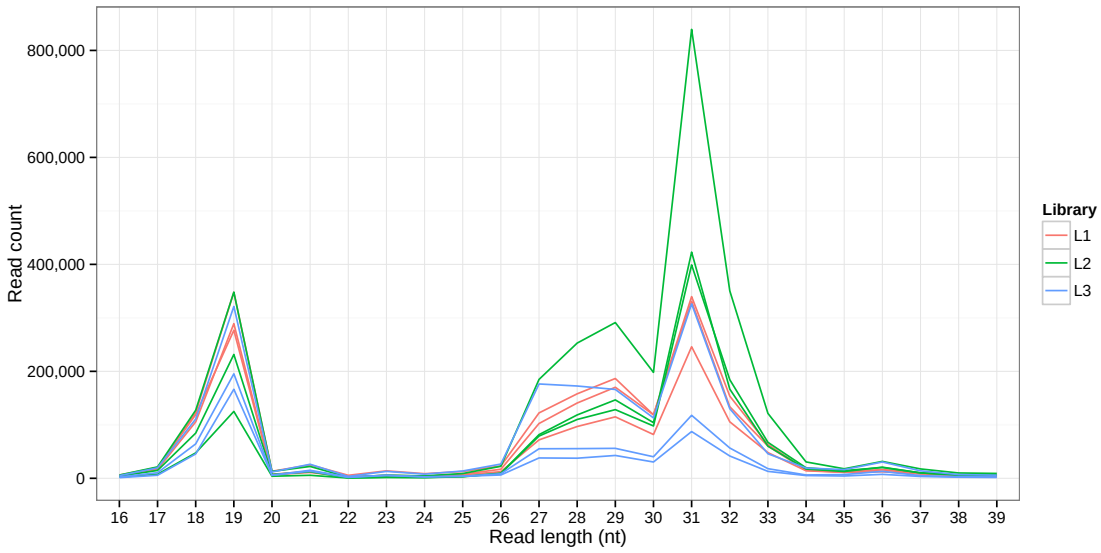


Figure 4.13: Length and abundance of redundant reads mapping to the mouse genome.
The different colors represent the different plasmid pools (Graph prepared by Matthew Beckers).

Figure 4.15 displays the hY5 mapping reads in the dataset. The graph shows that the size of the mutant pool 1 sRNA reads is mainly 32 nucleotides, but that for mutant pools 2 and 3 the sRNAs are mainly 30/31 nucleotides. It should be noted that the reads displayed are those as a direct result of sequencing, so the actual ydRNAs are likely to be 3 nucleotides longer than the reads displayed here.

As well as examining the ydRNAs by size, they were also plotted by start position (Figure 4.16). The distribution of peaks for mutant pools 2 and 3 is quite different

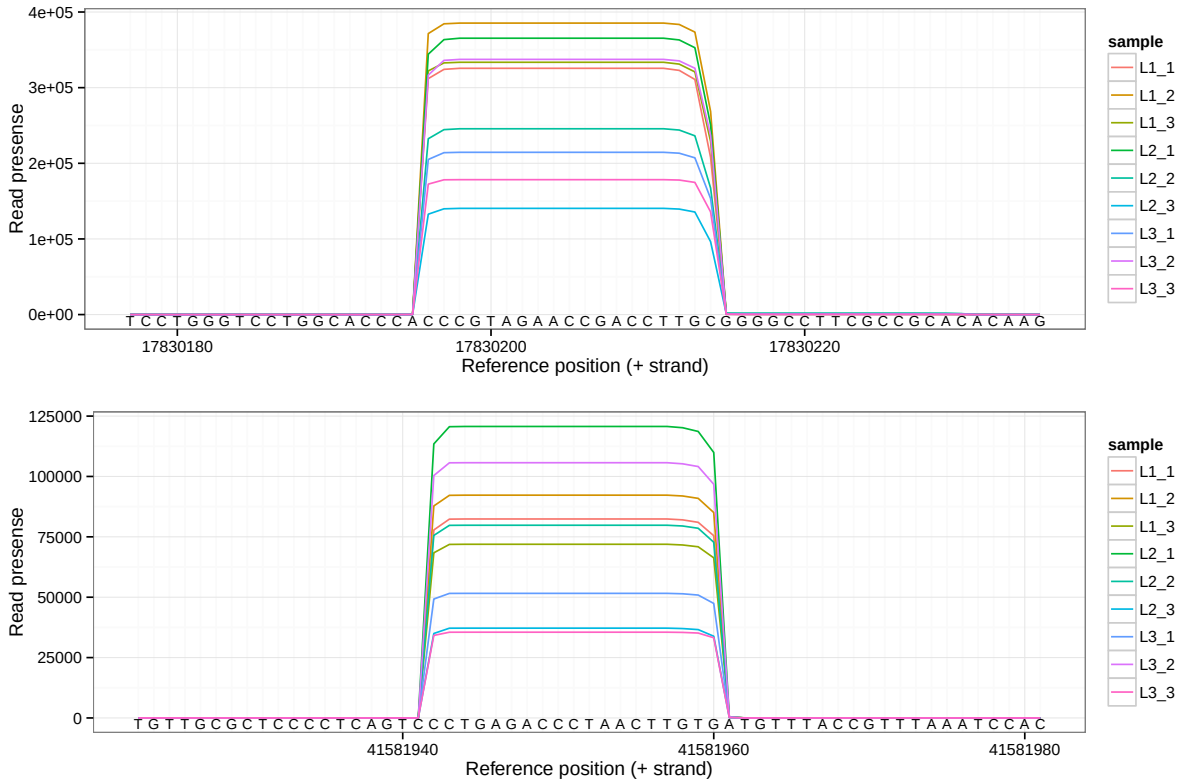


Figure 4.14: Mapping the two highly abundant miRNAs to the mouse genome. miR-99b accounted for 72.1% of the 19mer reads, while miR-125b accounted for 25.5% of the 19mer reads. (Graph prepared by Matthew Beckers).

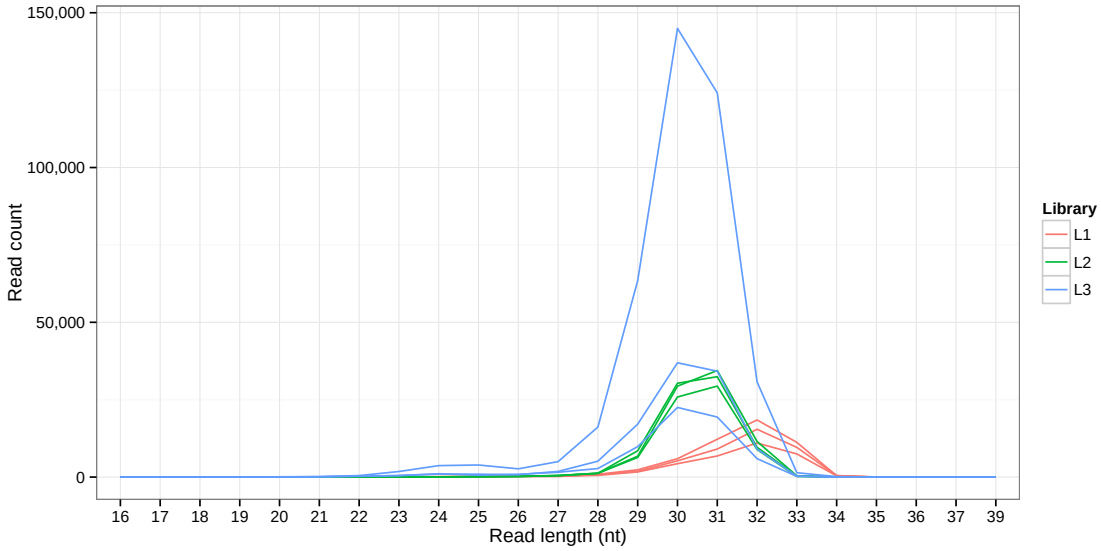


Figure 4.15: The length and abundance of redundant reads mapping to hY5. Reads were mapped to the hY5 mutant reference sequences. The different colors represent the different plasmid pools (Graph prepared by Matthew Beckers).

to the distribution for pool 1. For pools 2 and 3 the vast majority of sequences start at two discrete positions, with very little variation. This suggests that the majority of the cleavages to produce the reads detected from these pools occur between T49 and C50 or C50 and C51. This mirrors the main cleavage point seen in endogenous hY5 after Poly(I:C) treatment of human MCF7 cells, which occurs between C50 and C51. In the mutant pool 1 sRNA libraries, the start point of the sequencing reads is less well defined, with the majority of the reads spread over 3 positions. The highest peak is at position C51, which indicates that the majority of the cleavage events occurs between C48 and T49.

Rank	Rep 1	Rep 2
1	ACCGC	ACGTC
2	ACGTC	CACAC
3	ACTTC	ACCGC
4	CCACG	ATGGC
5	ATGGC	ACACC
6	CACAC	ACTGC

Mutant pool 1

Rank	Rep 1	Rep 2
1	CCGTT	ACGTC
2	ATCGT	CCGTC
3	TCGTC	CCGTT
4	ACGTC	CTGTC
5	ACTCT	ACTCT
6	CCGTC	CGTTG

Mutant pool 2

Rank	Rep 1	Rep 2
1	GTTTG	CGTCG
2	TGTAG	GTTTG
3	GATTG	GTTCG
4	TGTCG	TGTAG
5	CGTTG	TGTCG
6	GTTCG	GTTAG

Mutant pool 3

Table 4.3: Mutations sorted by abundance for each retained replicate (Rep).

The mutant ydRNA sequences were ranked in order of abundance for both replicates from each mutant pool. Table 4.3 displays the top 6 reads mapping to each replicate of a mutant pool, while Table 4.4 displays the top 6 reads mapping to both replicates combined. These tables show that the most abundant ydRNAs were detected at high levels in this dataset, highlighting the power of this technique when combined with NGS. When the top reads for each mutant pool are closely examined and the probabilities predicted and displayed in sequence logos, a preference for certain nucleotides can be seen at each position. This preference does not seem to correlate with the nucleotide present at that position in the wild type hY5 sequence. In the second mutant pool samples there is less of a preference at the first nucleotide, but for the rest of the sequence there is a preference for CGTC (Figure 4.17), while

in the third mutant pool samples there is a strong preference for (Figure 4.18) the sequence CTTTG. For mutant pool 1 the sequence reads were divided by start position and abundance plots and sequence logos were generated (Figures 4.21 to 4.26). These plots show that there is a different pattern of cleavage depending upon the mutant sequences present at each position (this will be explained in further detail later).

Mutation	Rep 1	Rep 2	Total
ACGTC	699	700	1399
ACCGC	700	698	1398
CACAC	695	699	1394
ATGGC	696	697	1393
ACTGC	692	695	1387
ACCTC	693	691	1384

Mutant pool 1

Mutation	Rep 1	Rep 2	Total
CCGTT	700	698	1398
ACGTC	697	700	1397
CCGTC	695	699	1394
ACTCT	696	696	1392
CTGTC	691	697	1388
CGATC	693	691	1384

Mutant pool 2

Mutation	Rep 1	Rep 2	Total
GTTTG	550	549	1099
TGTAG	549	547	1096
GTTCG	545	548	1093
TGTCG	547	546	1093
GATTG	548	544	1092
CGTCG	539	550	1089

Mutant pool 3

Table 4.4: Ranked abundant mutations for each pool.
The most abundant mutations for each pool after the rankings for each replicate have been combined.

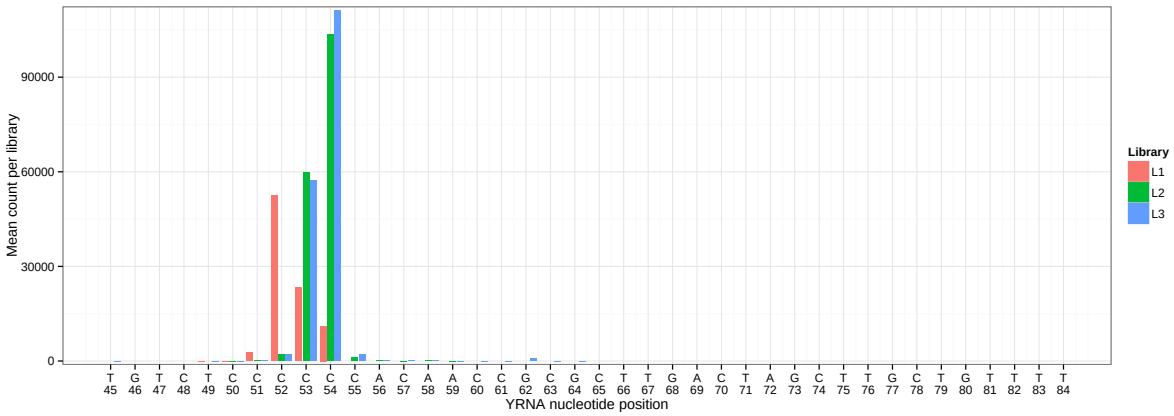
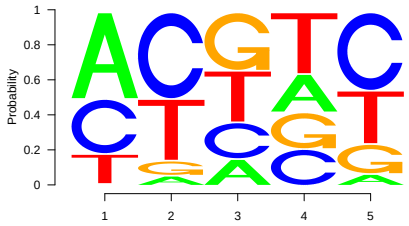
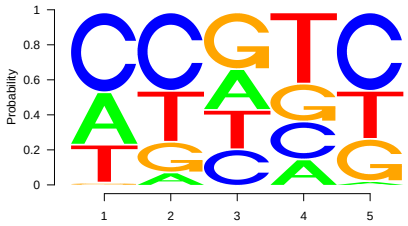


Figure 4.16: The abundance and start position of all of the ydRNAs mapping to Y5. The mean count at each position for each plasmid pool is displayed. The different colors represent the different plasmid pools (Graph prepared by Matthew Beckers).

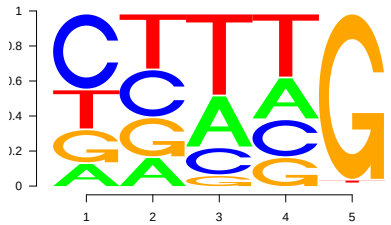


Mutant pool 2 start position 53

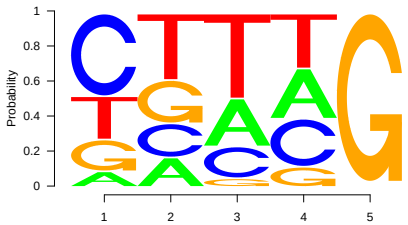


Mutant pool 2 start position 54

Figure 4.17: Sequence logos for mutant pool 2.
The sequence logos showing the probability of finding each nucleotide at each position within the mutation site for pool 2. This was calculated for each read start position using all of the abundant sequence reads, after both replicates had been combined (generated by Matthew Beckers).



Mutant pool 3 start position 53



Mutant pool 3 start position 54

Figure 4.18: Sequence logos for mutant pool 3.
The sequence logos showing the probability of finding each nucleotide at each position within the mutation site for pool 3. This was calculated for each read start position using all of the abundant sequence reads, after both replicates had been combined (generated by Matthew Beckers).

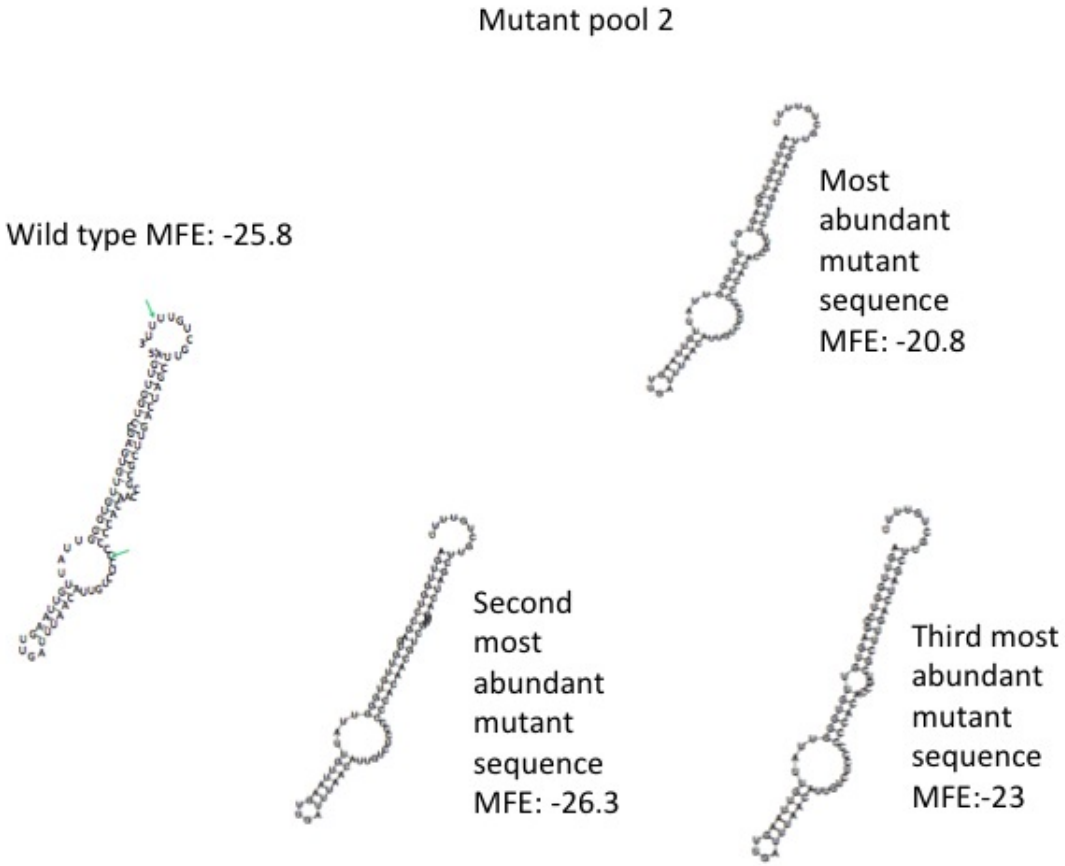


Figure 4.19: RNA structures for mutants yielding ydRNAs with the most abundant motifs in mutant pool 2 at C53 & C54.
The secondary structures of the wild type hY5 RNA and the full length mutant RNAs in pool 2 that yield ydRNAs with reads starting at positions C53 and C54 with the three most abundant mutations (as predicted by RNAfold) of the wild type hY5 RNA, and the three most abundant mutant sequences at start position C53 and C54. Green arrows represent the cleavage positions that produce the 32mer.

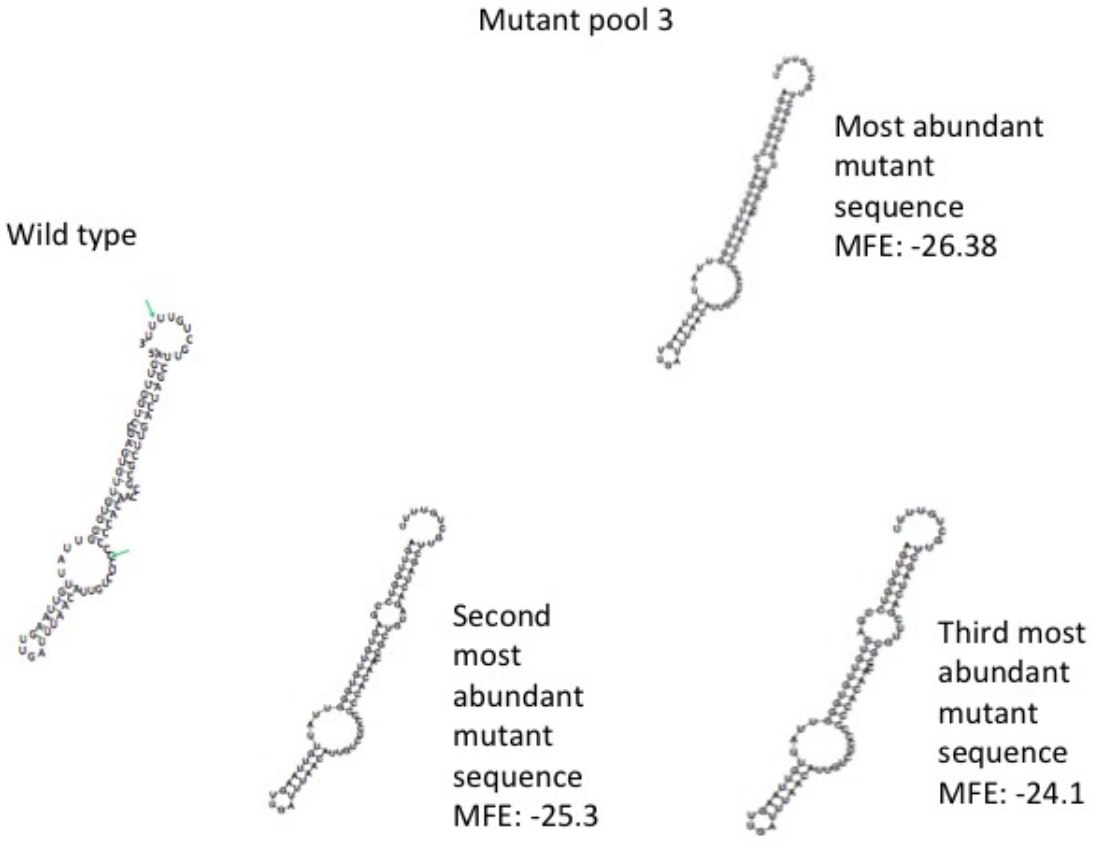


Figure 4.20: RNA structures for mutants yielding ydRNAs with the most abundant motifs in mutant pool 3 at C53 & C54.
The secondary structures of the wild type hY5 RNA and the full length mutant RNAs in pool 3 that yield ydRNAs with reads starting at positions C53 and C54 with the three most abundant mutations (as predicted by RNAfold) of the wild type hY5 RNA, and the three most abundant mutant sequences at start position C53 and C54. Green arrows represent the cleavage positions that produce the 32mer.

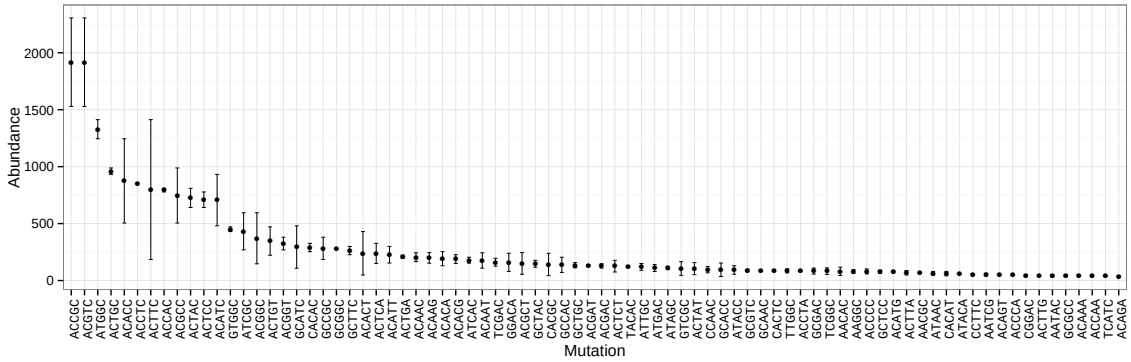


Figure 4.21: The ordered abundance of the mutant pool 1 ydRNAs starting at C52 (Graph prepared by Matthew Beckers).

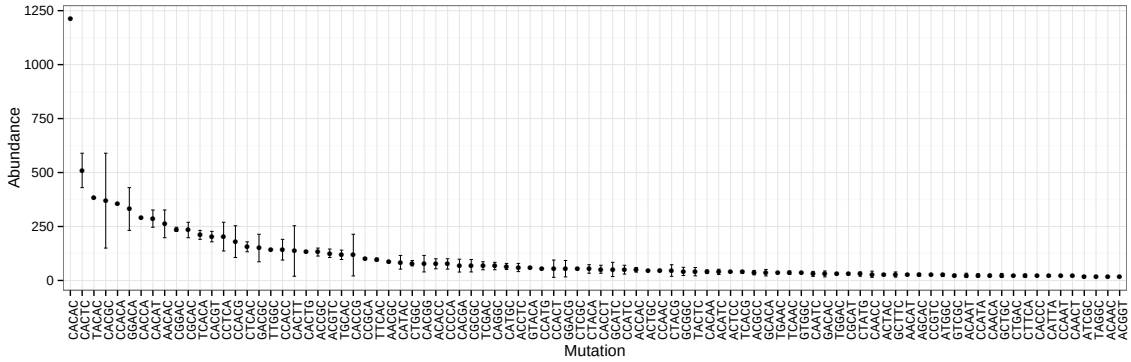


Figure 4.22: The ordered abundance of the mutant pool 1 ydRNAs starting at C53 (Graph prepared by Matthew Beckers).

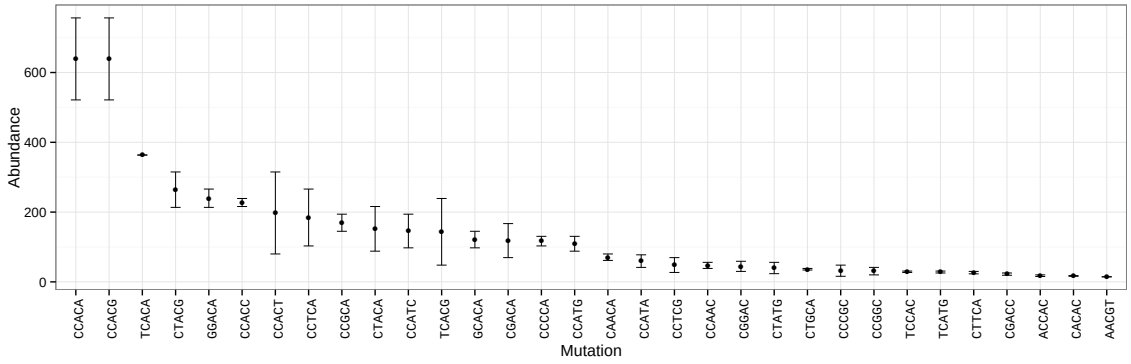


Figure 4.23: The ordered abundance of the mutant pool 1 ydRNAs starting at C54 (Graph prepared by Matthew Beckers).

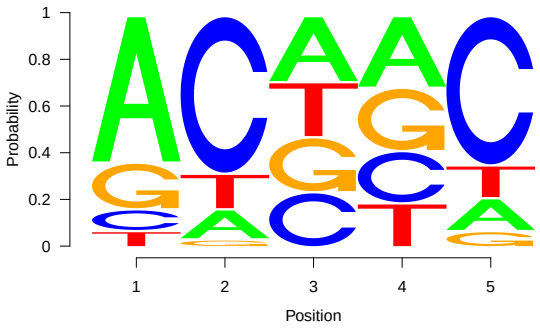


Figure 4.24: Sequence logos for mutant pool 1.
The sequence logos showing the probability of finding each nucleotide at each position within the mutation site for pool 1. This was calculated for read start position C52 all of the abundant sequence reads, after both replicates had been combined (generated by Matthew Beckers).

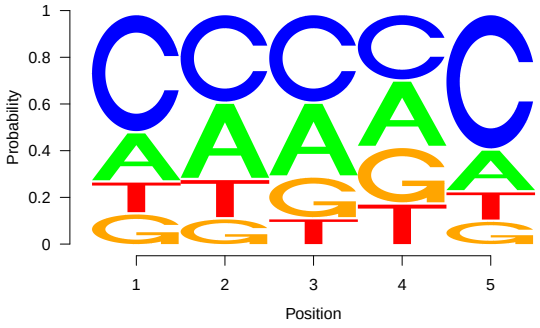


Figure 4.25: Sequence logos for mutant pool 1.
The sequence logos showing the probability of finding each nucleotide at each position within the mutation site for pool 1. This was calculated for read start position C53 using all of the abundant sequence reads, after both replicates had been combined (generated by Matthew Beckers).

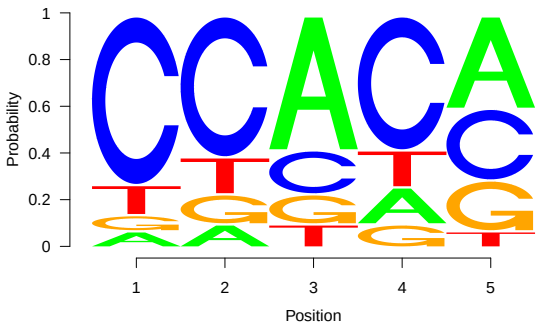


Figure 4.26: Sequence logos for mutant pool 1.
The sequence logos showing the probability of finding each nucleotide at each position within the mutation site for pool 1. This was calculated for read start position C54 using all of the abundant sequence reads (generated by Matthew Beckers).

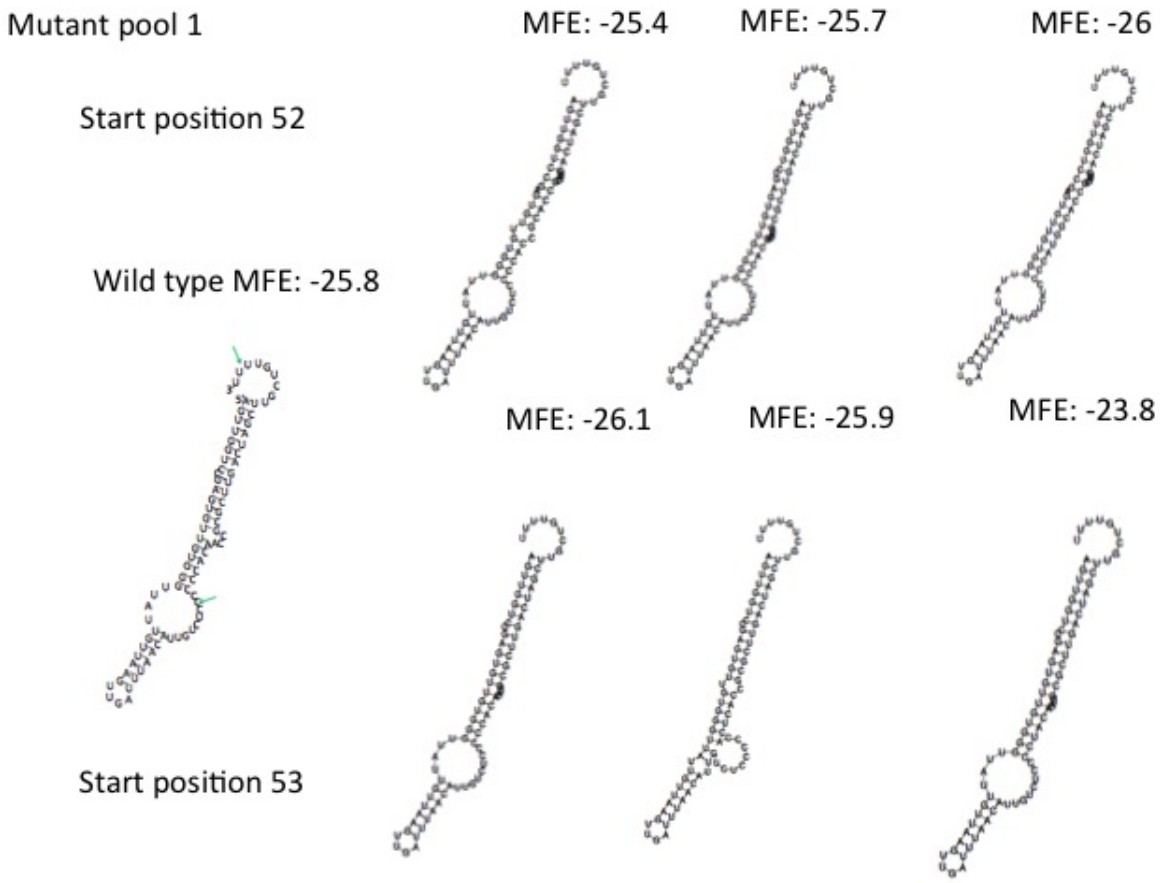


Figure 4.27: RNA structures for mutants yielding ydRNAs with the most abundant motifs in mutant pool 1 at C52 & C53.
The secondary structures of the wild type hY5 RNA and the full length mutant RNAs in pool 1 that yield ydRNAs with reads starting at positions C52 and C53 with the three most abundant mutations (as predicted by RNAfold) of the wild type hY5 RNA, and the three most abundant mutant sequences at start position C52 and C52. Green arrows represent the cleavage positions that produce the 32mer.

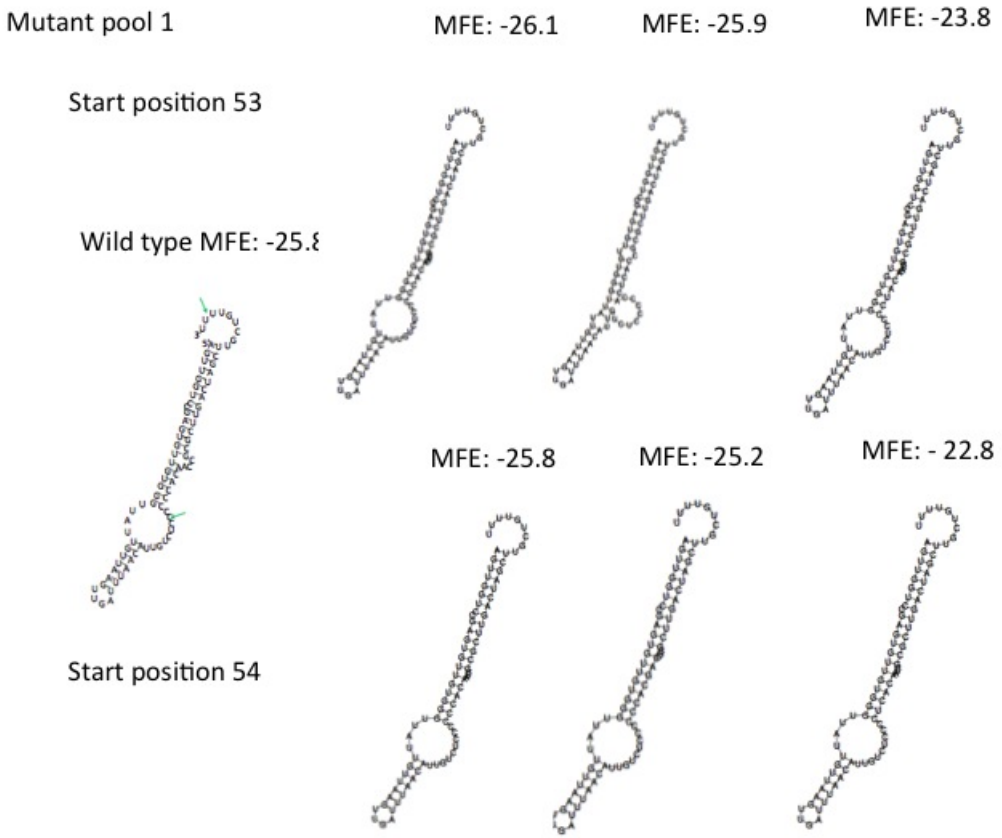


Figure 4.28: RNA structures for mutants yielding ydRNAs with the most abundant motifs in mutant pool 1 at C53 & C54.
The secondary structures of the wild type hY5 RNA and the full length mutant RNAs in pool 1 that yield ydRNAs with reads starting at positions C53 and C54 with the three most abundant mutations (as predicted by RNAfold) of the wild type hY5 RNA, and the three most abundant mutant sequences at start position C53 and C54. Green arrows represent the cleavage positions that produce the 32mer.

4.3 Discussion

Analysis of hY5 mutant full length and ydRNA levels in the NIH/3T3 cells transfected with each of the mutant pools showed detectable levels of both in all samples (Figure 4.5). With approximately a million full length Y5 mutant molecules per μg of total RNA, assuming that the ydRNAs are present at approximately ten times lower levels, each of the different ydRNA mutants would be expected to be present at an average of 975 copies per μg, giving a good chance of detection by next generation sequencing with an enrichment for 5' CCC sequences.

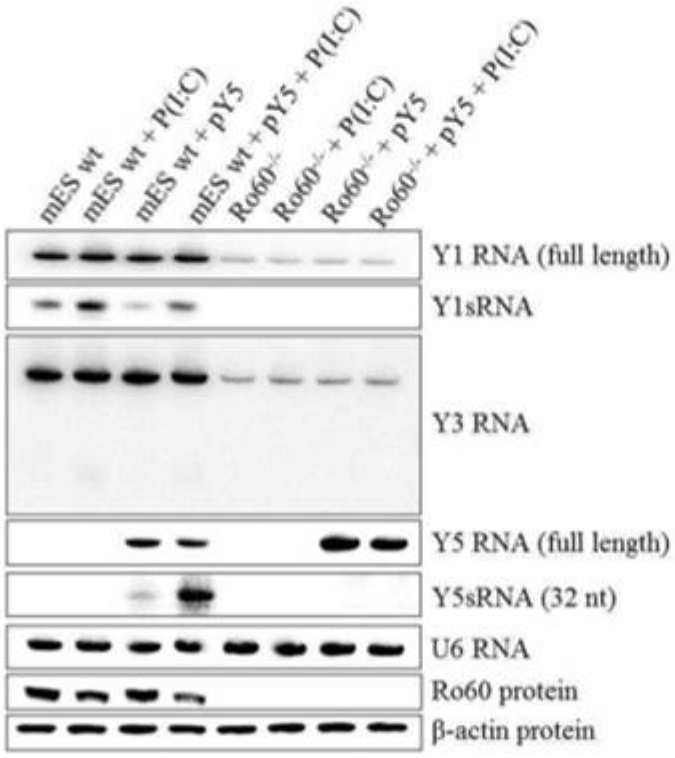


Figure 4.29: ydRNA generation is Ro60-dependent.
Wild type (wt) and Ro60$^{-/-}$ mouse embryonic stem cells (mES) were transfected with a wild type hY5 expressing plasmid (labeled here pY5) and treated with Poly(I:C) [P(I:C)]. Figure reproduced from Dr. Hall's thesis (Hall, 2013).

Surprisingly, there appears to be a size difference between the ydRNAs produced from each of the mutant pools. The mutant pool 1 ydRNAs appear resolved more slowly than those of the other two pools. It is also possible that the ydRNAs from mutant pool 2 resolved slightly more slowly than those of mutant pool 3, but this is less clear in this figure (Figure 4.5). The interpretation of any size difference is complicated by the position of the samples on the gel, as this size difference could be the result of smiling of the gel, where the samples at the edge of the gel resolve more slowly than those closer to the middle. This does not seem to be the case for

this gel, as the oligonucleotides loaded in the lanes on the right of the gel are all present at the same position. It is not possible to tell whether there is also a size difference between full length RNAs produced from these mutant pools, as RNAs of 84-86 nucleotides would be unlikely to separate on a gel of this percentage. From the single mutant analysis in Chapter 3, it seems unlikely that any of the mutations made here would have had an effect on the size of the full length RNA, as none of the mutated regions are close to either end of the RNA so would not have affected transcription. The potential size difference in full length Y5 or ydRNA lengths between the mutant pools will be discussed further, during the interpretation of the NGS data.

A preliminary analysis of the sRNA data has been conducted, but the full length and plasmid bias datasets are required for any final conclusions to be made. These libraries are currently being sequenced. After removal of the outlier replicate for each mutant pool sample set, all reads were mapped to the mouse genome and to a set of hY5 mutant reference sequences. The presence of many mouse genome-derived 30mers (as seen by the peak in Figure 4.13) in the NGS data is unsurprising, as it was necessary to include non-hY5-derived sequences in order to introduce enough heterogeneity for the samples to be processed by NGS. The presence of the 19 nucleotide peak is more unexpected, however, as the majority of sequences of this size were removed from the samples in two rounds of gel extraction. This indicates that the two microRNA sequences present in this peak, miR-99b and miR-125b-1, were very highly abundant in these samples. The sequence reads mapping to these miRNAs were 19 nucleotides in length, but taking into account the enrichment PCR, these sRNAs are more likely to be 21-22 nucleotides in length, meaning that these reads likely represent the full length miRNAs. Mir-99b contains the sequence CAC upstream of the sequencing reads (MiRBase), suggesting that the A may have looped out during the enrichment PCR. Mir-125b-1 contains CC upstream of the region sequenced, suggesting that the enrichment PCR works inefficiently to amplify both of these sequences and that they may be underrepresented here. There is evidence in the literature that both of these miRNAs play a role in cellular response to apoptosis. miR-99b regulates the expression of the mTOR protein (Sun et al., 2011). miR-125b has an opposing function to miR-99, as it inhibits apoptosis (Yang et al., 2013).

It is generally accepted that two different methods of detection are required to confirm that a sRNA is produced under a certain condition. Often these two methods do not entirely correlate and sometimes a clear conclusion is difficult. These discrepancies may be down to differences in sample processing, but work by the Dalmay group has recently begun to unravel a bias inherent to protocols utilizing RNA ligases, such as most NGS technologies. Indeed the same bias may be present

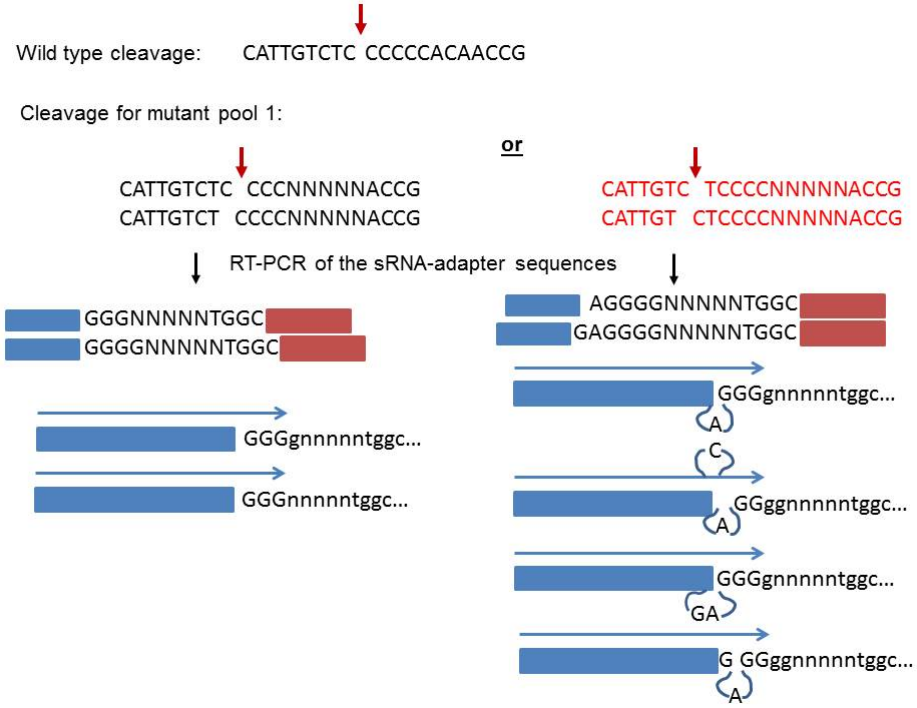


Figure 4.30: Possible sequencing reads created from mutant pool 1 sRNAs.
At the top the predominant cleavage position of the wild type hY5 RNA is shown for comparison. On the left of the diagram are the sequence reads from which the original ydRNA is easy to predict. On the right of the diagram are the ydRNAs that would not be as efficiently enriched during sRNA library preparation. During RT-PCR the sequences are converted to cDNA, so the complementary cDNA sequences are shown in the lower part of the diagram. the blue boxes represent the 5' adapter and the red boxes represent the 3' adapter. The lowercase letters show the regions present in the sequencing reads and the upper case letters are the regions amplified by the alternative RP1 primer, which are therefore not present in the sequencing reads.

in all methods which utilize a ligation step. Dr. Sorefan found that the ligase enzymes commonly used in the preparation of cDNA libraries for NGS have a bias for ligating together RNAs that can anneal to each other in such a way that the ligation point falls within a loop close to the base of the stem (Sorefan et al., 2012). As the adapters used in these techniques are set sequences provided in the kit, the ligases RNAs that are capable of interacting with these adapters to form preferred structures.

For this reason it is important to bear in mind that, although the fold change and general trends seen for abundant RNAs under different test conditions are usually correct, a direct comparison of the number of reads for one RNA with the number of reads for another is likely to be inaccurate. For example, if an RNA is favored by the ligase then it may be present in the data set at very high levels, but in reality it is present in the cell at below average expression levels. To overcome this problem, 'HD' adapters were designed to reduce the ligation bias (See Sorefan et al. 2012 for details). The use of these 'HD' adapters will be important when interpreting the results of this chapter. Due to the necessity of the enrichment PCR, it was not possible to use the 5' 'HD' adapter, but use of the 3' 'HD' adapter alone should have dramatically reduced the ligation bias.

The high-throughput method of mutagenesis has several advantages over the original method. Firstly, each of the 1024 different plasmids should enter the cells with the same efficiency, as they will have been transfected under the same conditions. This was a drawback of the low-throughput method, as different plasmid concentrations and transfection efficiencies could not be accounted for. Differences in concentration between each of the mutant plasmids prior to transfection can be normalized using the plasmid bias library dataset, when it is available.

It is possible that some of the mutated hY5 RNAs will either not be transcribed correctly or may misfold in such a way that they are rapidly targeted for degradation, by complexes such as the exosome, and will not be represent in the full length hY5 libraries dataset. This means that differences in the levels of full length mutant RNAs identified in the NGS data are likely to be a result of differences in the expression and processing of the RNA within the cells, rather than differences in the levels of mutant plasmids. Mutation of particular regions of hY5, and then subsequent examination of the effect on cleavage, was useful for determining a region of the RNA that is important in the interaction with the protein or proteins involved in the cleavage, but this method is limited by the time consuming design and generation of each mutant. The high-throughput method also has the advantage that no assumptions are made on the structure that the mutants might form, although secondary structure prediction programs can be used to hypothesize on the effect of abundant or absent mutant sequences on the secondary structure.

On examining only the sequence reads in the sRNA libraries that map to the 3' end of the hY5 RNA and contain the full mutant region, the NGS data confirms what was seen on the northern blot membrane, as the reads for mutant pool 1 are 1-2 nucleotides longer than those seen for mutant pools 2 and 3 (Figure 4.15). The length of these reads does not take into account the Cs included in the PCR and sequencing primers, so the true ydRNA lengths are likely to be 33-35 nucleotides in length. At the 3' end of the hY5 RNA there is a run of Us, and there is known length variation here. The sizes of the mutant pools 2 and 3 derived ydRNAs are therefore likely to be the same as the wild type ydRNAs. Initial analysis of the Poly(I:C)/control libraries prepared in MCF7 and SW1353 cells indicates that the true length of endogenous ydRNAs is 33-34 nucleotides (data not shown). These libraries were sequenced with reads of a length up to 50 nucleotides, so the sizing of the sRNAs is likely to be more accurate than the original libraries, which had a read length of 36 nucleotides.

When examining the start position of the sRNA reads from the different mutant pools (Figure 4.16), again a clear difference can be seen between the pools. It appears that pools 2 and 3 are very similar, with the majority of sequence reads starting at one of two distinct positions, C53 or C54. More reads start at position C54 than C53 for both of these pools. From this data we can assume that the cleavage positions for these mutant pools are most often either between T49 and C50 or between C50 and C51. This mirrors the wild type cleavage position, which is most often between C50 and C51, but can also occur between T49 and C50. For ydRNAs produced from both of these cleavage positions, the enrichment PCR would be expected to work very efficiently, as both of these cleavage events would produce ydRNAs with 3 C residues at their 5' ends.

An alternative method of analyzing the data was to sort all of the reads from each replicate of each library in order of abundance (displayed in Table 4.1). These mutant sequences were then ranked by the order they appeared in the list and the data was further sorted by combining the ranks of the two remaining replicates in each library (Table 4.2). When examining the top 6 mutations for each of the 3 mutant pools, what was unexpected was that the wild type sequence does not seem to be present in any of the hits. This suggests that the sequence requirements for Y RNA cleavage are not rigid and selection of mutant RNAs by the cleavage machinery is more likely to be dictated by the secondary structure of each of the mutants than the sequence. To examine this more closely, the sequences of the full length mutants producing the top 10 mutant hits for each of the three pools were entered into RNAfold and their secondary structures were predicted. Figures 4.19 and 4.20 show the predicted secondary structures for the top three mutant sequences for mutant pools 2 and 3, the structures of the most abundant mutants in pool 1

will be discussed later. What is noted in all of these secondary structures is that there is a 5 base pair stem above the large internal loop. On closer examination of each of the structures, the cleavage position is most often 2 nucleotides away from this stem, falling within the large internal loop. Our hypothesis is that the cleavage machinery binds the 5 base pair stem, possibly recruited by Ro60, and cleaves the stem between the second and third unpaired nucleotide. What seems likely, from these secondary structures, is that a 5 base pair stem, followed by a large internal loop, is required by the cleavage machinery because the most abundant motifs all maintain that stem.

Another surprising finding from these secondary structure predictions is that there is much variation at the top of the molecule, towards the 5' and 3' ends. In human cells, the Ro60 and La proteins are required for the stability of Y RNAs. This does not appear to be the case for hY5 RNAs expressed in mouse cells as, although the Ro60 binding site mutant tested in Chapter 3 is not cleaved, it appears to be as stable as the wild type hY5 RNA. This was confirmed when the wild type hY5 plasmid was transfected into Ro60 null mouse embryonic stem cells [See Figure 4.29, (Hall, 2013)]. In this experiment, the hY5 RNA looks to be stable in mouse cells regardless of whether mRo60 is expressed. Two different scenarios could explain this. In the first scenario, the mRo60 gene could have diverged from the human Ro60 gene in such a way that it can no longer recognize and directly bind to the hY5 RNA. In the second scenario only Y RNAs expressed from genes within the genome may be bound by Ro60, and the plasmid-encoded Y RNAs are expressed in such a way that they can no longer be recognized by mRo60. Perhaps they are localized to a different area of the cell than endogenous mY RNAs. Further experiments would need to be conducted to determine which theory is correct. In order for either scenario to be true, the hY5 must be bound by other proteins within the cell and therefore be protected from degradation in some other way.

What is evident from Dr. Hall's work is that mRo60 is required for plasmid-encoded Y5 RNA cleavage, regardless of whether or not it can directly bind to this RNA (Hall, 2013). In the mouse Y5 transfected wild type cells a clear ydRNA product is seen, whereas in the Ro60 null cells no ydRNA product is seen, despite the fact that the full length hY5 is present at higher levels than in the wild type mouse cells. The levels of full length hY5 RNA most likely higher in the mutant cells because none of these RNAs are targeted for cleavage.

Initial interpretation of the secondary structures of full length RNAs producing abundant ydRNAs in the data from pools 2 and 3, in particular from pool 3 where the mutation is located close to the Ro60-binding site, suggested that this binding site was disrupted in many of the mutants and that they were unlikely to still bind to Ro60. The bulged C seen on the left of the terminal stem (at the top of the

diagram) in the wild type structure is required for Ro60-binding, but it has been suggested that this region can form a small internal loop instead of a bulged C and either of these forms can bind to Ro60. This opening of the stem is thought to be necessary to allow Ro60 access to the backbone for binding. Further examination of the predominant mutant sequences for pools 2 and 3 suggests that the region around the C bulge may form an internal loop, which may still allow Ro60 entry. The GCTTG sequence to the right of the bulged C on the wild type structure is the region targeted for mutagenesis in pool 3. Matthew Beckers generated sequence logos to show the probability of finding each nucleotide at each of the mutant positions in the most abundant mutant sequences. The logo for mutant pool 3 shows that very few Gs are present at positions 3 and 4 (Figure 4.17). In mutants where Gs are present at either of these two positions, these Gs may be able bind to the C involved in the bulge. If Ro60 truly is required for cleavage, these mutants may be underrepresented here as they do not produce ydRNAs. What is also interesting is that the nucleotide at position 5 appears to be almost exclusively a G. Perhaps this nucleotide stabilizes the terminal stem and is required to maintain the structure for Ro60 binding. The diversity of sequences present in the sRNA library derived from mutant pool 3 was unexpected, as this mutation is so close to the Ro60-binding site. The role of Ro60 in Y RNA stability and cleavage warrants further investigation.

Mutant pool 1 has a very different start position pattern than pools 2 and 3 (Figure 4.16). Rather than having two distinct start sites, the majority of the detected sequence reads are spread across three different positions. For pools 2 and 3 the second position of the two is more common than the first, while this is not seen for mutant pool 1, where the majority of the reads start at the first position of the three. It is harder to predict the most common cleavage position for mutant pool 1, due to its close proximity to the edge of the run of Cs in the full length Y5 RNA. It seems likely that it falls either between C48 and T49, or T49 and C50. Figure 4.30 shows the possible ydRNAs and the way they might expected to be amplified by the enrichment PCR. If T49 is the first nucleotide present in the ydRNA, then it would be removed during the enrichment PCR, as either the A49 of the cDNA or a C of the alternative RP1 primer would be likely to loop out during the amplification. As the mutation site for mutant pool 1 is closer to the 5' end of the ydRNA, it was hypothesized that this mutant would be the most likely to affect the position of cleavage. What was unexpected was how much longer the ydRNA population produced from this pool would be compared to the other mutant pools and to the wild type. The fact that any reads which do not contain 2 or 3 5' Cs will not be detected is a limitation of this method. Although it is possible that there are abundant mutant ydRNAs that were present in the samples transfected with mutant pool 1 that have been missed in this analysis, it is important to highlight

that those that will be further investigated from this analysis are abundant and are informative in the investigation of the requirements for cleavage. This limitation does not apply to mutant pools 2 and 3, as the predominant cleavage position for these libraries is suitably positioned so that it falls within the run of Cs and should be efficiently detected in the enrichment PCR.

In order to examine the variation in cleavage positions in the full length RNAs from pool 1 compared with pools 2 and 3, which only have two distinct cleavage sites, the full length mutant RNA sequences from pool 1 were divided by read start site. Those mutant sequences producing sequencing reads starting at positions C52, C53 and C54 were selected and separately plotted according to abundance. Figures 4.22 and 4.23 show sequences that resemble the wild type sequence at the mutation site. In figure 4.23, the most abundant sequence at the mutation site, CCACA, is actually the wild type sequence. A nucleotides are rare in the first two positions of the mutation of reads starting at this position. In Figure 4.22, the sequences are more variable, but again the wild type sequence can be seen amongst the most abundant sequences (it is the 5^{th} most abundant read). The sequence logo for the pool 1 reads starting at position 54 shows that there is a clear preference for the wild type sequence (Figure 4.26). The logo for the reads starting at position 53 shows that there is a slightly weaker, but still a clear preference for the wild type sequence.

The ordered abundance plot for the pool 1 mutant reads starting at position 52 is very different to the other two plots for pool 1. In this plot there is a high probability of getting an A nucleotide at the first position of the mutation site, even though this is not present in the wild type sequence. The wild type sequence at the mutation site is not abundant in these reads as the logo for sequence reads at this start position shows that there is a preference against the wild type sequence.

To explain the reason for the distribution of mutant sequences across the start sites of the reads, the full length sequences of the three most abundant reads for each start sites were entered into RNAfold and the secondary structures predicted. Figure 4.27 shows the secondary structures of the full length RNAs with reads starting at C52 and C53. Although these are only predictions, and the true secondary structure of these mutant RNAs may vary, a clear pattern is visible when compared to the wild type structure. Bearing in mind that the predominant cleavage position in the wild type hY5 RNA in human cells is between C50 and C51, according to the structure prediction this is between the second and third unpaired nucleotide in the loop. In all three mutant RNAs at the top of the figure the internal loop has shifted along the RNA molecule and, although the sequence of the loop is altered, the cleavage occurs at the same position in the loop. This cleavage position is between C48 and T49, resulting in ydRNAs that are 2 nucleotides longer than the

wild type, with sequence reads starting at position C52. Apart from the middle structure prediction, the most abundant read sequences at position C53 also follow this pattern. This time the position of the internal loop is shifted in such a way that the nucleotides within the stem are one position away from where they are in the wild type. The resulting sequencing reads are 1 nucleotide longer than the wild type sequence. For the most abundant reads in mutant pool 1 with the same predominant cleavage position as the wild type RNA (Figure 4.28) the position of the nucleotides within the internal loop is the same as for the wild type structure prediction. This fits with the hypothesis that cleavage occurs two nucleotides into the internal loop of the hY5 RNA structure, regardless of which nucleotide is present at that position. As a size difference in the predominant sequences produced from the three mutant pools was also indicated by the northern blot membrane, this size difference is the result of the area targeted for mutation, rather than any selection resulting from the enrichment PCR.

4.4 Further work

To determine the most common cleavage positions of each of the plasmid pools, the top 5 and bottom 5 reads will be validated by creating single mutants and testing their expression, as well as their cleavage pattern, by northern blotting. Higher resolution PAGE will be used to size the cleavage products as accurately as possible and size marker differing in length by 1 nucleotide will be run alongside the mutants. These validation mutants will be prepared in a similar manner to the high-throughput mutants, using reverse primers containing specific mutants, rather than by hybridizing primers together and cloning them into pGEM-T Easy. This is important to follow the methods of high-throughput mutagenesis as closely as possible.

On generation of these mutants, it would also be interesting to examine their ability to interact with both human and mouse Ro60. This could be done by immunoprecipitating the Ro60 protein from either mouse or human cells and determining if the transfected mutant RNA is bound. As a control, co-immunoprecipitation of the endogenous human or mouse Y RNAs with Ro60 should also be examined. Determining the ability of mRo60 to bind to the wild type hY5 and a substantial number of mutant sequences would go some way towards determining Ro60 involvement in cleavage. If Ro60 is not detectably bound to the mutants, then it must play a role in cleavage by being recruited by other proteins or RNAs that interact with the hY5 molecule.

Another immunoprecipitation experiment that would be interesting would be to clone a number of mutants containing nucleotides other than G at position 5 of the

mutation region in mutant pool 3. This G was enriched in the mutant sequences that were cleaved, so it would provide information on the role of this G nucleotide in Ro60 binding.

To determine whether the difference in endogenous hY5 cleavage and cleavage of exogenously expressed hY5 in mouse cells is the result of expression from a plasmid or a difference between the human and mouse Ro60 protein, a hY5 mutant producing an abundant ydRNA in the NGS data could be tested in human cells. The mutant would need to be chosen carefully so that it would be unlikely to affect Ro60 binding, but be different enough in sequence so that a mutant-specific probe could be designed that did not cross-hybridize with the endogenous wild type hY5 sequence. shRNAs designed against the human Ro60 protein could be transfected in combination with the mutant plasmid and, if the difference in stability of the full length hY5 RNA is a result of a difference between the mouse and human Ro60 protein, then the mutant hY5 should not be stable in the absence of Ro60. If, however, this difference in stability is the result of inherent differences in expression from genomic DNA versus expression from a plasmid, the mutant RNA levels should be unaffected by the absence of the Ro60 protein. This mutant RNA would also need to be expressed in human cells in the absence of the Ro60-targeting shRNA, to confirm that it is highly expressed when the human Ro60 protein is present.

Once the full length and plasmid bias datasets are obtained, the sRNA library data will need to be normalized and will need to be re-examined in comparison to the expression levels of the full length RNA. Those full length mutant RNAs that are present at very high levels, but do not appear to produce and cleavage products should also be validated by northern blotting, as these could be equally important in providing information as to the requirements for hY5 cleavage.

Aside from examining the specific cleavage pattern of hY5 during apoptosis, this high-throughput method could be utilized to examine cleavage of many of the other 30mer RNAs identified in this project, as well as those that may be identified in the new Poly(I:C)/control data when it is available. For apoptosis-treated cells a similar enrichment method would be required in order to be able to detect these sequences, especially as the majority of the other 30mers appear to be less abundant than ydRNAs. This method could also potentially be used to examine cleavage patterns of other ncRNAs. The volume of information provided by a method such as this could be extremely powerful in more accurately characterizing the processing of many RNAs or, indeed, the activity of proteins targeting these RNAs.

5 POLY(I:C)-, AND POSSIBLY APOPTOSIS-, DEPENDENT 30MER sRNA PRODUCTION

5.1 INTRODUCTION

In 1941, Beadle and Tatum published the "one gene, one enzyme" hypothesis, suggesting that each human gene encodes one protein with a particular function (Beadle and Tatum, 1941). This theory was an over-simplification and has been repeatedly challenged since its conception. For example, protein-coding genes within the human genome can encode different isoforms of proteins, which are produced by alternative splicing events following transcription (DeNoto et al., 1981). The "one gene, one enzyme" hypothesis (Beadle and Tatum, 1941) was further challenged by the discovery that protein-coding genes can also contain alternative initiation and termination sites, potentially producing yet more protein isoforms (Early et al., 1980; Hagenbüchle et al., 1981). In recent years, a re-evaluation of protein-coding gene expression has highlighted further levels of complexity in the form of ncRNAs produced from the start and end sites of these genes (Kapranov et al., 2007).

Although less than 3% of the human genome encodes proteins, the majority of it is transcribed (Hangauer et al., 2013). ncRNAs that are produced at very low levels are likely to be by-products of transcription, but many ncRNAs present at higher levels, which were missed by previous methods of sequencing, have recently come into focus (Hangauer et al., 2013). Rapid cellular adaptation means that even poorly conserved RNAs may have new but important functions (Pang et al., 2006).

The majority of ncRNAs within mammalian cells are processed and modified. For example, stress-induced small RNAs derived from tRNAs are generated by processing enzymes acting on mature tRNAs, rather than targeting precursor tRNAs (Fu et al., 2009). Investigation into the generation of sRNAs from other ncRNAs is ongoing.

5.1.1 RNA processing

RNAs can be targeted by ribonucleases during their biogenesis, as the majority of RNAs are processed from precursor RNAs. RNAs can also be cleaved as part of a cellular quality control process. If they are recognized as mutant or misfolded, they must be prevented from entering the translation machinery. Finally, RNAs come to the end of their life and must be degraded; this process also involves ribonucleases.

There are RNA-targeting enzymes that have endonucleolytic activity, cutting within the RNA, or exonucleolytic activity, degrading the RNA from the ends. Exonucleases can digest RNA from either the 5' or the 3' end. Many nucleases, however, have more than one of these functions; some may have one activity on one RNA target and a different activity on another RNA target. In mammalian cells, RNA degradation pathways are often overlapping. For example, Angiogenin is mainly responsible for tRNA half production in human cells during stress (Yamasaki

et al., 2009; Fu et al., 2009), but work by Thompson *et al.* showed that the human RNASET2 enzyme could cleave yeast tRNA to produce these fragments, suggesting that some redundancy between these two enzymes is likely to occur in human cells (Thompson and Parker, 2009a).

In mammalian cells RNA degradation is extremely complex, with many enzymes capable of degrading RNA, as well as cofactors that unwind the RNA secondary structures or aid recruitment of the degradation machinery to the RNAs. The eukaryotic exosome complex can either act as an endoribonuclease (Lebreton et al., 2008; Schaeffer et al., 2009) or an exoribonuclease, degrading RNA from the 3' end (Mitchell et al., 1997). An example of an exoribonuclease known to digest RNAs from the 5' end after cap removal is Xrn1 (Houseley and Tollervey, 2009). In eukaryotic cells, a number of helicases work in combination with the exosome to aid its activity. A protein complex known as the TRAMP complex adds a run of As to the end of RNAs that require degradation and this enables access for the exosome (Vaňáčová et al., 2005). Recognition of an RNA with a poly(A) tail depends upon its cellular localization; the addition of a poly(A) tail to RNAs present in the cytoplasm generally results in increased stability (Decker and Parker, 1993) while in the nucleus poly(A)-tailed RNA is degraded (Das et al., 2003). Terminal poly(U) stretches may also act as targets for degrading enzyme complexes in the nucleus (Mullen and Marzluff, 2008). RNA degradation is extremely rapid, as the presence of unwanted RNA within the cell can have deleterious consequences on gene expression. As we know from miRNAs and siRNAs, even very small fragments can have very dramatic effects on the overall expression patterns of mammalian cells (Lee et al., 1993).Along with blocking translation, ncRNAs can bind to DNA molecules within the cell and regulate DNA repair (Pothof et al., 2009). For these reasons, the decay of unnecessary RNAs within each cell is vital and time is of the essence.

5.1.2 Virus response in mammalian cells

When viruses infect host cells, dsRNA is often formed during viral genome replication. This dsRNA is recognized by the host cell as foreign and the cell mounts an immune response. Poly(I:C) is a synthetic mimic of viral dsRNA and is used to stimulate an innate immune response (Larson et al., 1969). A variety of receptor proteins are present within, and on the surface of cells to detect the presence of viral dsRNA. Upon virus detection, cells can induce mechanisms to block viral genome replication, or they can undergo apoptosis to prevent the spread of new viral particles. Dendritic cells produce cytokines in response to viral dsRNA (Cella et al., 1999), alerting surrounding cells to the threat and inducing an inflammatory

response. Using *in vitro* systems, Poly(I:C) treatment can stimulate differentiation or apoptosis, depending upon the cell line used and the concentration and duration of treatment (Tomida et al., 1980; Flikke et al., 1970).

In our dataset the 30-35 nucleotide sRNAs derived from human Y RNAs, in particular hY5, were induced after Poly(I:C) treatment. Rutjes *et al.* found that human Y RNAs were cleaved in an apoptosis-dependent manner (Rutjes et al., 1999b). It seems likely that, at least for Y RNAs, the sRNAs detected after Poly(I:C) treatment are a more general consequence of apoptosis, rather than a virus-specific response (see Chapter 3 for further discussion). Dr. Nicolas visualized Poly(I:C)-treated cells under the microscope and noted signs of apoptosis, such as cell-rounding and enucleation. Whether or not these conditions caused apoptosis had not been tested experimentally. The first aim of this chapter was to measure the levels of apoptotsis in Poly(I:C)-treated samples, and in control cells, to determine whether Poly(I:C) treatment caused apoptosis in the cell line tested. Stress-induced fragments from tRNAs and snoRNAs are abundant in the literature, but at the start of this project no-one had examined virus or apoptosis induction of tRNA or snoRNA cleavage. The second aim of this chapter was to more closely analyze sRNAs differentially expressed between Poly(I:C) and control treatments and validate a number of these by northern blotting. Four potential sRNAs were chosen for validation, based on their differential expression between the two conditions, the pattern of small RNAs mapping to the gene, and the type of RNA gene they mapped to. In addition, a number of these novel sRNAs were further investigated to determine whether they were induced by other apoptotic stimulants.

5.2 RESULTS

Apart from snoRNA and mitochondrial tRNA-derived 31 nucleotide fragments, sRNAs of this size were generally up-regulated after Poly(I:C) treatment (Figure 5.1). This resulted in a higher overall read number of sRNAs 31 nucleotides in length in the Poly(I:C)-treated sample than in the control sample (Figure 1.3 in Chapter 1).

The levels of 31 nucleotide Y RNA-derived sRNAs were notably increased in the Poly(I:C)-treated sample compared to the control (Figure 5.1). Certain sequences derived from Y RNAs have previously been miss-annotated as microRNAs. In this dataset, these Y RNA-derived sequences account for the majority of the up-regulated sequences annotated as miRNAs. Very few *bona fide* miRNA sequences were higher in the Poly(I:C)-treated sample than the control.

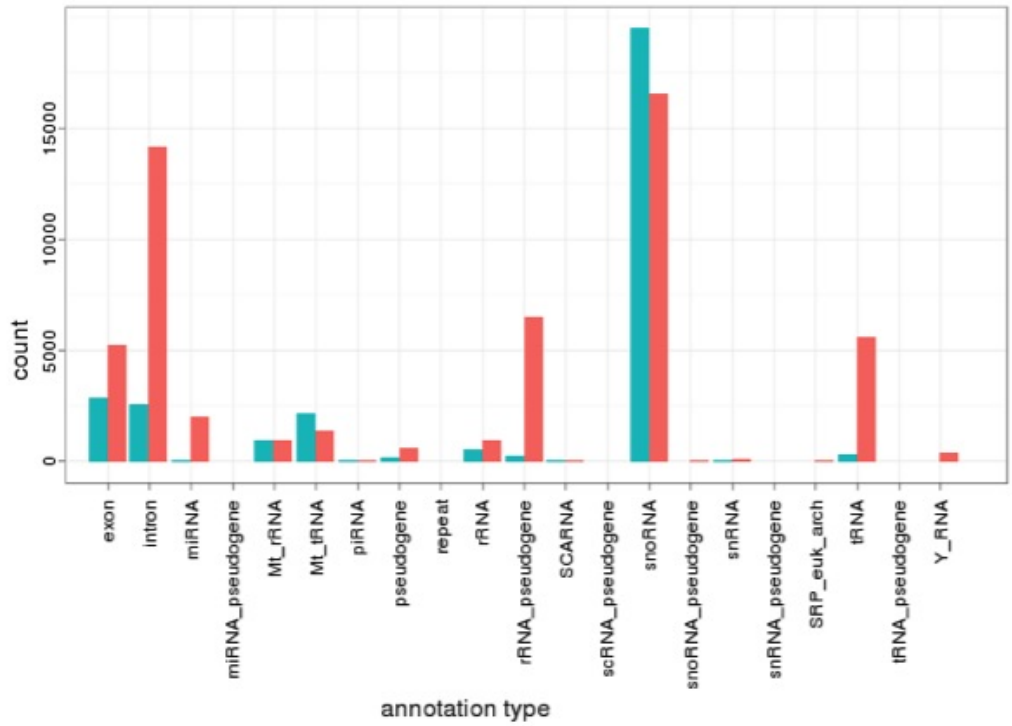


Figure 5.1: The annotation and abundance of 31 nucleotide reads in control vs. Poly(I:C)-treated MCF7 cells.
Reads from Poly(I:C)-treated cells (shown in red) and untreated control cells (shown in turquoise) were normalized and grouped according to annotation type and plotted by read count (produced by Dr. Pais). Count displayed in reads per million.

5.2.1 Validation of 30mers

In order to conduct a direct comparison of sRNA levels between Poly(I:C)-treated and untreated samples, a number of sRNAs were selected for validation by northern blotting (see Appendix VI for details). To choose sRNAs to validate, the full length RNAs from which they were derived were examined. All sequenced hits derived from each full length RNA were mapped and the read counts were compared. Only those sRNAs that represented distinct reads were further investigated. Where a large number of staggered reads were found, indicating a generally high level of degradation across the RNA rather than specific cleavage, these sRNAs were ignored. As not all of the sRNAs were found to be exactly 31 nucleotides in size, denaturing PAGE gels were run containing DNA oligos of 30 and 35 nucleotides in length, and were used to approximately size the detected sRNAs. These sRNAs will be referred to as 30mers, as this novel class of sRNAs is 30-35 nucleotides in size.

5.2.1.1 Small cajal body RNA 2-derived sRNAs

Three sRNA sequences derived from small cajal body RNA 2 (scaRNA2) were detected in the next generation sequencing data (Figure 5.2). All three sequences had a higher number of reads in the Poly(I:C)-treated sample than in the control sample. Two of these sequences, which have a 1 nucleotide difference in length, were mapped to a C box at the 5' end of the U2-61 domain of this scaRNA [see Figure 5.3 (B)]. These two sequences are likely to correspond to the same sRNA, which will be referred to the C box scaRNA2-derived 30mer. The second most abundant sRNA derived from this scaRNA mapped to a D box closer to the 3' end of the transcript; this will be referred to as the D box scaRNA2-derived 30mer.

Figure 5.3 shows that both of the scaRNA2-derived sRNAs, detected on the membranes in (A) and (B), fall between the two size markers and are therefore between 30-35 nucleotides in size. These membranes also confirm that both of these sRNAs are dramatically up-regulated after Poly(I:C) treatment. The intermediate length scaRNA band, which is expected to be 82 nucleotides in size, resolved above the 100 nucleotide marker band.

In order to determine whether the C box 30mer sRNA derived from scaRNA2 is a response specific to Poly(I:C), or whether it is induced as a general consequence of apoptosis, MCF7 cells undergoing varying lengths of treatment with Hygromyin were compared with cells undergoing 4 or 8 hours Poly(I:C) treatment, untreated cells, and Lipofectamine-treated cells (see Figure 5.4). An sRNA band is visible in the sample treated with Poly(I:C) for 4 hours and a band of a higher intensity is visible after 8 hours. No sRNA band is visible after any of the Hygromycin treatments, or in any of the control lanes, suggesting that Hygromycin did not induce scaRNA2

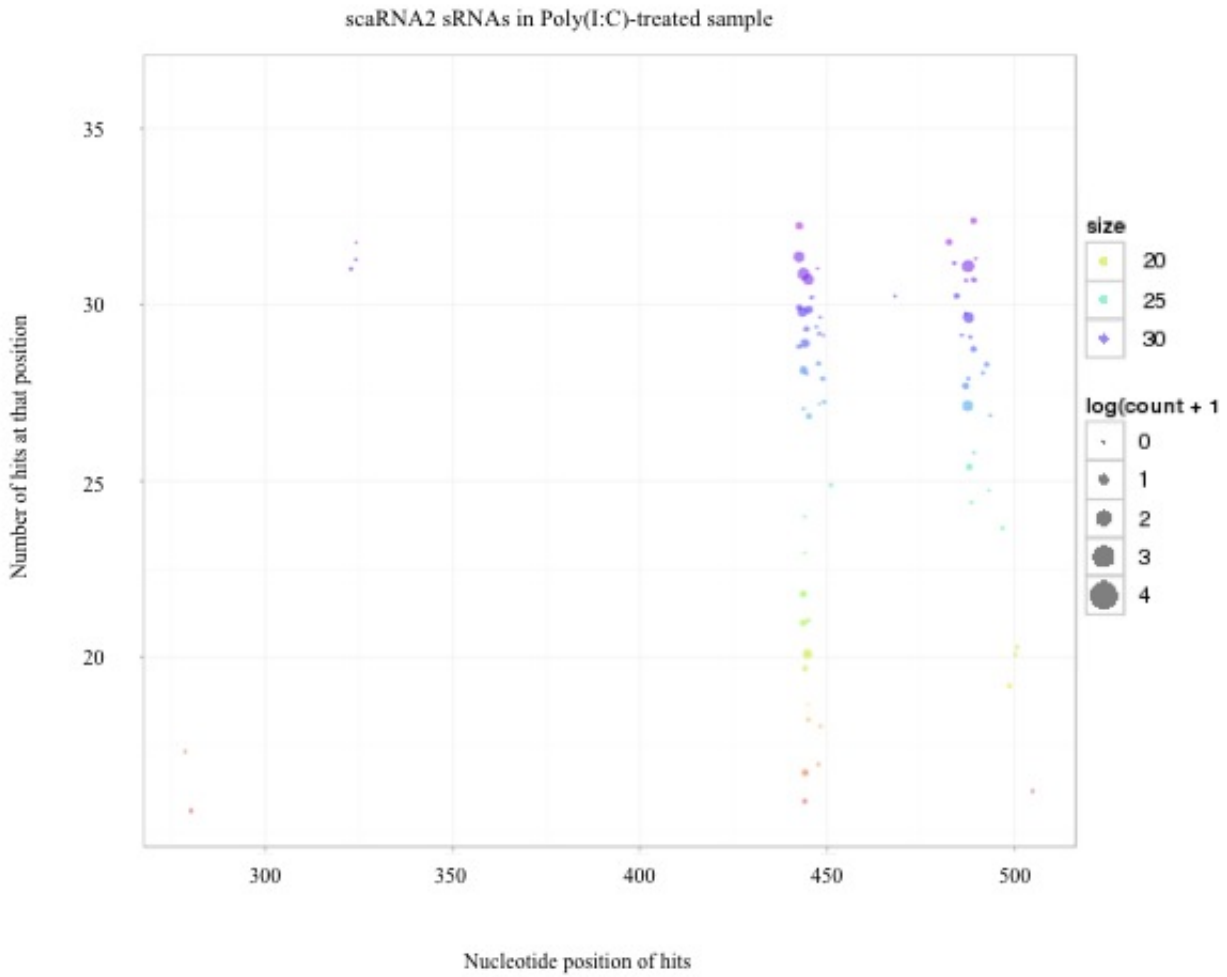


Figure 5.2: The most abundant reads mapping to the scaRNA2 gene.
Reads were mapped according to size and position from the start of the gene (produced by Dr. Pais). The colour of each dot indicates the size of each sequence detected and the size of each dot represents the number of counts (logged) of each sequence.

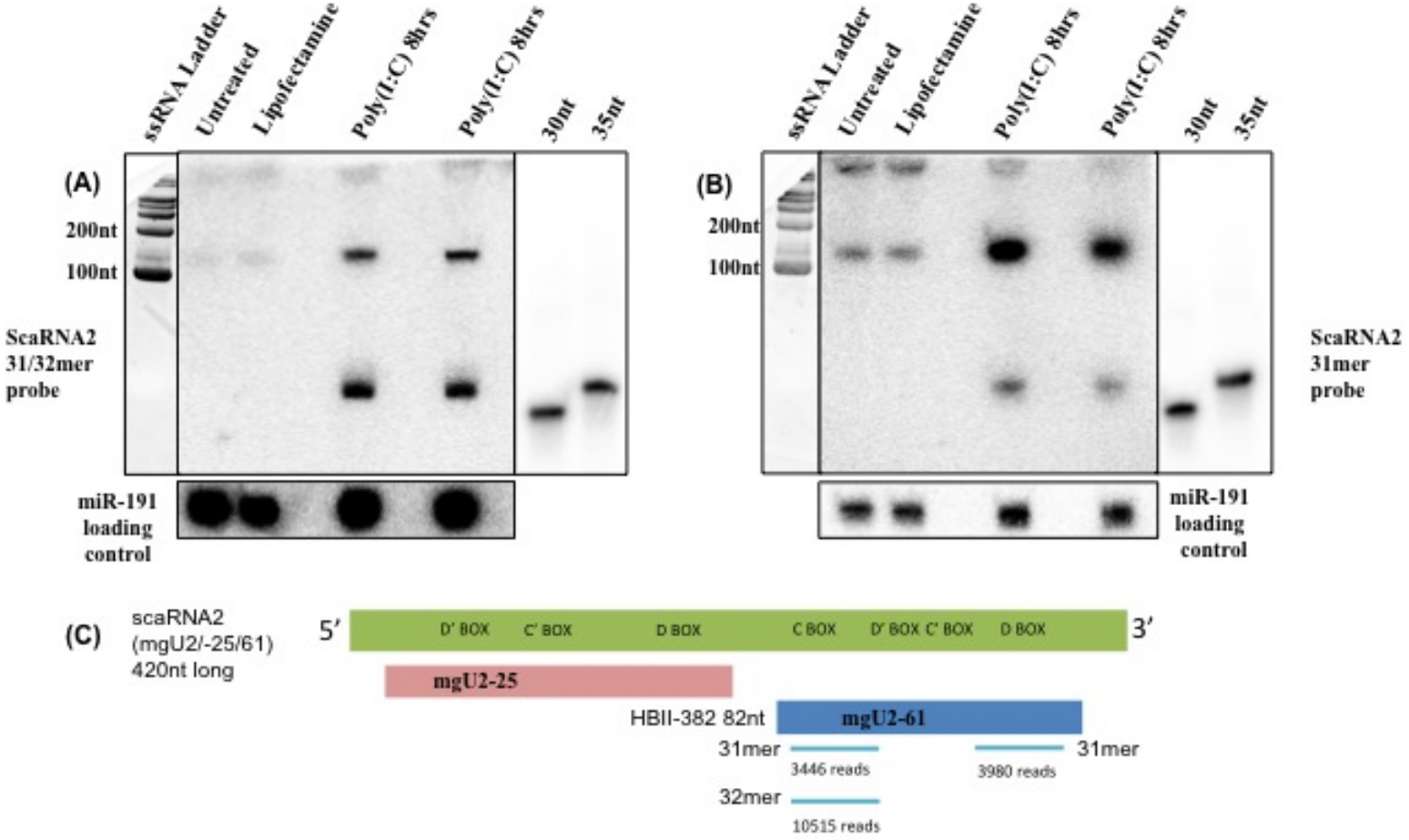


Figure 5.3: Detection of scaRNA2-derived sRNAs in MCF7 Poly(I:C)-treated and control cells.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), Lipofectamine alone, or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 10 μg was used for northern blotting. Membranes were probed with (A) a scaRNA2 C box 30mer-specific radiolabeled probe or (B) a scaRNA2 D box 30mer-specific radiolabeled probe. (C) The 3 most abundant 30mer sequences fall within the U2-61 domain of scaRNA2.

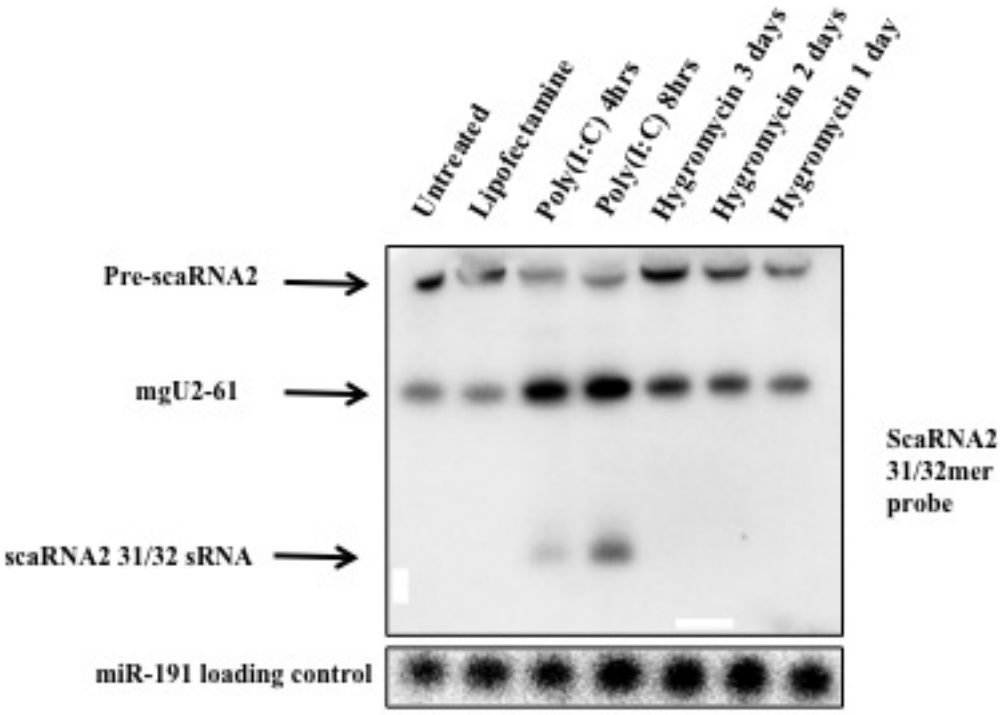


Figure 5.4: Detection of the scaRNA2-derived C box 30mer in MCF7 cells grown under various conditions.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with inducers of apoptosis or media was replaced with no treatment. After the appropriate length of time cells were harvested, total RNA was extracted and 10 μg was used for northern blotting. Membranes were probed with a scaRNA2 C box 30mer-specific radiolabeled probe.

cleavage, at least under these treatment conditions.

5.2.1.2 Small nucleolar RNA C/D Box 7-derived sRNAs

A number of small nucleolar RNA C/D Box 7 (SNORD7)-derived sRNAs were also dramatically up-regulated after Poly(I:C) treatment in the NGS data. Figure 5.5 displays the location and the abundance of the reads mapping to this snoRNA in the Poly(I:C)-treated sample. The three SNORD7 30mer sequences with the highest read number in the Poly(I:C)-treated sample mapped to the 5' end of this RNA. The two smaller sequences are derived from the same region of SNORD7 as the larger sequnce, so they are likely to represent one sRNA, with a slight variation in length. This region contains the C box (see Figure 5.6).

A probe was designed to detect all three of these sequences and was hybridized to a membrane containing control and Poly(I:C)-treated MCF7 samples. Figure 5.6 shows that in the Poly(I:C)-treated sample lane what looks to be a doublet band is visible, probably made up of RNAs varying by at least two nucleotides. All of these SNORD7-derived small RNAs fall between the two size markers, so are *bona fide* 30mers. They are not detectable in the control lanes, confirming that they are Poly(I:C)-induced 30mers. The upper band on the blot is the mature SNORD7 RNA. This RNA should be 94 nucleotides in size, but resolved above the 100 nucleotide ssRNA ladder. This is likely to be a result of the ladder being a pure RNA solution, while the total RNA samples were extracted from cells.

5.2.1.3 Zinc finger MYM-type 2-derived sRNAs

Unlike the other validated sRNAs, a sequence derived from the exon of an mRNA was selected for validation. This sequence mapped to the Zinc finger MYM-type 2 (ZMYM2) mRNA. It was one of the very few 30mer sequences to be down-regulated after Poly(I:C) treatment according to the NGS data. The sRNA detected is derived from exon 4 of the gene (see Figure 5.7), an exon that is present in 4 of the 5 protein-coding isoforms of ZMYM2. On northern blotting with a ZMYM2 30mer-specific probe (Figure 5.8), an sRNA is present at relatively low levels in both the control and poly(I:C)-treated lanes, although it looks slightly more abundant in the control samples when compared to the Poly(I:C)-treated samples. No full length ZMYM2 RNA is visible on this membrane, as it was too long to run into a gel of this percentage.

5.2.1.4 tRNA half validation

As seen by many other groups, tRNA halves were abundant in our NGS dataset. Since they were not novel, they were not the focus of this project, but a number of

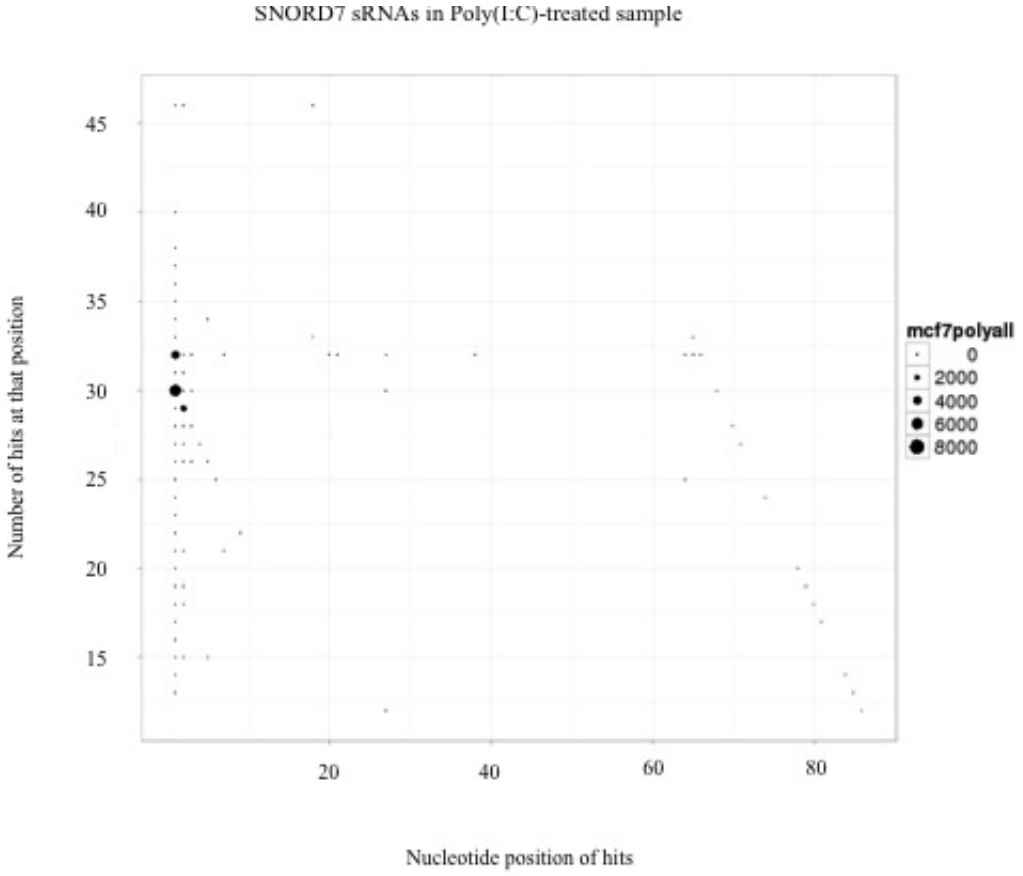


Figure 5.5: The most abundant reads mapping to the SNORD7 gene.
Reads were mapped according to size and position from the start of the gene (produced by Dr. Pais). The size of the dots represents the number of reads in the Poly(I:C)-treated sample

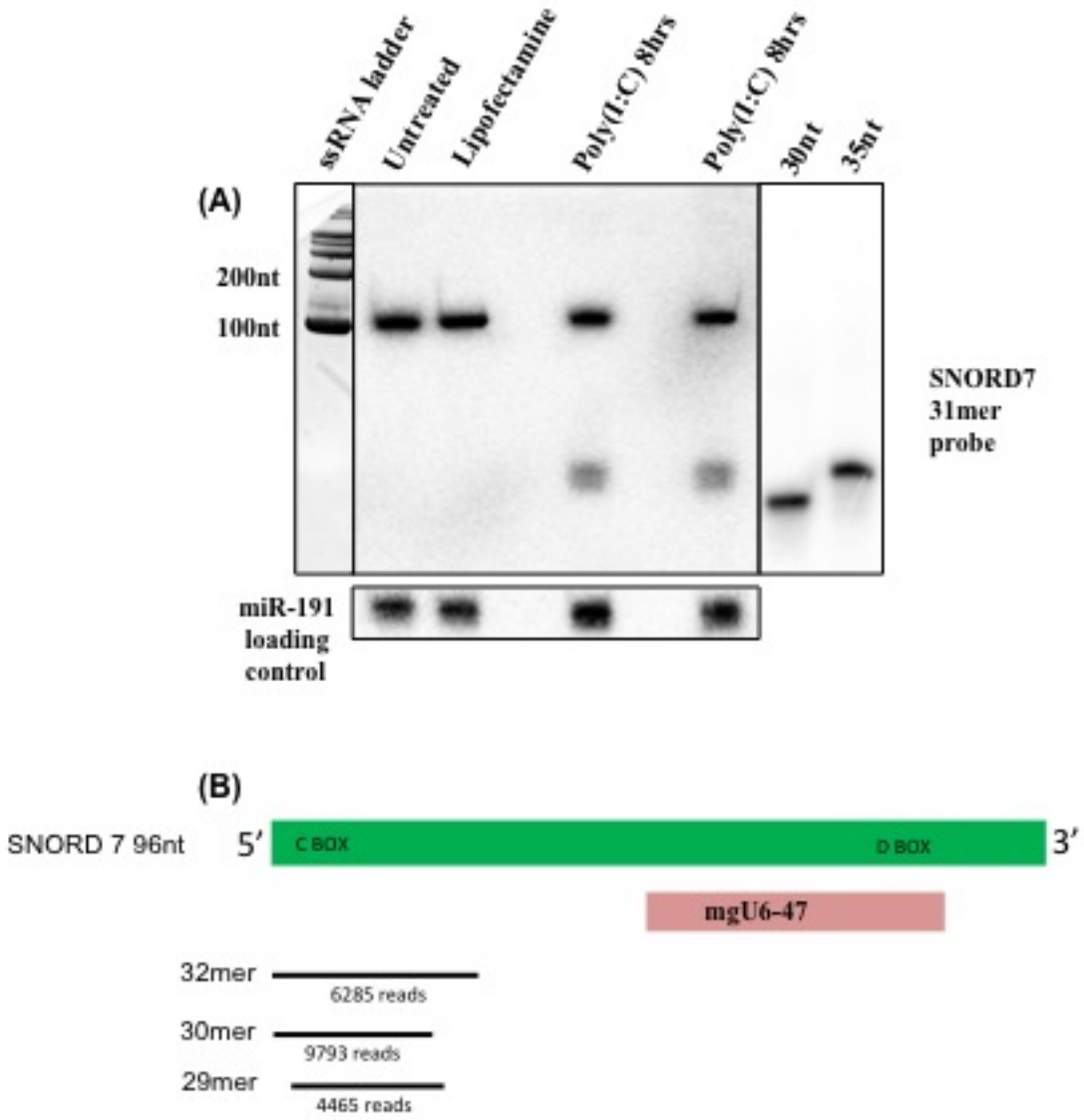


Figure 5.6: Detection of SNORD7-derived sRNAs in MCF7 Poly(I:C)-treated and control cells.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), Lipofectamine alone, or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 10 μg was used for northern blot analysis. Membranes were probed with (A) a SNORD7 C box 30mer-specific radiolabeled probe. (B) The 30mer SNORD7-derived sRNA encompasses the C box region of the full length SNORD7 RNA.

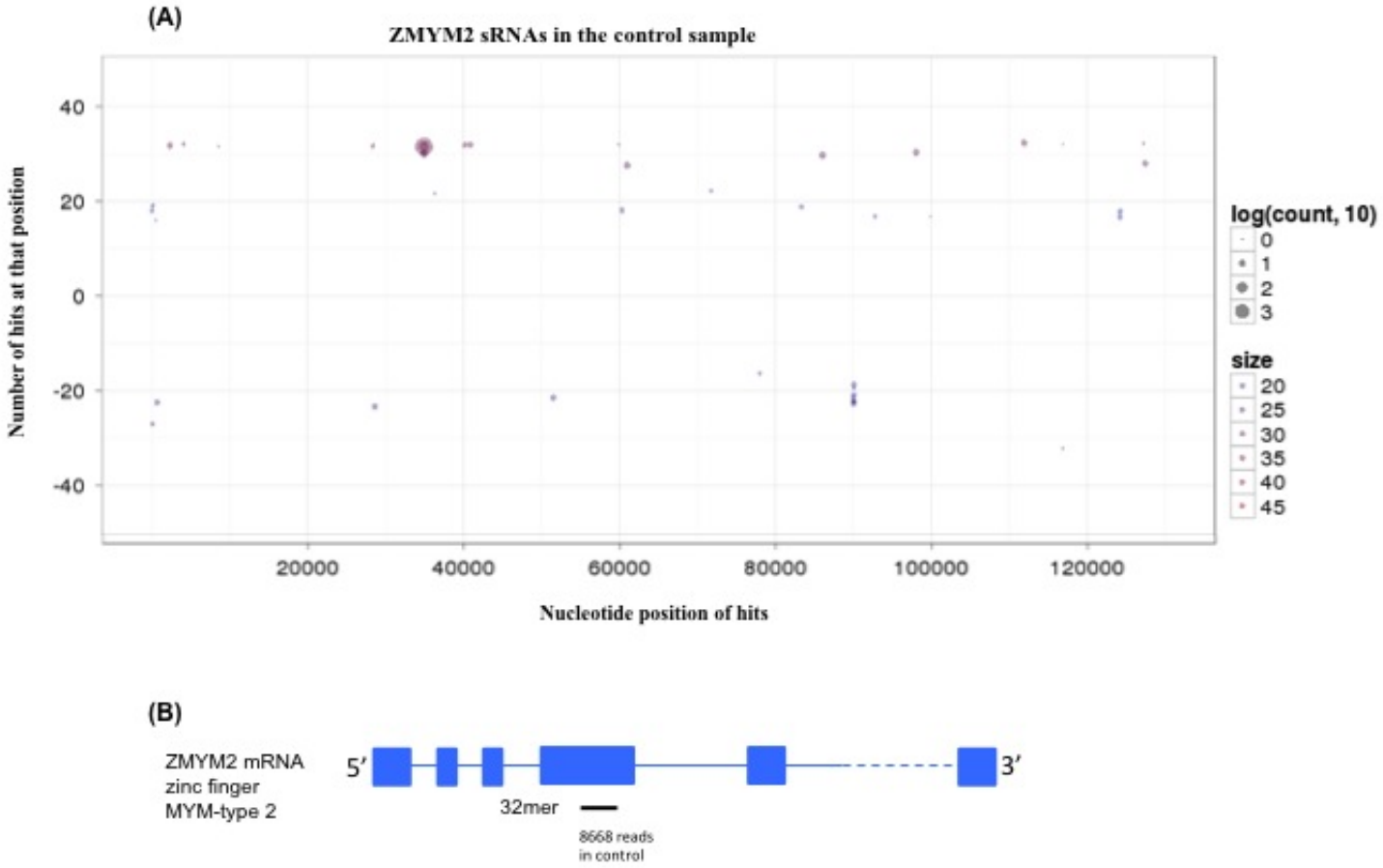


Figure 5.7: The most abundant reads mapping to the ZMYM2 gene.
(A) Reads were mapped according to size and position from the start of the gene (produced by Dr. Pais). The colour of the dots represents the size of the sequences derived from these gene and the size of the dots represents the read count (logged) of each sequence. (B) The position of the most abundant sRNA read within the ZMYM2 gene. This 30mer maps to exon 4, which is present in the majority of the protein-coding isoforms of this mRNA.

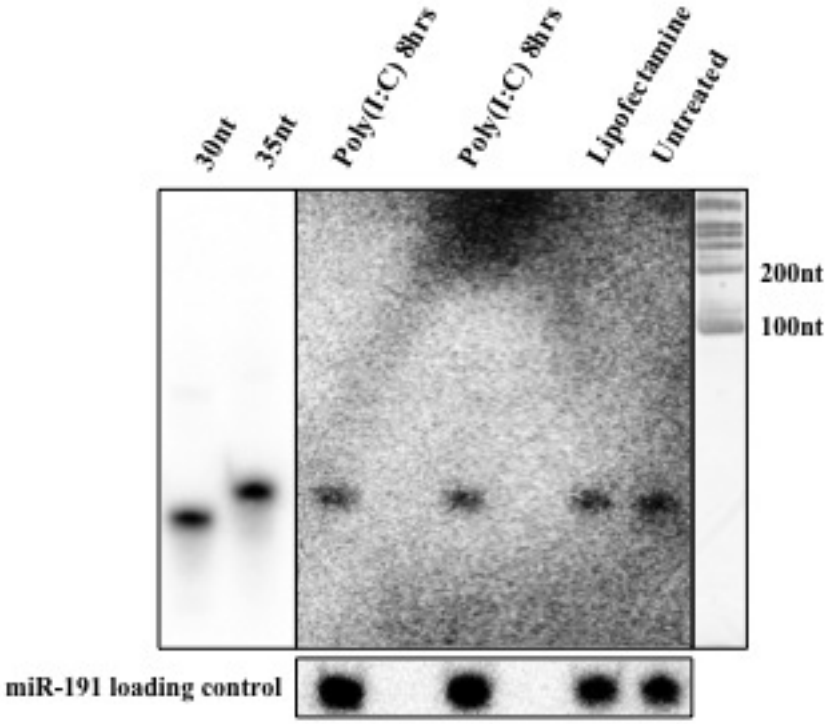


Figure 5.8: Detection of a ZMYM2-derived sRNA in MCF7 Poly(I:C)-treated and control cells.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), Lipofectamine alone, or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 10 μg was used for northern blot analysis. Membranes were probed with a ZMYM2 30mer-specific radiolabeled probe.

halves were validated by northern blotting to determine the accuracy of the NGS data. Northern blot analysis, unlike NGS, is unable to distinguish between tRNA-derived sequences with one or two nucleotide differences, as northern probes can bind to target RNAs with a number of mismatches. This is particularly problematic for those RNAs of very high abundance, such as tRNAs. Northern blotting, however, gives a more accurate read out of the difference in levels of an sRNA between experimental conditions, as there is no sequence bias in probe hybridization.

Although not the most differentially expressed tRNA halves in the NGS dataset, Ala-tRNA^{AGC} and Cys-tRNA^{GCA} were chosen for validation as they have been shown to inhibit translation initiation (Emara et al., 2010; Ivanov et al., 2011a). The Ala-tRNA^{AGC} 5' tRNA half was up-regulated 4.8 fold in the Poly(I:C)-treated sample compared to the control sample, according to the NGS data. The membrane probed for this tRNA half confirms the NGS data as a clear band of 30-35 nucleotides is seen in the Poly(I:C)-treated but not the control lanes. A single band of the expected size is present higher up the gel, corresponding to the full length tRNA. No other bands are seen, suggesting that this probe is likely to be specific for this tRNA isoacceptor type.

The 5' Cys-tRNA^{GCA} half was chosen for validation as it was present at low levels in both the Poly(I:C)-treated and the control NGS data, although there was still a slight increase in the Poly(I:C)-treated sample compared to the control sample. Despite the low level of reads from this tRNA half in the NGS data, the membrane probed for this tRNA-derived 30mer showed dramatic differential expression between the two conditions. The sRNA is of the correct size and its production appears to be Poly(I:C)-induced. A single band, corresponding to the full length Cys-tRNA^{GCA} is visible at the top of the membrane and no cross-hybridization is evident.

A probe was also designed against a Leu-tRNA-derived sequence that was highly up-regulated in the Poly(I:C)-treated sample. When membrane was hybridized with this probe, an sRNA was detected, but it resolved above the 35 nucleotide size marker and is probably around 40 nucleotides in size. Interestingly, although this sRNA is not a 30mer, it is Poly(I:C)-induced. It should be noted that a double-band is visible at the top of the membrane, indicating that the probe cross-hybridized with another full length tRNA, as well as Leu-tRNA^{AAG}.

In order to compare tRNA cleavage under different conditions, MCF7 cells were treated with various stresses, RNA was extracted and used for northern blotting, and the membrane was probed with a Met-tRNA 30mer-specific probe. The probe used in northern experiments in this project was the same as that used by Thompson *et al.*, it was also derived from the same sequence as the probe used by Fu *et al.* except that it was slightly shorter and was shifted along the tRNA by one nucleotide at the 3' end (Fu et al., 2009; Thompson and Parker, 2009a). This probe is specific to the 5'

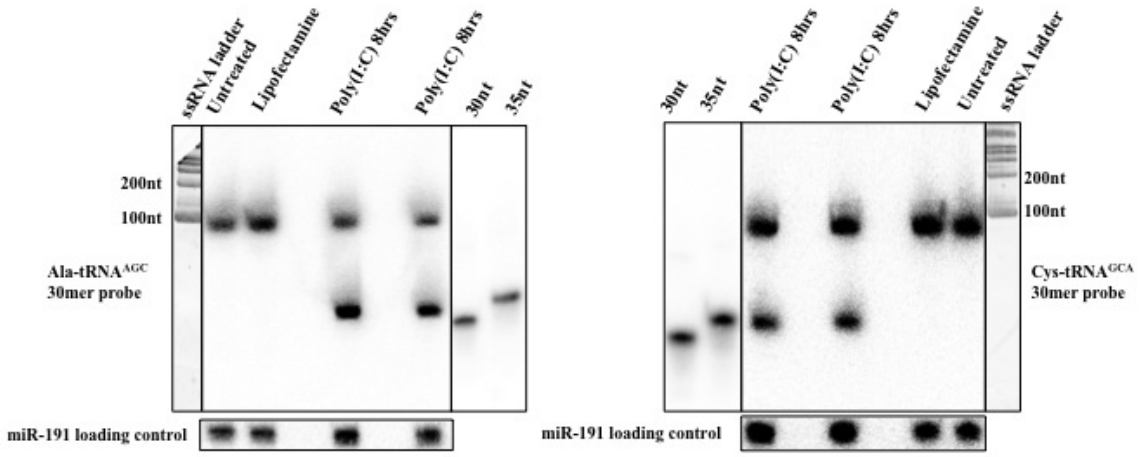


Figure 5.9: Detection of Ala-tRNA^{AGC} and Cys-tRNA^{GCA}-derived sRNAs in MCF7 Poly(I:C)-treated and control cells.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), Lipofectamine alone, or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 10 μg was used for northern blotting. Membranes were probed with (A) an Ala-tRNA^{AGC} or (B) a Cys-tRNA^{GCA} 30mer-specific radiolabeled probe.

Met-tRNA half. Unlike the other 30mers tested in this chapter, Met-tRNA is clearly cleaved after both Hygromycin and Poly(I:C) treatment, although the presence of Poly(I:C) in the cell clearly results in greater cleavage levels than Hygromycin under treatments conditions used here. No function for this Met-tRNA-derived sRNA has yet been shown.

5.2.2 Northern blot analysis of apoptosis in DLD-1 cells

DLD-1 colorectal adenocarcinoma cells were grown under various stress conditions before harvesting. After harvesting, cells were split into two and one half was used for flow cytometry, while the other half was used for northern blot analysis (See Chapter 3 for flow cytometry results and Appendix 7 for data table). RNA was extracted and 3 μg of each sample was used for northern blot analysis. The membrane was hybridized with the hY5-specific probe as a control (Figure 5.12). Using this probe, a high level of 30mer ydRNA is seen in the Poly(I:C)-treated sample and a lower level is visible in the Hygromycin-treated sample, although still higher than the levels in the control cells. The cells grown in PBS for 30 minutes have levels of 30mer ydRNA similar to the control cells. Another tRNA fragment that was considerably differentially expressed between the two conditions tested in the NGS data

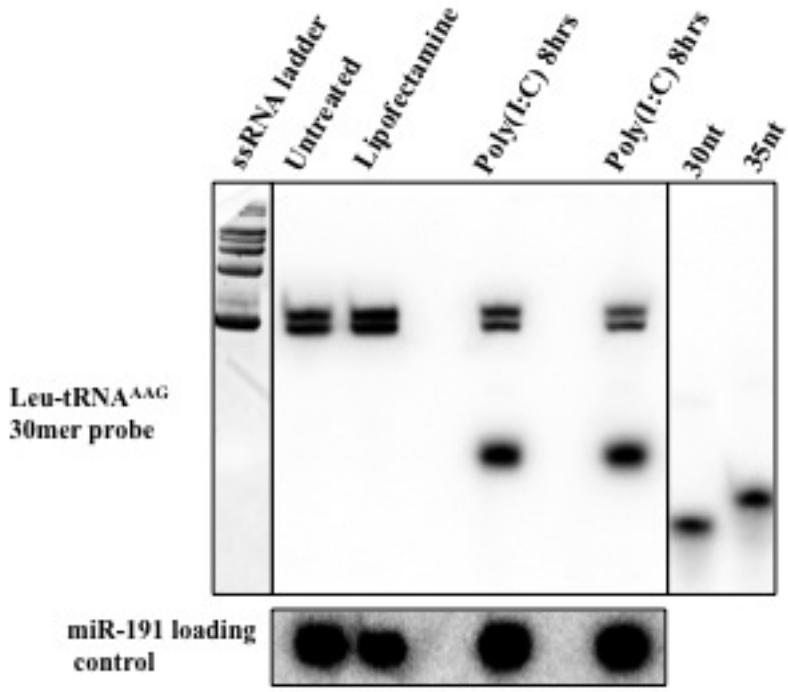


Figure 5.10: Detection of Leu-tRNA^{AAG}-derived sRNAs in MCF7 Poly(I:C)-treated and control cells.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), Lipofectamine alone, or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 10 μg was used for northern blotting. Membranes were probed with a Leu-tRNA^{AAG} 30mer-specific radiolabeled probe.

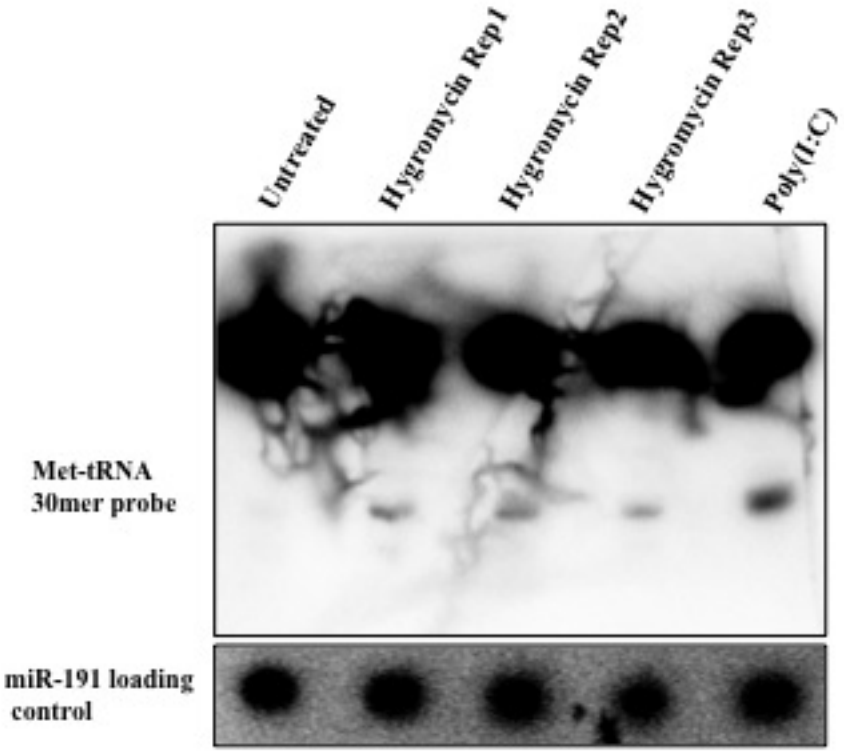


Figure 5.11: Detection of Met-tRNA-derived sRNAs in MCF7 stress-treated cells. Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), Hygromycin, or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 10 μg was used for northern blot analysis. Membranes were probed with a tRNA^{Met}-specific radiolabeled probe.

was the Gln-tRNA^{TTG} 5' tRNA half. When the same membrane was hybridized with a Gln-tRNA^{TTG} 30mer-specific probe, a strong 30mer band is visible in the Poly(I:C)-treated lane, confirming the results seen in the NGS data. A 30mer band is not present in any of the other lanes, including the Hygromycin-treated lane. It should be noted that less total RNA was used for this membrane compared with the membranes used for validation of the other NGS tRNA-derived 30mer sequences.

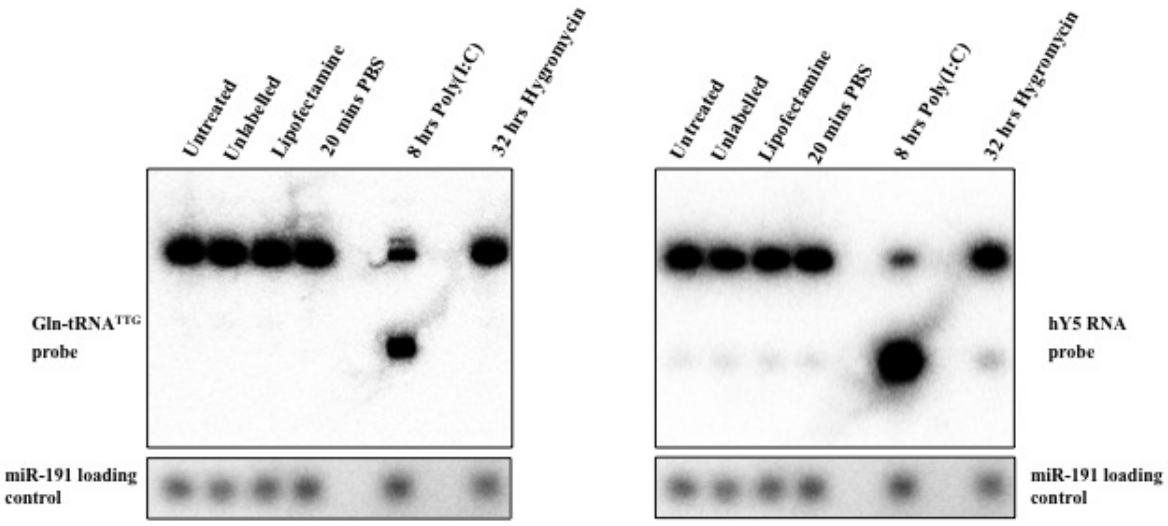


Figure 5.12: Detection of Gln-tRNA^{TGG}-derived sRNAs in DLD-1 stress-treated cells.
Cells were plated in T75cm^2 flasks. After 24 hours cells were treated with 800 mg/mL of Hygromycin or media was replaced. After a further 24 hours cells were treated with Poly(I:C) or Lipofectamine alone, left in Hygromycin-containing media, or media was replaced. 20 minutes prior to harvest, media was removed from one of the flasks and replaced with DPBS. Cells were harvested 56 hours after plating, total RNA was extracted and 3 μg was used for northern blotting. The membrane was probed with (A) a Gln-tRNA^{TTG}-specific radiolabeled probe or (B) a hY5-specific radiolabeled probe.

5.2.3 Generation of further cDNA libraries of sRNAs following Poly(I:C) or control treatment

At the initiation of this project, NGS experiments were very expensive. Over the past few years the cost of running these samples has reduced and, in order to publish NGS data, a number of biological or technical replicates are now required. To

overcome this problem, cDNA libraries of sRNAs were prepared from two different cell lines, after growth in the presence or absence of Poly(I:C) for 8 hours prior to harvesting. Three biological replicates were prepared for each condition. During the library preparation process, 'HD' adapters were ligated to the ends of the small RNAs and these were then converted to cDNA (Sorefan et al., 2012). cDNA samples were separated on polyacrylamide gels and the region containing the 30mers was cut out of the gel for each sample (Figures 5.13 and 5.14). The miRNA band was cut out separately. 30mer samples were subjected to a second gel extraction, to ensure that any adapter-adapter and miRNA sequences had been removed, and then sent for NGS analysis.

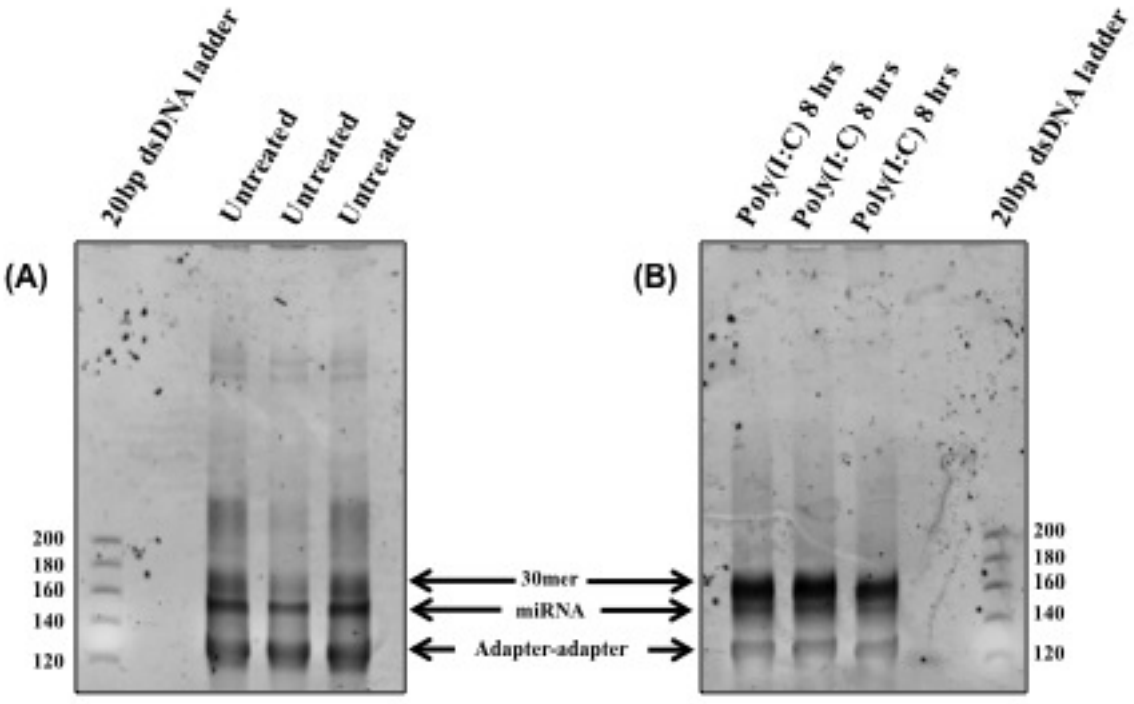


Figure 5.13: cDNA library preparation of sRNAs from (A) untreated and (B) Poly(I:C)-treated MCF7 cells.
Cells were plated in T75cm^2 flasks. After 48 hours they were treated with Poly(I:C), or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 2 μg was used for library preparation. 3 PCR reactions were prepared for each sample, each run for a different number of cycles. The upper band corresponding to the 30mers was cut from this gel for all three replicates and combined, before a second gel extraction to ensure the removal of both the adapter-adapter (lower band) and the miRNA band (middle band).

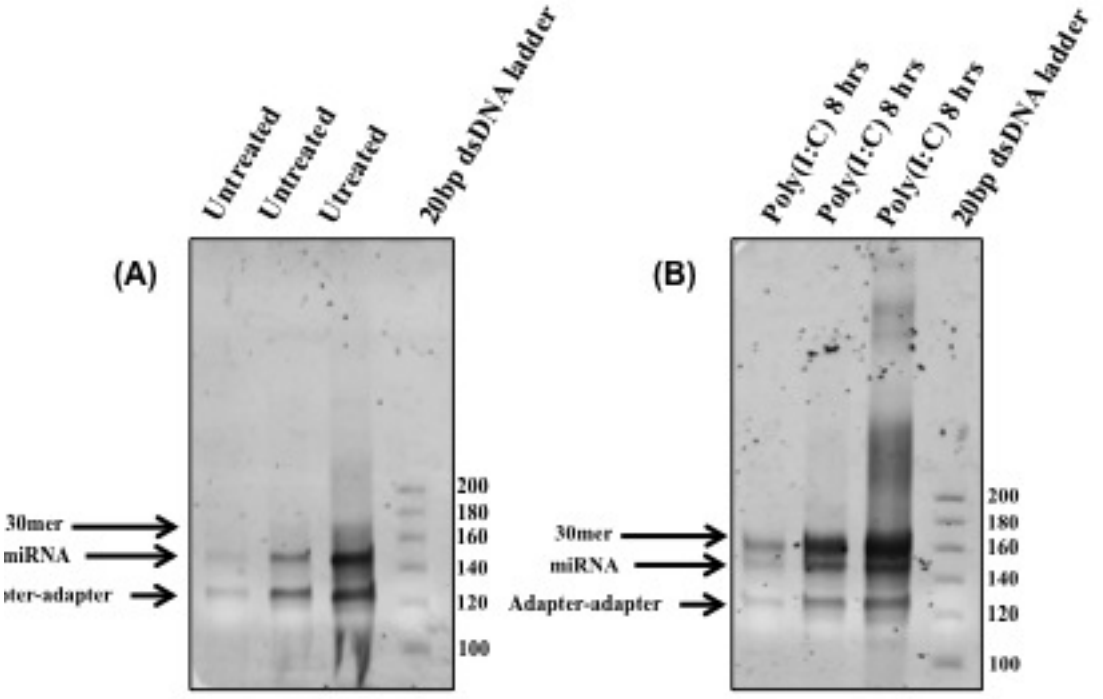


Figure 5.14: Small RNA library preparation of (A) untreated and (B) Poly(I:C)-treated SW1353 cells.
Cells were plated in T75cm^2 flasks. After 24 hours they were treated with Poly(I:C), or media was replaced with fresh DMEM. After 8 hours cells were harvested, total RNA was extracted and 2 μg was used for library preparation. 3 PCR reactions were prepared for each sample, each run for a different number of cycles. The upper band corresponding to the 30mers was cut from this gel for all three replicates and combined, before a second gel extraction to ensure the removal of both the adapter-adapter (lower band) and the miRNA band (middle band).

5.3 DISCUSSION

The majority of sRNAs derived from ncRNA were up-regulated after Poly(I:C) treatment in the NGS data (Figure 5.1). Most of the Poly(I:C)-induced 30mers were probably derived from the ncRNAs directly, as they were sense in relation to the gene. Annotation of the data showed that many of the up-regulated sequences were derived from tRNAs and rRNAs. In fact, on closer inspection, a large proportion of the sequences annotated as hits in introns or exons of protein coding genes also mapped to poorly annotated rRNA genes, tRNA genes, or pseudogenes derived from these ncRNA genes.

5.3.1 snoRNAs

C/D box snoRNAs are often given an alternative name beginning with mg, for methylation guide RNA. The rest of the name indicates the target RNA and nucleotide position that is methylated. C/D box snoRNAs are usually involved in the methylation of rRNAs or small nuclear RNAs (Kiss-Laszlo et al., 1996; Nicoloso et al., 1996). C/D box snoRNAs contain C and D box motifs, which are close to each other as the RNA forms a step-loop structure (Tyc and Steitz, 1989). C' and D' boxes are less well conserved forms of C/D boxes and are usually found more internally on the RNA molecule (Tycowski et al., 1996). A region of the RNA slightly upstream of the D box, known as the antisense element, is involved in base-pairing with the target RNA and is highly complementary to this target (Kiss-Laszlo et al., 1996; Nicoloso et al., 1996).

Although two C/D box snoRNAs were chosen for validation, it should be noted that H/ACA box snoRNA-derived 30mers were also present in the NGS dataset. Using a large number of publicly available small RNA NGS datasets, Taft *et al.* showed that animal H/ACA box snoRNAs are generally cleaved into sRNAs of 20-24 nucleotides in size (Figure 5.15) (Taft et al., 2009b). Certain microRNA-size fragments have also been shown to be derived from snoRNAs (Ender et al., 2008). In fact, many miRNAs in MiRBase have subsequently been located to H/ACA snoRNA genes (Scott et al., 2009; Brameier et al., 2011). C/D box snoRNAs are cleaved into two distinct size classes of 17-19 nucleotides and greater than 27 nucleotides (Taft et al., 2009b). In many animals, C/D box snoRNA-derived sRNAs are more abundant than H/ACA box snoRNA-derived sRNAs (Taft et al., 2009b). As a general rule, these fragments also appear to be derived from different regions of the snoRNAs (Figure 5.16) (Taft et al., 2009b). The majority of C/D box snoRNA fragments are produced in a Dicer-independent manner and were therefore not thought to act as miRNAs (Babiarz et al., 2008; Taft et al., 2009b), but a number of microRNA-like C/D box snoRNA-derived fragments have recently been identified, indicating that

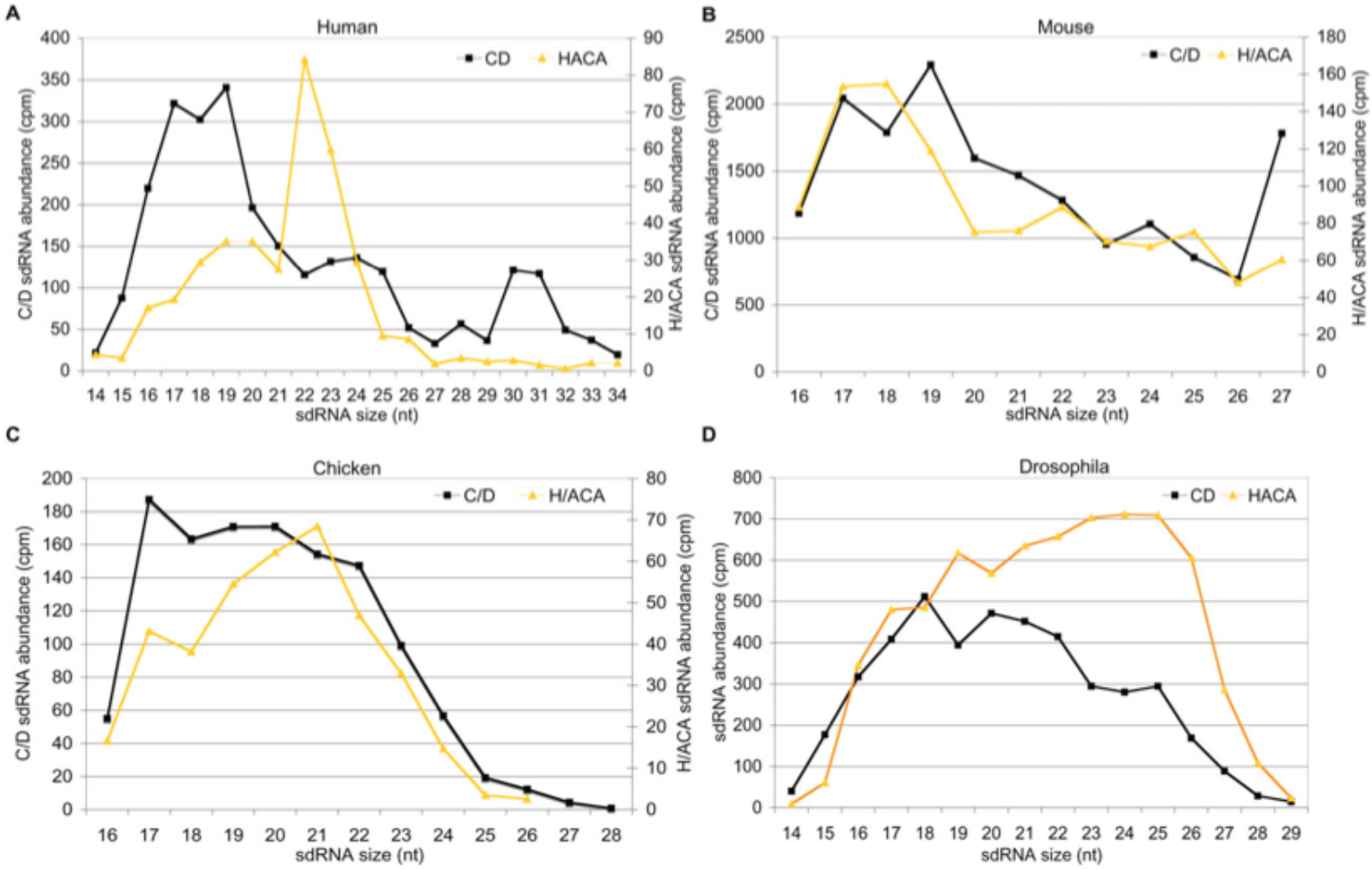


Figure 5.15: Size distribution of snoRNA-derived fragments in different animals.
The size and abundance (in counts per million) of snoRNA-derived fragments, grouped by either H/ACA or C/D box type. The abundance of the C/D box type is shown on the left and the abundance of the H/ACA type is shown on the right. Reproduced from Taft *et al.* (2009b).

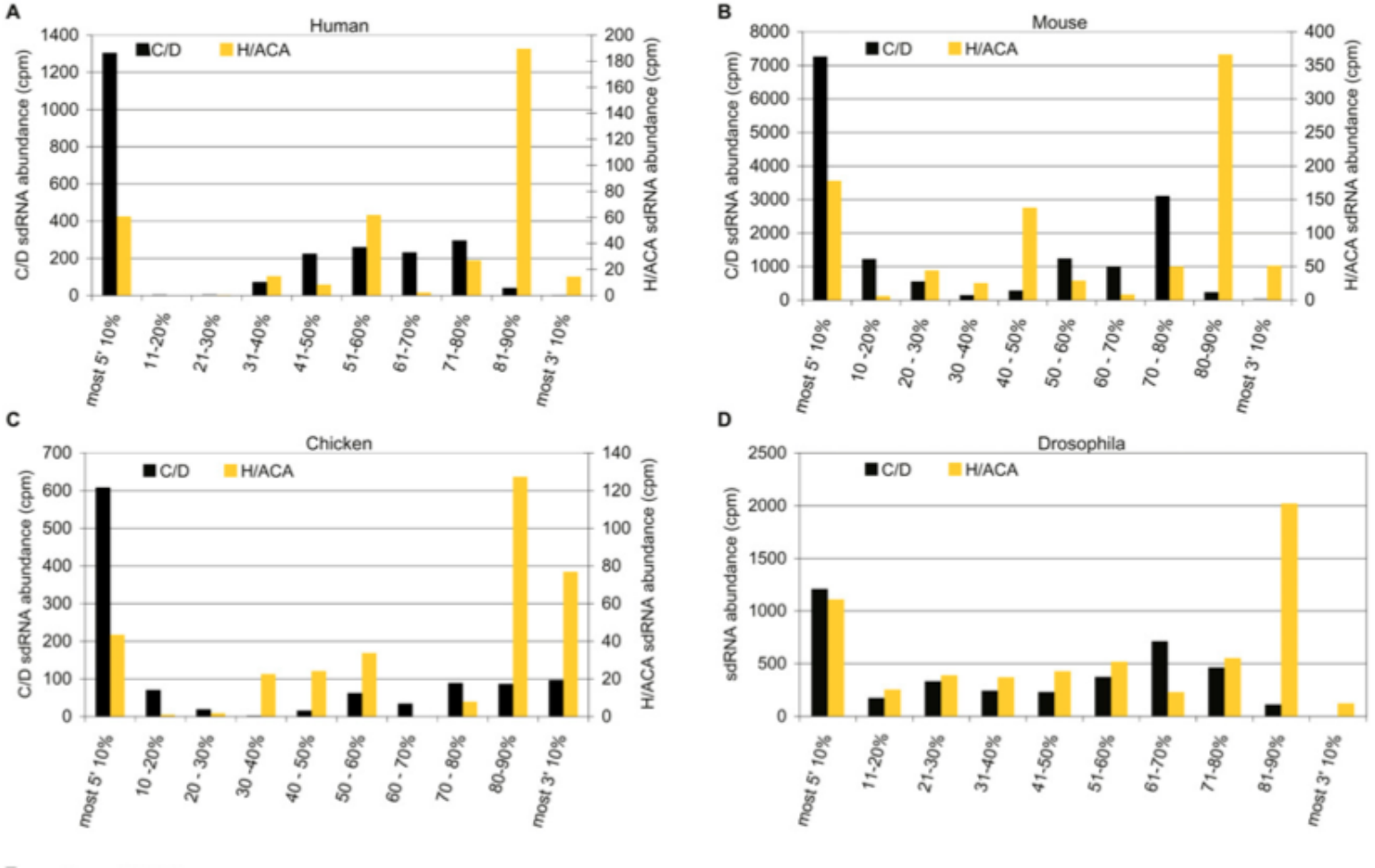


Figure 5.16: Position of fragments derived from snoRNAs in different animals.
The two types of snoRNA-derived fragments are plotted according to their abundance and position within the full length snoRNA. Reproduced from Taft *et al.* (2009b).

this type of snoRNA can also alter gene expression (Brameier et al., 2011). With a less uniform structure than the full length C/D box snoRNA, C/D box snoRNA-derived sRNAs have been shown to bind to different mRNA transcripts than the parent snoRNA (Scott et al., 2012). The HBII-180C C/D box snoRNA has C and D boxes and targets 28S rRNA, but also contains an M box, which is complementary to an intron of the FGFR3 gene (Scott et al., 2012). Cleavage of this HBII-180C gives a fragment that contains the M box and is inversely correlated to the levels of a short isoform of FGFR3 in a number of cell lines, suggesting that this fragment regulates splicing of this mRNA (Scott et al., 2012). Transient transfection of cells with reporter plasmids containing the intronic region of FGRF3 confirmed this function (Scott et al., 2012).

5.3.1.1 scaRNA2

scaRNA2 (mgU2-25/61) targeting of U2 snRNA at positions G25 and C61 results in 2'-O-methylation at these positions. Figure 5.4 shows the positions of the box motifs on the secondary structure of scaRNA2, along with the regions of complementarity with U2 snRNA. scaRNA2 is unusual, as it is one of the few snoRNAs that is transcribed independently by RNA polymerase II (Gerard et al., 2010; Tycowski et al., 2004), and is encoded in an intergenic region of the human genome (Gerard et al., 2010). scaRNA2 has a snoRNA domain containing a C and a D box (the 3' mgU2-61 domain) and a snoRNA-like domain containing a C' and a D' box and a C and D box (the 5' mgU2-25 domain) (Tycowski et al., 2004). The U2-61 domain is present as a cleavage product of the full length scaRNA2 and is a homolog of the mouse MBII-382 RNA (Tycowski et al., 2004). Tycowski *et al.* found that the mgU2-25 domain is highly conserved among vertebrates (Tycowski et al., 2004). The full length scaRNA2 was localized to the nucleoplasm, which fits with its presence in cajal bodies, the site where the U2 snRNA is 2'-O-methylated (Tycowski et al., 2004). The mgU2-61 domain was found in the nucleolus (Tycowski et al., 2004). Whether this domain alone is responsible for U2 snRNA C61 methylation, or whether this role is carried out by the full length scaRNA2 has not yet been determined. The mgU2-25 domain has not been detected on its own by northern blot and is not believed to be stable in the absence of the mgU2-61 domain (Tycowski et al., 2004). This was not investigated here as all probes were specific to the mgU2-61 domain.

Two sRNAs were found to be highly differentially expressed in the NGS data. The 5' 30mer scaRNA2 sequence encompasses the C box of domain mgU2-61 (denoted C box scaRNA2-derived 30mer), while the 31mer sequence includes the D box motif (denoted D box scaRNA2-derived 30mer). The C box 30mer starts at the same position as the start of the mgU2-61 domain and includes most of the U2 snRNA

complementary region, so it may still interact with the U2 snRNA after cleavage. The start position of this sRNA suggests that its biogenesis is linked to the mgU2-61 domain. As there is no evidence that the mgU2-61 domain is transcribed separately from the full length scaRNA, it seems likely that this domain is cleaved from the mature scaRNA2. The C box 30mer sRNA could be either a cleavage product of the mature scaRNA2 or of the mgU2-61 domain directly. As they both share the same 5' cleavage site, the most plausible explanation is that the C box 30mer sRNA is produced directly from the mgU2-61 domain. If this domain is targeted for cleavage in the presence of Poly(I:C), interaction of the C box 30mer region with proteins or with the U2 snRNA may result in protection and prevent the full degradation of this sRNA. Whether this sRNA still interacts with U2 snRNA and performs some function requires further investigation. In the NGS data the C box 30mer sRNA is present at higher levels after Poly(I:C) treatment than the D box 30mer, but this could be an artifact resulting from sequencing bias. Its presence at lower levels in the NGS data does, however, fit with its production after a second processing event of the mgU2-61 domain.

The band at the top of the membrane (Figure 5.4) in each lane of both of the scaRNA2 probed northerns is the full length 420 nucleotide scaRNA2, which did not run into the gel very far and is therefore difficult to size correctly. The intermediate-sized band visible on the membranes for both scaRNA2-derived sRNAs ran more slowly than the 100 nucleotide band of the ladder, although the mgU2-61 domain has previously been shown to be 82 nucleotides. This is probably the result of differences in the purity of the samples, as the ladder was pure ssRNA, which the samples were processed from cells and were not as pure. Interestingly, the intermediate band appears to be increased in the Poly(I:C)-treated samples when compared to the controls, despite the fact that a smaller band is also visible. This is true for both the C box 30mer and the D box 30mer sRNA northerns, suggesting that both sRNAs are derived from this intermediate RNA; giving us more confidence that this intermediate band is the mgU2-61 domain. The discrepancy with the expected size and the visible size requires further investigation. The C box 30mer band appears to be stronger in intensity than the intermediate band, while the D box 30mer band appears weaker. This mirrors the NGS data, where the C box 30mer sequence had more reads than the D box 30mer. An accurate comparison of the levels of sRNAs by northern blot can only be achieved by comparing the band intensity of each to the same set of size standards at set concentration. Given more time, this would be done to further characterize these scaRNA2-derived 30mers. Both of these scaRNA2-derived sRNAs fall between the 30 and 35 nucleotide size markers, although on the C box 30mer-probed membrane there is a wider band visible, as would be expected if this sRNA varied by one nucleotide in length. In order for the mgU2-61 domain

to be stable after a cleavage event from the full length scaRNA2, it is possible that it undergoes some sort of 5' modification. This could explain why the C box 30mer, which has the same 5' end as this domain, is at higher concentration than the D box 30mer, which has a different and probably less stable 5' end.

The difference in levels of the intermediate band between the Poly(I:C)-treated and control samples could be investigated by mutating the scaRNA2 gene in a similar manner to the hY5 gene in Chapter 3, introducing a mutation that prevents 30mer production. The levels of the intermediate band could then be compared between control and Poly(I:C)-treated cells expressing this mutant construct, to determine whether its biogenesis is altered during apoptosis. As the levels of this scaRNA are much lower in humans than the levels of hY5, this might be more effectively investigated in a human cell line, where we have confidence that all of the cleavage machinery is present.

5.3.1.2 SNORD7

SNORD7, otherwise known as mgU6-47 or Z30, methylates the A47 nucelotide in the U6 snRNA (Zhou et al., 2002). The human SNORD7 is located in the nucleolus and is 105 nucleotides in length (Zhou et al., 2002). It is a C/D box snoRNA containing one C and one D box, with the 3' and 5' ends forming a stem (Zhou et al., 2002), it may also contain a C' and a D' box (**?**). The upper band on the membrane resolved just above the 100 nucleotide band in the ssRNA ladder, which is the expected size for the full length SNORD7 RNA. Three 30mer sequences were identified that mapped to the SNORD7 gene. As they are all derived from the same region of the gene and vary in length by only a few nucleotides, they are likely to represent one sRNA. The band pattern on the northern blot for the SNORD7-derived sequence confirms the NGS data, as a diffuse band is visible, probably representing an sRNA sequence with several nucleotides variation in length. The sRNA is derived from the 5' end of SNORD7, but could occur as the result of a novel transcription event or of processing of the mature SNORD7 RNA. This sRNA contains the C box region of the SNORD7 RNA, so is unlikely to be able to interact with the U6 snRNA in a similar manner to the full length SNORD7 RNA. Perhaps this snoRNA, like HBII-180C (Scott et al., 2012), contains an alternative box complementary to a different RNA. It may interact with the target RNA to regulate RNA splicing or gene expression.

5.3.2 mRNAs

5.3.2.1 ZMYM2

The ZMYM2 protein (also known as ZNF198) can interact directly with chromatin and with the LSD1-coREST-HDAC1 ternary complex via its MYM-type zinc finger domains (Gocke and Yu, 2008). The ZMYM2 protein recruits this complex to an as yet uncharacterized group of genes, one of which is E cadherin, resulting in transcriptional repression (Gocke and Yu, 2008). It also prevents transcriptional repression of RE1-silencing transcription factor (REST) -responsive genes (Andres et al., 1999; Gocke and Yu, 2008) Alterations in expression of the ZMYM2 protein, therefore, could have a dramatic effect on the overall expression levels of the cell. The ZMYM2 mRNA is 10199 nucleotides long, which explains why no mRNA band can be seen on the membrane (Figure 5.8). According to the NGS data, the ZMYM2-derived sRNA of interest was considerably higher in the control sample than the Poly(I:C)-treated sample. This is not as clear when examined by northern blot analysis. A clear band of the correct size is visible, but it is only slightly more intense in the control samples than the Poly(I:C)-treated samples, if at all. This 30mer is somewhat unusual as it is produced from a middle exon, but does not share a terminus with this exon. Its production may be linked to the splicing of the full length ZMYM2 mRNA. The relationship between biogenesis of this 30mer and the ZMYM2 mRNA cannot be separated, as this 30mer is not differentially expressed between the two experimental conditions tested here. The exon encoding the 30mer is present in the majority of the protein-coding isoforms of this mRNA. The 30mer looks to be present at low levels according to the northern data. miRNAs, even when expressed at relatively low levels compared to mRNAs, can have dramatic effects on gene expression. Although this 30mer is unlikely to act in a similar manner to miRNAs due to a difference in size, its production could alter ZMYM2 mRNA isoform ratios: this could have dramatic effects on the full length mRNA levels and therefore on the chromatin modification status of the cell.

It would be interesting to investigate the biogenesis of this 30mer further, and to determine whether its production has any effect on the levels of the full length RNA. Validation of this 30mer shows that not all 30mers are apoptosis-dependent, even though very few are derived from mRNAs. Does this 30mer represent an exception to the rule or are there many other 30mers present in cells at low levels that are not up-regulated after Poly(I:C) treatment? This question warrants further investigation.

5.3.3 tRNAs

At the point of analysis of this dataset, tRNA cleavage was already being investigated by a number of other groups. Therefore, cleavage of other ncRNAs became the focus of this part of the project. As tRNA halves made up such a large part of the dataset, however, it was important to validate a number of them. The most differentially expressed tRNAs between the two conditions in the NGS data were validated, as well as two tRNAs known to play a role in blocking translation initiation, which were also differentially expressed between the two conditions. A tRNA that has been shown to be cleaved during cellular stress, but does not play a role in translation initiation was also analyzed here. tRNAs are very abundant in the cell and, as the majority of the sRNAs derived from these tRNAs were present at similar levels to the full length tRNAs, we can conclude that they two must be abundant.

Both Leu-tRNA^{AAG} and Gln-tRNA^{TTG} 5' halves, which were up-regulated after poly(I:C) treatment in the NGS data, were validated by northern blot analysis. As the tRNA code is degenerate, the majority of tRNAs are present with a number of isoacceptor sequences. The human genome contains not only many tRNA genes with different isoacceptor sequences, but also many copies of the same tRNA gene with the same isoacceptor sequence. The sequence similarity between tRNA genes of different types means that northern blot data should be interpreted with caution, as it is possible that a probe designed to detect one tRNA may also bind less stringently to a different tRNA sequence. This appears to be the case for the Leu-tRNA^{AAG} probe as, although only one sRNA band is visible on this gel, two full length tRNA bands are visible. This suggests that this probe hybridized to either two different isoacceptors of Leu-tRNA, or to a Leu-tRNA and a different tRNA with a similar sequence but different length. Using the University of California, Santa Cruz blat search (UCSC Genome Bioinformatics and International Human Genome Project), the most abundant NGS read was mapped to four different Leu-tRNA genes, all predicted to be 82 nucleotides in size, so hybridization of the probe to any or all of these would not explain the size difference seen on the membrane. What is also interesting is that the tRNA half detected is clearly larger than the predominant read in the NGS data, as it resolved notably more slowly than the 30 and 35 nucleotide size markers. It is hard to determine whether a 30mer is present at lower levels on this membrane, but this larger sized tRNA half is definitely more abundant. As this tRNA is 82 nucleotides in size, a tRNA half derived from it might be expected to be around 40 nucleotides. This 40mer sequence is Poly(I:C)-induced.

The northern blot membrane for Gln-tRNA^{TTG}, containing DLD-1 samples, showed that this tRNA is cleaved to give a 5' tRNA half after Poly(I:C) treatment. As no 30mer bands are visible in the Hygromycin-treated samples, we can make no con-

clusions about whether this tRNA is cleaved under general apoptosis conditions, or whether cleavage is Poly(I:C)-specific. The Y5-probed image for this membrane shows that ydRNAs in the Hygromycin-treated samples were at much lower levels than in the Poly(I:C)-treated samples, so it is possible that a Gln-tRNA^{TTG} 30mer was present after Hygromycin treatment, but was below the limit of detection.

Although Ala-tRNA^{AGC} was not the most significantly differentially expressed tRNA between the two conditions, it was chosen for validation as it is one of the tRNA halves shown by Ivanov *et al.* to inhibit translation initiation of certain proteins (Emara et al., 2010; Ivanov et al., 2011a). As this tRNA half was detected at such high levels in the presence of Poly(I:C), it is possible that it has a similar function under these conditions. The northern blot validation confirms that a tRNA half of 30-35 nucleotides is derived from this tRNA and is dramatically up-regulated after Poly(I:C) treatment. A 5' Cys-tRNA^{GCA} 30mer was present at low levels in the control sample and slightly higher levels in the presence of Poly(I:C), according to the NGS data. This tRNA half was also shown to inhibit translation by Ivanov (Ivanov et al., 2011a). Northern blot results did not confirm this data, as there was a striking difference in expression between the two conditions on the northern membrane probed for this tRNA half, with levels in the Poly(I:C)-treated sample on the membrane much higher than in the NGS data. As there were so many 30mers up-regulated in the Poly(I:C) dataset compared to the control, this would have resulted in a dramatic difference between the overall numbers of 30mers present in the two conditions. It is therefore possible that the 30mers which remain at the same level under both conditions may be lost from the Poly(I:C) dataset. Alternatively, unfavorable sequences may be lost from the Poly(I:C) dataset as a result of the inherent sequence bias of the Truseq method. This may explain the discrepancy between the NGS data and the northern blot results for the Cys-tRNA^{GCA} tRNA half.

5.3.4 Virus-dependent or Apoptosis-dependent?

As seen in Chapter 3, the flow cytometry experiment, although it only contained one replicate of each condition, provides strong evidence that Poly(I:C) treatment induces apoptosis in a large proportion of the cells. This experiment was conducted in DLD-1 cells rather than MCF7 cells, and DLD1 cells appear to show a slower phenotypic response to Poly(I:C) than MCF7 cells. This indicates that the proportion of MCF7 cells undergoing apoptosis after the same treatment is likely to be even higher. In RNA derived from the same samples as those analysed by flow cytometry, Poly(I:C) treatment, like many other stress treatments, induces the cleavage of a number of tRNAs to produce tRNA halves. Another apoptosis inducer, Hy-

gromycin, did not produce detectable cleavage products of Gln-tRNA^{TTG}. This fits with data by Thompson *et al.*, who found that oxidative stress induced by hydrogen peroxide treatment induced His-tRNA^{GTG} cleavage, but a number of other apoptosis inducers did not (Thompson et al., 2008). Met-tRNA, however, is cleaved after Hygromycin treatment. Our NGS data, along with the corresponding northern blots, suggest that apoptosis-induced tRNA cleavage patterns are tRNA isotype specific. PBS treatment, which has been shown by Fu *et al.* to considerably increase tRNA half levels for a number of different tRNAs (Fu et al., 2009), also did not produce any detectable Gln-tRNA^{TTG} halves in these experiments.

Most cancer cell lines have obtained mutations during the process of transformation, on top of the mutations accrued that are characteristic of the cancers from which these cells originate. MCF7 cells have increased proliferation rates as a result of an inactivation of caspase 3, an effector caspase that is key in apoptosis. For this reason the DLD1 cell line was utilized for the flow cytometry analysis. Any conclusions made on apoptosis-induced cleavage of RNA in MCF7 cells should be interpreted with caution. The cleavage of tRNAs during apoptosis requires further investigation, and should be conducted in the DLD-1 cell line using an inducer such as Staursporine, which has been used to induce apoptosis in NIH/3T3 cells in this project.

In 1999 Rutjes *et al.* found that Y RNAs are cleaved in an apoptosis-dependent manner (Rutjes et al., 1999b), and the data presented here confirms these findings. The other sRNAs, however, have not been examined for apoptosis dependence. Numerous experiments using Hygromycin were conducted to determine whether these sRNAs are produced after Poly(I:C) treatment specifically, or under more general apoptosis conditions. These sRNAs were undetectable in Hygromycin-treated cells to date (data not shown). The full investigation of all 30mers during apoptosis was hindered by the fact that Hygromycin is a less potent inducer of apoptosis than Poly(I:C), unless used at very high concentrations. Given more time, validation experiments would be repeated in a number of cell lines with another apoptosis inducer.

No Poly(I:C)-induced sRNAs derived from mRNAs could be validated, suggesting that sRNAs of this size class may be solely of ncRNA origin. Interestingly, the only mRNA-derived 30mer that could be validated appeared to be unaltered in the presence of Poly(I:C), possibly even decreasing in concentration in response to this dsRNA mimic. Do the sRNA sequences identified in this dataset represent a ncRNA-derived class of 30mers that are entirely up-regulated in the presence of viral nucleic acid? It seems likely that this is the case, with a small number of 30mers, such as the ZMYM2-derived sRNA, acting as the exception that proves the rule.

The high overall read levels for each of the scaRNA- and YRNA- derived 30mers

can be explained by very high levels of one particular sequence in each class present in the Poly(I:C)-treated sample. Surprisingly, very few *bona fide* microRNA sequences were increased in the Poly(I:C)-treated sample over the control, despite the fact that microRNAs are known to be involved in cellular response to viral infection (Pedersen et al., 2007). One such miRNA, miR-1246, was up-regulated in the Poly(I:C)-treated sample in our dataset and Dr. Nicolas further identified it in Argonaute 2 immunoprecipitates, indicating that it is a functional miRNA. This miRNA has recently been identified as p53-regulated and pro-apoptotic (Zhang et al., 2011b). This finding matches with our data and gives us greater confidence that our NGS data is accurate. A number of other miRNA sequences were detected in the Poly(I:C)-treated sample, but these sequences actually mapped to Y RNAs. Our lab has shown that these sequences are not miRNAs (Nicolas et al., 2012).

The Poly(I:C)-treated and control NGS data were based upon a standard Illumina sequencing method, utilizing the provided illumina 5' and 3' adapters. As a result, an inherent bias was introduced into the data as certain sRNAs, once ligated to the 3' adapter, form preferred secondary structures and are more likely to be ligated to the 5' adapter than less favored sequences. This can be seen by the fact that the predominant human Y5-derived 30mer sequence investigated in Chapter 3 can be detected by northern blot after only an hour, while it is only present at around 800 copies per million in the Poly(I:C) NGS dataset; this is the result of a preference by the RNA ligase enzyme for other sequences over this one. With such a dramatic difference in overall levels of 30mers between the control and test conditions, it is therefore more likely that the sequencing bias will affect interpretation of the differential expression of individual sRNA sequences. In the Dalmay lab, many efforts have been made to reduce this sequencing bias through the use of specially designed 'HD' adapters that increase the ability of the RNA ligase to act upon all sequences equally (Sorefan et al., 2012). These 'HD' adapters were utilized in the new NGS sample libraries prepared from MCF7 and SW1353 samples.

Wild type MCF7 cells do undergo apoptosis, but not as efficiently as those transfected with caspase 3. It is therefore somewhat puzzling that these cells show such a dramatic response to Poly(I:C). A particularly interesting follow-up experiment would be to transfect wild type MCF7 cells with a caspase 3 construct, and determine whether this has any effect on 30mer production in the presence of Poly(I:C). The flow cytometry experiments, combined with northern blot analysis of the corresponding total RNA samples, showed many of the cells were undergoing apoptosis in response to Poly(I:C) treatment. Although the levels of cleavage of Met-tRNA and hY5 RNA correlated with the levels of apoptosis, their apoptosis-dependence had not yet been proved.

Work by Xue *et al.* showed that, although MCF7 cells take longer to undergo

apoptosis when caspase 3 is absent, they are still reproductively dead and are committed to the apoptotic process (Xue et al., 2003b). Perhaps the cleavage of numerous classes of ncRNAs is an early event in apoptosis and these sRNAs accumulate to higher levels in MCF7 cells as a result of the delay in the apoptotic process.

Poly(I:C) was initially used in this project as a viral mimic. Further work needs to be done to determine if any or all of the sRNAs detected here are apoptosis or virus dependent. Alongside experiments in a number of other cell lines using a range of apoptosis inducers, experiments should be conducted to infect cells with viruses under physiological conditions. This would allow us to determine whether 30mer biogenesis is a true physiological response to viral infection.

5.3.5 Why are 30mers produced?

ANG cleaves tRNAs during stress to produce tRNA halves. 5' halves from Ala and Cys inhibit cap-dependent translation initiation (Emara et al., 2010) by binding to the eIF4F complex, but IRES-dependent translation initiation appears to be unaffected. This mode of translation regulation is likely to be vital for the cell's survival. In our NGS dataset only the 5' Ala-tRNA half is dramatically up-regulated although both Ala-tRNA and Cys-tRNA show up-regulation by northern blot. So are all 30mers produced solely to create 5' Ala- and Cys-tRNA halves, even though these are not the most abundant halves? A run of terminal G residues is thought to be essential for the function of these tRNA halves in blocking translation (Ivanov et al., 2011a). It might, therefore, be expected that a more specific mechanism for their production would have evolved, if only these two tRNA halves were required during apoptosis. In support of this hypothesis, other functions for tRNA fragments have recently been identified; shorter fragments have microRNA-like functions, while the 30-35 nucleotide-sized fragments have been found to act as small guide RNAs. These fragments interact with the long version of the tRNaseZ enzyme (tRNaseZ^L). Intriguingly, the tRNA halves identified with this function, like those involved in inhibiting translation, are mainly derived from the 5' end of the full length tRNA. Perhaps, like Ala-tRNA and Cys-tRNA halves, the 5' monophosphate is necessary for this function (Ivanov et al., 2011a). It seems likely that yet more functions will be identified for these tRNA halves, suggesting that these cleavage products are not simply generated to directly block the translation stage of gene expression. rRNAs fragments were also isolated from tRNaseZ^L complexes; we found many rRNA fragments in our dataset and perhaps these should be investigated further. Full length tRNAs relocate to the nucleus to halt gene expression during cellular stress conditions. It is possible that tRNA cleavage is another method of blocking gene expression, as it is possible that the number of receptors available to transport

tRNAs into the nucleus limits the rate of tRNA relocation and cleavage may remove tRNAs more rapidly. The levels of Met-tRNA cleavage appeared to be lower than for the other tRNAs tested. A number of different versions of Met-tRNA genes are known to be encoded in the human genome to counter-act any mutations that may occur in the initiation region of mRNAs. This variety in the Met-tRNA sequence may explain why such a small proportion of full length Met-tRNA molecules are targeted. It is also possible that other versions of Met-tRNA are targeted by the cleavage machinery, but are not detected by the Met-tRNA specific probe used here.

As well as their relocation to the nucleus, tRNAs have also been found to interact with cytochrome C during apoptosis and prevent caspase 9 activation (Mei et al., 2010). tRNA cleavage may provide a mechanism to allow tRNAs to dissociate from cytochrome C and enable apoptosis to progress. tRNA halves, as a whole, could be produced to either block apoptosis or allow it to progress. Further experiments could be conducted to transfect cells with varying concentrations of synthetic versions of different 5' tRNA halves and to determine the effect on apoptosis, as induced by a chemical such as STS. It is probable that transfecting in these tRNA halves at the point when the cell would naturally produce them is vital to see their physiological effect on this process, so experiments should be set up in parallel where tRNA halves are added at various time points before and after apoptosis induction. The aim would be to determine if any noticeable effect is seen on the progression of apoptosis and the overall survival of the cells.

Peng et al investigated the response of mouse lung cancer cells to viral infection, and the effect on sRNA expression (Peng et al., 2011). This group identified similar cleavage products to those seen in our data, as they found a 30mer produced from mouse Y RNAs. They also identified a cleavage product derived from the SNORD76 gene (Peng et al., 2011). A similar 30mer was identified in our NGS data, but was less differentially expressed than those sRNAs validated here. This snord76 30mer is particularly interesting as it is encoded within the GAS5, lncRNA. GAS5 is up-regulated during apoptosis (Mourtada-Maarabouni et al., 2009). The production of this sRNA may affect the levels of the lncRNA. The overlap between 30mers detected in mouse lung cancer cells infected with severe acute respiratory syndrome coronavirus or influenza virus and our NGS data, which is derived from human breast cancer cells treated with a viral mimic, provides strong evidence that this class of small RNAs is a *bona fide* cellular response to viral infection. This overlap should be further investigated using the new NGS data to identify 30mers conserved across different organisms and cellular stress responses. Perhaps moving the focus away from tRNAs will enable more rapid progress in determining a function for this class of sRNAs, as a whole, in the near future.

6 SUMMARY

The first part of this project involved further characterization of the requirements for hY5 RNA cleavage during apoptosis. Through mutagenesis, a double-stranded stem region close to the cleavage site was identified as a requirement for cleavage. Deletion or alteration of this region to remove the stem structure resulted in an absence of detectable cleavage products. But, when the nucleotides within this region were substituted for nucleotides that maintain the secondary structure, cleavage still occurred. This indicates that the protein complex that cleaves this RNA recognises the secondary structure of this region rather than a particular nucleotide sequence. The effect of the mutations on the secondary structure was predicted using the RNAfold and Mfold prediction programs. As neither of these programs take into account tertiary structure, it is possible that as yet unknown tertiary interactions may occur that are not taken into account when interpreting the results discussed here.

Mutation of the Ro60-binding site within the hY5 RNA also abrogated cleavage, confirming studies previously conducted by Dr. Hall (Hall, 2013). Surprisingly, the Ro60 protein does not appear to be necessary for exogenous hY5 stability, as the full length RNA containing the mutated Ro60-binding site was as abundant as the wild type. Further investigation is necessary to determine whether the difference in the requirement of the Ro60 protein for stability results from the expression from the genome compared to a plasmid, or from the expression in mouse cells as opposed to human cells.

High-throughput mutagenesis experiments allowed analysis of a larger region of the hY5 molecule, with no reliance upon secondary structure predictions. Data on the full length mutants expressed in NIH/3T3 cells, along with any plasmid bias that might have been introduced during cloning and precipitation of the mutant plasmids, is still in preparation. Preliminary analysis of the sRNA libraries prepared from NIH/3T3 cells transfected with mutant pool 1 plasmids highlighted an interesting pattern. In these samples, transfected with plasmids where the mutation site was only a few nucleotides away from the wild type cleavage site, the cleavage position appeared to be shifted. The most abundant mutant sequence reads were 2 nucleotides longer than the wild type sequence read.

A more in depth analysis of the most abundant pool 1 sequence reads involved

grouping them by read start site. From the results of this analysis another pattern emerged. The nucleotide preferences at each of the mutated positions within the 32 nucleotide reads were very different to the preferences in both the 31 and 30 nucleotide reads. Using RNAfold to predict the secondary structures of the full length RNAs that the most abundant reads were derived from, showed the reason for this. For the abundant 30 nucleotide reads, the secondary structure was similar to the wild type structure and was cleaved in the same position (i.e. between the second and third nucleotide within the internal loop). For the abundant 31 nucleotide reads, the secondary structure again was similar to the wild type, although a slight shift was seen in the position of the nucleotides within the internal loop, so that cleavage occurred at the same position within the loop, but led to a sequencing read 1 nucleotide longer than the wild type sequence. For those mutants with nucleotide preferences at the mutation site that were very different to the wild type sequence, the position of the nucleotides within the RNA molecule was shifted even more, so that the nucleotides were moved two positions along the internal loop. This predicted difference in secondary structure offers a convincing explanation for the difference in the start sites of these groups of mutant sequence reads. The secondary structure suggests that, as long as a double-stranded stem of five base pairs within the RNA is conserved, the cleavage machinery is able to bind and to cut the RNA. These mutations indicate that, when cleavage occurs, the cleavage position is most commonly between the second and third nucleotides within the stem.

The most abundant sRNA reads detected in the samples transfected with mutant pools 2 and 3 were all 30-31 nucleotides in length and were similar in sequence to the wild type RNA. For these mutants the position of the stem was also similar to the wild type RNA, as was the overall secondary structure. The caveat with these interpretations is that only a single secondary structure prediction algorithm was used to predict the structure of each of the mutants and chemical and enzymatic investigations have previously suggested that these predictions are not entirely accurate. As a clear pattern was seen that explains the cleavage position shift in the mutant pool 1 mutants, it seems likely that a similar pattern would occur in the physiological structure of these mutants, even if the secondary structure was not exactly as displayed here.

Along with examining the different cleavage patterns of these mutants, they were also useful for investigating the role of Ro60 in cleavage. Data by Dr. Hall showed that, when the wild type hY5 plasmid was expressed in mouse cells, mRo60 was necessary for cleavage: however, in the same experiment it was clear that mRo60 was not required for stability of hY5 (Hall, 2013). It is possible that the role of Ro60 in cleavage is independent of its binding to Y RNAs. Interpretation of the secondary structures of the most abundant mutants from each mutant pool shows

much variation in the region where Ro60 binds, especially in those mutants where the mutation site is close to the Ro60-binding site. Again inaccuracies in secondary structure prediction could explain this variation, but an alternative hypothesis is that all of these mutations open up the structure in such a way that a small internal loop is formed, which still allows Ro60 to bind. Only a detailed examination of secondary structure and Ro60-binding for each of the mutants can determine whether this is the case.

Although only in the preliminary stages of analysis, the high-throughput method of RNA processing analysis used here shows promise. The production and identification of mutants that are cleaved is rapid and has the benefit of being independent of secondary structure prediction, even though these predictions may be necessary for interpretation of the data. Combination of the full length and plasmid bias data with the sRNA library data may enable identification of full length mutants produced at high levels which are not targeted by the cleavage machinery, providing further information on the requirements for cleavage.

This high-throughput method could be implemented to look at the processing of many other small ncRNAs or, indeed, the RNAs targeted by particular enzyme complexes. The expression of PCR products within mammalian cells could increase the consistency between replicates and therefore increase the number of different mutants that could be investigated at once, making this method more powerful. In the final chapter a more general consequence of apoptosis treatment was investigated, as a number of 30mer RNAs generated under apoptosis conditions were validated. The majority of these 30mers appeared to be up-regulated during apoptosis, suggesting a general function during the apoptotic process. In this chapter 30-35 nucleotide sRNAs derived from snoRNAs, tRNAs and mRNAs during Poly(I:C) treatment were validated, although their production as a more general consequence of apoptosis has not yet been confirmed due to problems with variation in the sensitivity of each cell line to different apoptosis inducers. Analysis of the newest NGS data, which comprises of three replicate samples for each condition and two different human cell lines, should provide many more 30mers to validate and provide more clues as to whether the majority of these 30mers are likely to be functional or simply a by-product of the apoptotic process.

7 APPENDICES

APPENDIX I

Component	Amount
Bacto-tryptone	10 g
NaCl	5 g
Yeast extract	5g
H2O	Up to 1 L

Table 7.1: Bacterial LB medium

Component	Amount
5 x M9 salts	40 mL
2M glucose	2.75 mL
Water agar	157.25 mL

Table 7.2: Bacterial M9 minimal medium

Component	Amount
Na2HPO4	6.78 g
KH2PO4	3.0 g
NaCl	0.5 g
NH4Cl	1.0 g

Table 7.3: 5 x M9 salts

Component	Amount
Bactoagar	3 g
H2O	175.25 mL

Table 7.4: Water agar

APPENDIX II

Component	Amount
0.1 M KOAc	30 mL
0.5 M MnCl2	10 mL
1 M KCl	10 mL
1 M CaCl2	1 mL
Glycerol	15 mL
H2O	Up to 100 mL

Table 7.5: TFBI

Component	Amount
0.1 M Na-MOPs	2.5 mL
1 M CaCl2	1.875 mL
1 M KCl	0.25 mL
Glycerol	3.75 mL
H2O	Up to 25 mL

Table 7.6: TFBII

Component	Amount
Bacto-tryptone	10 g
Yeast extract	2.5g
NaCl	2.92 g
1 M MgCl2	5 mL
H2O	Up to 1 L

Table 7.7: TYM broth

APPENDIX III

Mutant oligo name and description	Sequence 5 to 3
Δ1S F	[Phos]GCTAAAACAGCAAGCTAGTCAAGGGGGGAGACAATGTTAAATC
(Deletion one side of stem forward)	AACTTAACAATAACCCACAACACTCGGACCAACTGTGTTA
Δ1S R	[Phos]CCGGTAACACAGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTT
(Deletion one side of stem reverse)	GATTTAACATTGTCTCCCCCCTTGACTAGCTTGCTGTTTTAGC
Δ2S F	[Phos]GCTAAAACAGCAAGCTAGTCAAGGGGGGAGACAATGTTAAATCAAC
(Deletion both sides of stem forward)	TTA ACAATAACCCTCGGACCAACTGTGTTA
Δ2S R	[Phos]CCGGTAACACAGTTGGTCCGAGGGTTATTGTTAAGTTGATTTAACA
(Deletion both sides of stem reverse)	TTGTCTCCCCCCTTGACTAGCTTGCTGTTTTAGC
Sub1S F	GCTAAAACAGCAAGCTAGTCAAGATGGGTCTCAGGGGGAGACAATGTTAAA
(Change one side of stem forward)	TCAACTTAACAATAACCCACAACACTCGGACCAACTGTGTTA
Sub1S R	CCGGTAACACAGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAA
(Change one side of stem reverse)	CATTGTCTCCCCCTGAGACCCATCTTGACTAGCTTGCTGTTTTAGC
Sub2S F	GCTAAAACAGCAAGCTAGTCAAGATGGGTCTCAGGGGGAGACAATGTTAA
(Change both sides of stem forward)	ATCAACTTAA CAATAACCTGAGCATCTCGGACCAACTGTGTTA
Sub2S R	CCGGTAACACAGTTGGTCCGAGATGCTCAGGTTATTGTTAAGTTGATTTAA
(Change both sides of stem reverse)	CATTGTCTCCCCCTGAGACCCATCTTGACTAGCTTGCTGTTTTAGC
ΔACC F	[Phos]GCTAAAACAGCAAGCTAGTCAAGCGCTGTGGGGGGAGACAA
(ACC deletion forward)	TGTTAAATCAACTTA ACAATAACCCACAACACTCGGACCAACTGTGTTA
ΔACC R	[Phos]CGATTTTGTCGTTCGATCAGTTCGCGACACCCCCCTCTGTTA
(ACC deletion reverse)	CAATTTAGTTGAATTGTTATTGGGTGTTGTGAGCCTGGTTGACACAATGGCC
Δ3' end F	[Phos]GCTAAAACAAGCTAGTCAAGCGCGGTTGTGGGGGGAGACAATG
(3' end deletion forward)	TTAAATCAACTTAACAATAACCCACAACACTCGGACCAACTGTGTTA
Δ3' end R	[Phos]CGATTTTGTTCGATCAGTTCGCGCCAACACCCCCCTCTGTTAC
(3' end deletion reverse)	AATTTAGTTGAATTGTTATTGGGTGTTGTGAGCCTGGTTGACACAATGGCC
ΔRo F	[Phos]GCTAAAACAGCAAGCTCAAGCGCGGTTGTGGGGGGAGACAATGTTAA
(Ro60 binding site deletion forward)	ATCAACTTAACAATAACCCACAACACTCGAACTGTGTTA
ΔRo R	[Phos]CGATTTTGTCGTTCGAGTTCGCGCCAACACCCCCCTCTGTTACAATTT
Ro60 binding site deletion reverse)	AGTTGAATTGTTATTGGGTGTTGTGAGCTTGACACAATGGCC

Table 7.8: Probes used for low-throughput mutagenesis

APPENDIX IV

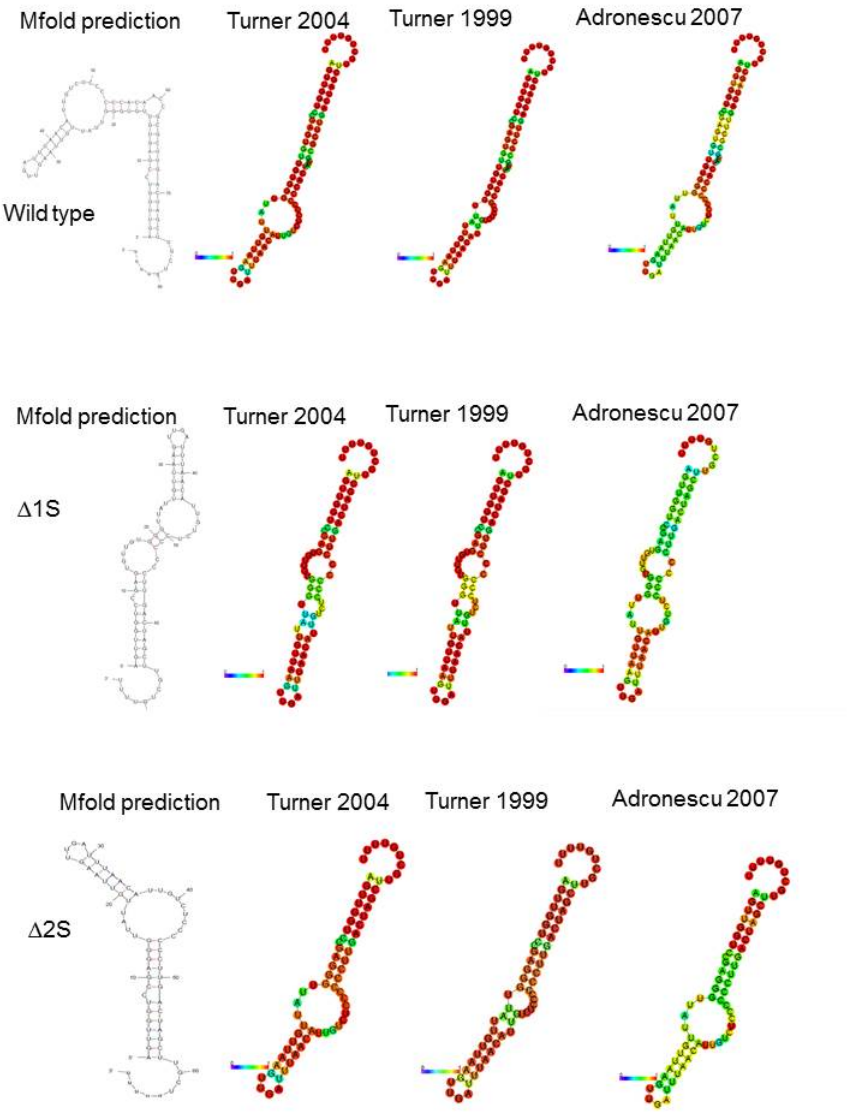


Figure 7.1: The secondary structure predictions of the deletion mutants, as predicted by all algorithms in Mfold and RNAfold.

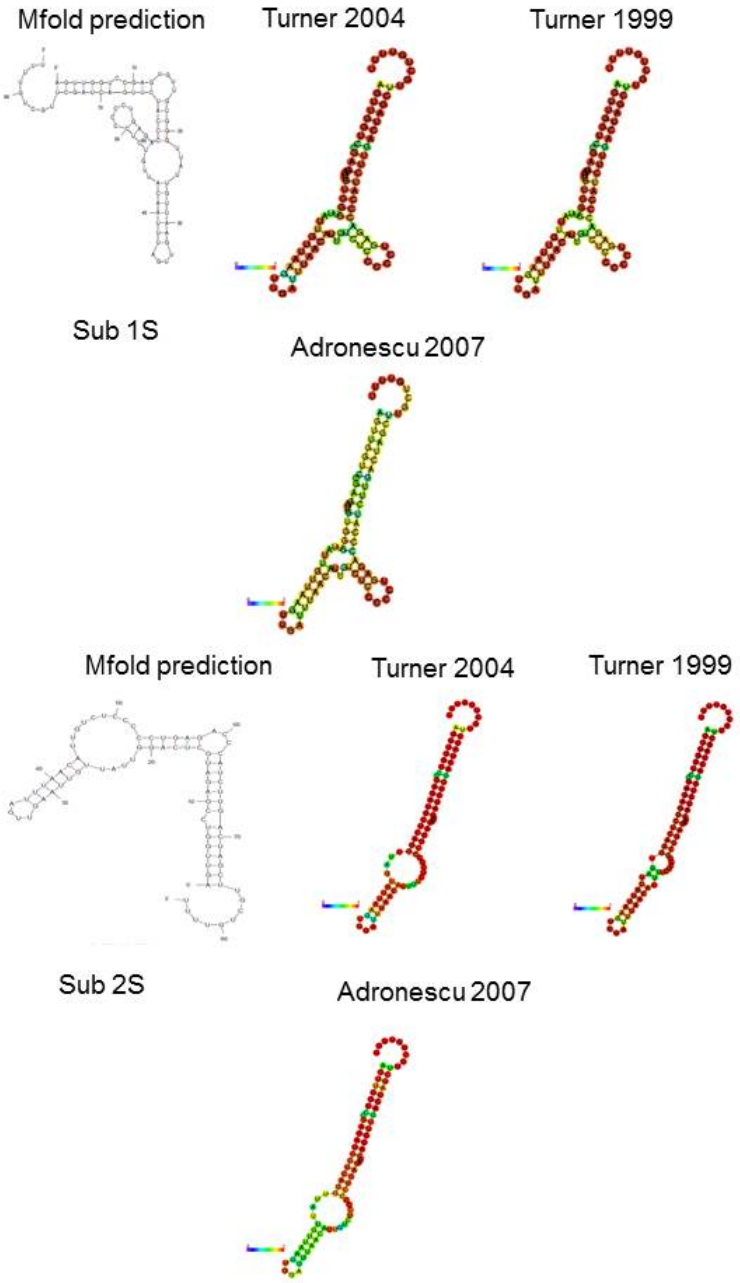


Figure 7.2: The secondary structure predictions of the substitution mutants, as predicted by all algorithms in Mfold and RNAfold.

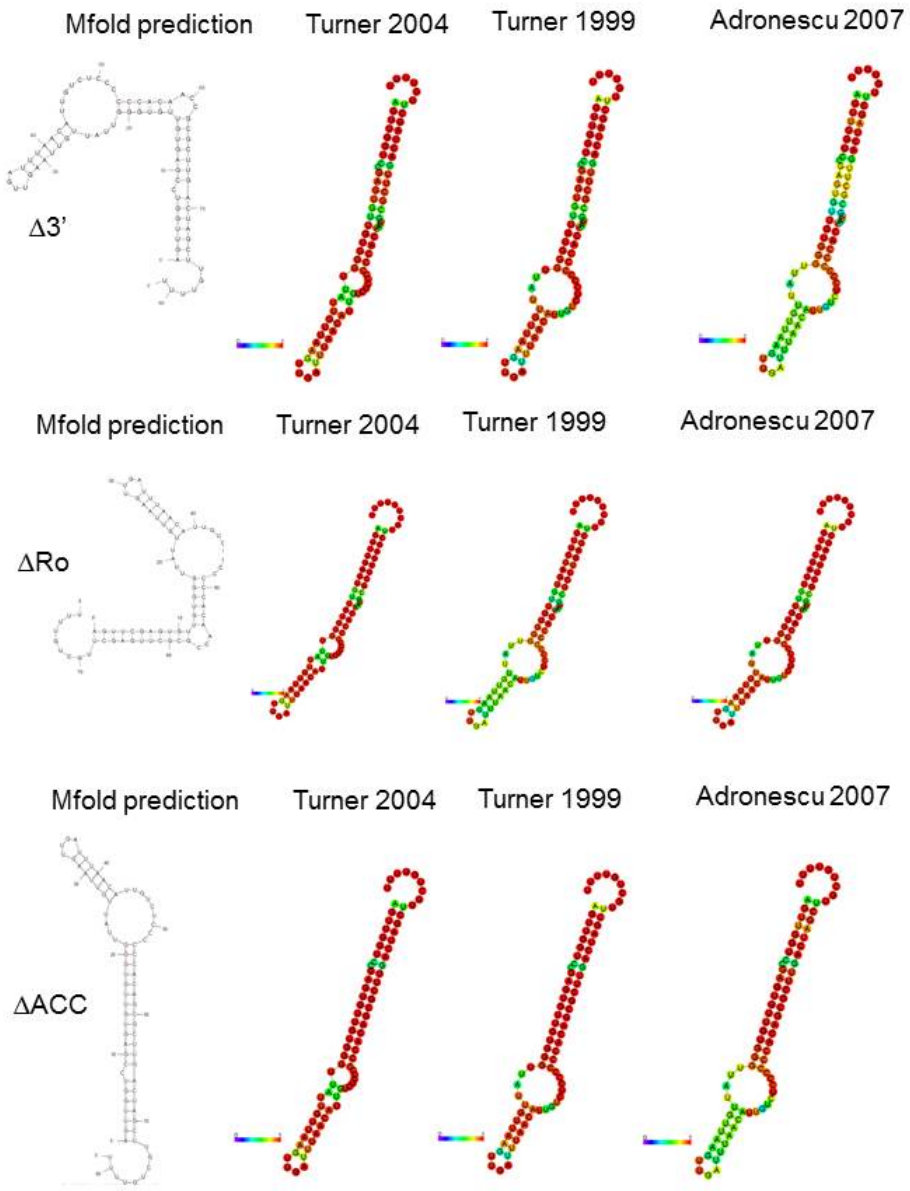


Figure 7.3: The secondary structure predictions of the other mutants, as predicted by all algorithms in Mfold and RNAfold.

APPENDIX V

	Q1-UL		Q1-UR		Q1-LL		Q1-LR	
	% of This Plot	Median FL1-A	% of This Plot	Median FL1-A	% of This Plot	Median FL1-A	% of This Plot	Median FL1-A
8 houra Poly(I:C)	0.07%	19,482.00	1.56%	1,078,639.00	10.94%	11,313.00	87.44%	521,748.00
Unlabelled	0.40%	10,120.00	0.09%	50,285.00	99.06%	2,982.00	0.46%	43,386.00
Control	1.14%	15,385.50	4.63%	511,057.50	74.08%	8,766.00	20.15%	90,405.50
Untreated	0.31%	20,265.00	4.46%	564,862.00	81.24%	8,877.00	13.98%	64,864.00
Lipofectamine	0.32%	18,692.50	4.46%	605,010.00	82.08%	9,266.00	13.14%	54,613.00
20 mins PBS	1.17%	19,446.00	3.07%	169,063.00	88.74%	8,578.50	7.02%	58,873.00
32 hours Hygromycin	0.25%	19,791.00	14.46%	453,772.00	13.61%	21,062.00	71.69%	56,385.50

Figure 7.4: The flow cytometry percentages for cells undergoing apoptosis. After gating, the percentage of cells undergoing apopotsis, healthy cells and cells undergoing necrosis was calculated.

APPENDIX VI

Sequence	Chr	Start site	End site	Annotation	Antype	sense	Hits	Control ave	Poly(I:C) ave	Log fold change
AGTTGGTCCGAGTGTTGTGGGTTATTGTTAA	7	148638580	148638610	hY5	Y RNA	sense	1	6.821363829	831.8223515	6.734444783
CAGAAGTGATGAATTGATCAGATAGACGAGG	1	109643158	109643188	ScaRNA2	SnoRNA	sense	1	0.974480547	193.7189851	6.62377659
CGGCCACTGATTATCGAGGCGATTCTGATCT	1	109643202	109643232	ScaRNA2	SnoRNA	sense	1	1.315548739	223.7381198	6.60074692
GGTCCCATGGTGTAATGGTTAGCACTCTGGA	17	47269890	47269920	trna16-GlnTTG	tRNA	sense	2	2.825993586	327.9056414	6.425695488
GGTCCCATGGTGTAATGGTTAGCACTCTGGA	6	28557156	28557186	trna64-GlnTTG	tRNA	sense	2	2.825993586	327.9056414	6.425695488
ATGCGATGATGAGTGAAGTAGAGCCTGACCT	17	33900676	33900706	SNORD7	SnoRNA	antisense	1	0.389792219	114.5674091	6.377721585
GGTAGCGTGGCCGAGCGGTCTAAGGCGCTGG	5	180524525	180524555	trna16-LeuAAG	tRNA	sense	9	14.90955237	1295.769764	6.348885303
GGTAGCGTGGCCGAGCGGTCTAAGGCGCTGG	5	180524525	180524555	trna19-LeuAAG	tRNA	sense	9	14.90955237	1295.769764	6.348885303
CCCCCACAACCGCGCTTGACTAGCTTGCTGT	7	148638630	148638660	hY5	Y RNA	sense	1	10.96290615	936.0460887	6.291480197
CCCCCCACAACCGCGCTTGACTAGCTTGCTG	7	148638629	148638659	hY5	Y RNA	sense	1	6.675191747	221.2084175	4.855566865
GGGGGTGTAGCTCAGTGGTAGAGCGCGTGCT	6	28806262	28806292	trna102-AlaAGC	tRNA	sense	6	4.677506626	160.4393452	4.829590933
GGGGGTATAGCTCAGTGGTAGAGCATTTGAC	17	37310028	37310058	trna27-CysGCA	tRNA	sense	6	0.828308465	1.911330672	0.671169215
GAGACAGCAGAGTGGCGCAGCGGAAGCGTGC	6	26313398	26313428	trna171-MetCAT	tRNA	sense	1	0	0.112431216	0.153716134

Figure 7.5: The abundant reads from the NGS data form MCF7 Poly(I:C)-treated and untreated samples.
The annotation, start and end positions and the number of hits for each sequence read. The normalised read numbers and log fold change between the untreated and Poly(I:C)-treated sample is also displayed. Colours indicate ncRNA type.

8 PUBLICATIONS

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

Hall, A.E.; Turnbull, C.; and Dalmay, T. Y RNAs: recent developments. *BioMolecular Concepts*, 4: 103-111, 2013.

Bibliography

Abbas, Y.M.; Pichlmair, A.; Górna, M.W.; Superti-Furga, G.; and Nagar, B. Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature*, 494:60–64, 2013.

Affymetrix ENCODE Transcriptome Project, and Cold Spring Harbor Laboratory ENCODE Transcriptome Project, . Post-transcriptional processing generates a diversity of 5' modified long and short RNAs. *Nature*, 457:1028–1032, 2009.

Ambros, V.; Lee, R.C.; Lavanway, A.; Williams, P.T.; and Jewell, D. MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. *Current Biology*, 13:807–818, 2003.

Andersen, K.L. and Collins, K. Several RNase T2 enzymes function in induced tRNA and rRNA turnover in the ciliate Tetrahymena. *Molecular Biology of the Cell*, 23:36–44, 2012.

Andres, M.E.; Burger, C.; Peral-Rubio, M.J.; Battaglioli, E.; and Anderson, M.E. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 96:9873–9878, 1999.

Angelov, D.; Bondarenko, V.A.; Almagro, S.; et al. Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. *The EMBO Journal*, 25:1669–1679, 2006.

Aravin, A. A. et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442:203–207, 2006.

Aravin, A.A.; Lagos-Quintana, M.; Yalcin, A.; et al. The small RNA profile during Drosophila melanogaster development. *Developmental Cell*, 5:337–350, 2003.

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

Bandziulis, R.J.; Swanson, M.S.; and Dreyfuss, G. RNA-binding proteins as developmental regulators. *Genes & Development*, 3:431–437, 1989.

Beadle, G.W. and Tatum, E.L. Genetic control of biochemical reactions in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America*, 27:499–506, 1941.

Belisova, A.; Semrad, K.; Mayer, O.; et al. RNA chaperone activity of protein components of human Ro RNPs. *RNA*, 11:1084–1094, 2005.

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

Boelens, W.C.; Palacios, I.; and Mattaj, I.W. Nuclear retention of RNA as a mechanism for localization. *RNA*, 1:273–283, 1995.

Boire, G. and Craft, J. Human Ro ribonucleoprotein particles: characterization of native structure and stable association with the La polypeptide. *Journal of Clinical Investigation*, 85:1182–1190, 1990.

Boire, G.; Gendron, M.; Monast, N.; Bastin, B.; and Menard, H. A. 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, 1995.

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

Boria, I.; Gruber, A.R.; Tanzer, A.; et al. Nematode sbRNAs: homologs of vertebrate Y RNAs. *Journal of Molecular Evolution*, 70:346–358, 2010.

Bouche, G.; Caizergues-Ferrer, M.; Bugler, B.; and Amalric, F. Interrelations between the maturation of a 100 kDa nucleolar protein and pre rRNA synthesis in CHO cells. *Nucleic Acids Research*, 12:3025–3035, 1984.

Bouffard, P.; Barbar, E.; Brière, F.; and Boire, G. Interaction cloning and Characterization of RoBPI, a novel protein binding to human Ro ribonucleoproteins. *RNA*, 6:66–78, 2000.

Boulanger, C.; Chabot, B.; Ménard, H.A.; and Boire, G. Autoantibodies in human anti-Ro sera specifically recognize deproteinized hY5 Ro RNA. *Clinical and Experimental Immunology*, 99:29–36, 1995.

Brameier, M.; Herwig, A.; Reinhardt, R.; Walter, L.; and Gruber, J. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Research*, 39:675–686, 2011.

Brennecke, J.; Aravin, A.A.; Stark, A.; et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, 128:1089–1103, 2007.

Burns, K.; Duggan, B.; Atkinson, E.A.; et al. Modulation of gene expression by calreticulin binding to the glucocorticoid receptor. *Nature*, 367:476–480, 1994.

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

Casciola-Rosen, L.A.; Miller, D.K.; Anhalt, G.J.; and Rosen, A. Specific cleavage of the 70-kDa protein component of the U1 small nuclear ribonucleoprotein is a characteristic biochemical feature of apoptotic cell death. *Journal of Biological Chemistry*, 269:30757–30760, 1994b.

Cella, M.; Salio, M.; Sakakibara, Y.; et al. Maturation, activation, and protection of dendritic cells induced by double-stranded RNA. *The Journal of Experimental Medicine*, 189:821–829, 1999.

Chambers, J.C.; Kenan, D.; Martin, B.J.; and Keene, J.D. Genomic structure and amino acid sequence domains of the human La autoantigen. *The Journal of Biological Chemistry*, 263:18043–18051, 1988.

Chan, E.K.; Francoeur, A.M.; and Tan, E.M. Epitopes, structural domains, and asymmetry of amino acid residues in SS-B/La nuclear protein. *The Journal of Immunology*, 136:3744–3749, 1986.

Chen, X.; Quinn, A.M.; and Wolin, S.L. Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation. *Genes & Development*, 14:777–782, 2000.

Chen, X.; Smith, J.D.; Shi, H.; et al. The Ro autoantigen binds misfolded U2 small nuclear RNAs and assists mammalian cell survival after UV Irradiation. *Current Biology*, 13:2206–2211, 2003.

Chen, X.; Wurtmann, E.J.; Van Batavia, J.; et al. An ortholog of the Ro autoantigen functions in 23S rRNA maturation in D. radiodurans. *Genes & Development*, 21: 1328–1339, 2007.

Chen, X.; Taylor, D.W.; Fowler, C.C.; et al. An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. *Cell*, 153:166–177, 2013.

Cheng, J.; Guo, J.-M.; Xiao, B.-X.; et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clinica Chimica Acta*, 412:1621–1625, 2011.

Cheng, S.T.; Nguyen, T.Q.; Yang, Y.S.; Capra, J.D.; and Sontheimer, R.D. Calreticulin binds hYRNA and the 52-kDa polypeptide component of the Ro/SS-A ribonucleoprotein autoantigen. *Journal of Immunology*, 156:4484–4491, 1996.

Christov, C.P.; Gardiner, T.J.; Szuts, D.; and Krude, T. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Molecular and Cellular Biolo*, 26:6993–7004, 2006.

Cole, C.; Sobala, A.; Lu, C.; et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA*, 15:2147–2160, 2009.

Collins, K.L. and Kelly, T.J. Effects of T antigen and replication protein A on the initiation of DNA synthesis by DNA polymerase alpha-primase. *Molecular Cell Biology*, 11:2108–2115, 1991.

Crosio, C.; Boyl, P.P.; Loreni, F.; Pierandrei-Amaldi, P.; and Amaldi, F. La protein has a positive effect on the translation of TOP mRNAs in vivo. *Nucleic Acids Research*, 28:2927–2934, 2000.

Cui, L.; Lou, Y.; Zhang, X.; et al. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. *Clinical Biochemistry*, 44:1050–1057, 2011.

Damgaard, C.K. and Lykke-Andersen, J. Translation coregulation of 5'TOP mRNAs by TIA-1 and TIAR. *Genes & Development*, 25:2057–2068, 2011.

Daniely, Y. and Borowiec, J.A. Formation of a complex between nucleolin and replication protein A after cell stress prevents initiation of DNA replication. *Journal of Cell Biology*, 149:799–810, 2000.

Das, B.; Bulter, J.S.; and Sherman, F. Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 23:5502–5515, 2003.

Decca, M.B.; Carpio, M.A.; Bosc, C.; et al. Post-translational arginylation of calreticulin: A new isospecies of calreticulin component of stress granules. *Journal of Biological Chemistry*, 282:8237–8245, 2007.

Decker, C.J. and Parker, R. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes & Development*, 7:1632–1643, 1993.

Dedhar, S. Novel functions for calreticulin: interaction with integrins and modulation of gene expression? *Trends in Biochemical Sciences*, 19:269–271, 1994.

Deng, J.S.; Takasaki, Y.; and Tan, E.M. Nonhistone nuclear antigens reactive with autoantibodies. Immunofluorescence studies on distribution in synchronized cells. *The Journal of Cell Biology*, 91:654–660, 1981.

DeNoto, F.M.; Moore, D.D.; and Goodman, H.M. Human growth hormone DNA sequence and mRNA structure: possible alternative splicing. *Nucleic Acids Research*, 9:3719–3730, 1981.

Deveraux, Q.L.; Takahashi, R.; Salvesen, G.S.; and Reed, J.C. X-linked IAP is a direct ininhibitor of cell-death proteases. *Nature*, 388:300–304, 1997.

Early, P.; Rogers, J.; Davis, M.; et al. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell*, 20:313–319, 1980.

Emara, M.M.; Ivanov, P.; Hickman, T.; et al. Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *Journal of Biological Chemistry*, 285:10959–10968, 2010.

Ender, C.; Krek, A.; Friedländer, M.R.; et al. A human snoRNA with microRNA-like functions. *Molecular Cell*, 32:519–528, 2008.

Engels, B.; Jannot, G.; Remenyi, J.; Simard, M.J.; and Hutvagner, G. Polypyrimidine tract binding protein (hnRNP I) is possibly a conserved modulator of miRNA-mediated gene regulation. *PLoS ONE*, 7:e33144, 2012.

Fabini, G.; Rutjes, S.A.; Zimmermann, C.; Pruijn, G.J.; and Steiner, G. Analysis of the molecular composition of Ro ribonucleoprotein complexes. Identification of novel Y RNA-binding proteins. *European Journal of Biochemistry*, 267:2778–2789, 2000.

Fabini, G.; Raijmakers, R.; Hayer, S.; et al. The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the Ro ribonucleoprotein-associated Y RNAs in vitro and in vivo. *Journal of Biological Chemistry*, 276:20711–20718, 2001.

Fan, H.; Sakulich, A.L; Goodier, J.L; et al. Phosphorylation of the human La antigen on serine 366 can regulate recycling of RNA polymerase III transcription complexes. *Cell*, 88:707–715, 1997.

Fan, H.; Goodier, J.L.; Chamberlain, J.R.; Engelke, D.R.; and Maraia, R.J. 5' processing of tRNA precursors can be modulated by the human La antigen phosphoprotein. *Molecular and Cellular Biology*, 18:3201–3211, 1998.

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

Feng, F.; Yuan, L.; Wang, Y.E.; et al. Crystal structure and nucleotide selectivity of human IFIT5/ISG58. *Cell Research*, 23:1055–1058, 2013.

Fennell, R.H.; Rodnan, G.P.; and Vazquez, J.J. Variability of tissue-localizing properties of serum from patients with different disease states. *Laboratory Investigation*, 2:24–31, 1962.

Flikke, M.; Kjeldsberg, E.; and Lahelle, O. Interference with cell viability and poliovirus multiplication by polyinosinic-polycytidylic acid. *Journal of General Virology*, 9:111–117, 1970.

Fouraux, M.A.; Bouvet, P.; Verkaart, S.; van Venrooij, W. J.; and Pruijn, G.J. Nucleolin associates with a subset of the human Ro Ribonucleoprotein complexes. *Journal of Molecular Biology*, 320:475–488, 2002.

Francoeur, A.M. and Mathews, M.B. Interaction between VA RNA and the lupus antigen La: formation of a ribonucleoprotein particle in vitro. *Proceedings of the National Academy of Sciences*, 79:6772–6776, 1982.

Fu, H.; Feng, J.; Liu, Q.; et al. Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Letters*, 583:437–442, 2009.

Fuchs, G.; Stein, A.J.; Fu, C.; Reinisch, K.M.; and Wolin, S.L. Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. *Nature Structural & Molecular Biology*, 13:1002–1009, 2006.

Galabru, J. and Hovanessian, A.G. Two interferon-induced proteins are involved in the protein kinase complex dependent on double-stranded RNA. *Cell*, 43:685–694, 1985.

Garcia, E.L.; Onafuwa-Nuga, A.; Sim, S.; et al. Packaging of host mY RNAs by murine leukemia virus may occur early in Y RNA biogenesis. *Journal of Virology*, 83:12526–12534, 2009.

Gardiner, T.J.; Christov, C.P.; Langley, A.R.; and Krude, T. A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. *RNA*, 15: 1375–1385, 2009.

Gendron, M.; Roberge, D.; and Boire, G. 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, 2001.

Gerard, M.A.; Myslinski, E.; Chylak, N.; et al. The scaRNA2 is produced by an independent transcription unit and its processing is directed by the encoding region. *Nucleic Acids Research*, 38:370–381, 2010.

Girard, A.; Sachidanandam, R.; Hannon, G.J.; and Carmell, M.A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442:199–202, 2006.

Gocke, C.B. and Yu, H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS ONE*, 3:e3255, 2008.

Gottlieb, E. and Steitz, J.A. The RNA binding protein La influences both the accuracy and the efficiency of RNA polymerase III transcription in vitro. *The EMBO Journal*, 8:841–850, 1989a.

Gottlieb, E. and Steitz, J.A. Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. *The EMBO Journal*, 8:851–861, 1989b.

Green, C.D.; Long, K.S.; Shi, H.; and Wolin, S.L. Binding of the 60-kDa Ro autoantigen to Y RNAs: evidence for recognition in the major groove of a conserved helix. *RNA*, 4:750–765, 1998.

Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Research*, 32:D109–D111, 2004.

Grishok, A.; Pasquinelli, A.E.; Conte, D.; et al. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell*, 106:23–34, 2001.

Grivna, S.T.; Beyret, E.; Wang, Z.; and Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development*, 20:1709–1714, 2006.

Gunawardane, L.S.; Saito, K.; Nishida, K.M.; et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science*, 315:1587–1590, 2007.

Guttman, M.; Amit, I.; Garber, M.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458: 223–227, 2009.

Hagenbüchle, O.; Tosi, M.; Schibler, U.; et al. Mouse liver and salivary gland alpha-amylase mRNAs differ only in 5' non-translated sequences. *Nature*, 289:643–646, 1981.

Hall, A.E. *Biogenesis of Y RNA-derived small RNAs.* PhD thesis, University of East Anglia, 2013.

Hall, A.E.; Turnbull, C.; and Dalmay, T. Y RNAs: recent developments. *BioMolecular Concepts*, 4:1226–1230, 2013.

Hamilton, A.; Voinnet, O.; Chappell, L.; and Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *The EMBO Journal*, 21:4671–4679, 2002.

Hamilton, A.J. and Baulcombe, D.C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286:950–952, 1999.

Hammond, S.M.; Bernstein, E.; Beach, D.; and Hannon, G.J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404:293–296, 2000.

Hangauer, M.J.; Vaughn, I.W.; and McManus, M.T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genetics*, 9:e1003569, 2013.

Hendrick, J.P.; Wolin, S.L.; Rinke, J.; Lerner, M.R.; and Steitz, J.A. 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, 1981.

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

Holcik, M and Korneluk, R.G. 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, 2000.

Houseley, J. and Tollervey, D. The many pathways of RNA degradation. *Cell*, 136: 763–776, 2009.

Houwing, S.; Kamminga, L.M.; Berezikov, E.; et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, 129:69–82, 2007.

Huichalaf, C.; Schoser, B.; Schneider-Gold, C.; et al. Reduction of the rate of protein translation in patients with Myotonic Dystrophy 2. *Journal of Neuroscience*, 29: 9042–9049, 2009.

Itine, R.V.; Sakulich, A.L.; Koduru, S.B.; et al. Control of transfer RNA maturation by phosphorylation of the human La antigen on serine 366. *Molecular Cell*, 6:339–348, 2000.

Itoh, Y.; Itoh, K.; Frank, M.B.; and Reichlin, M. Autoantibodies to the Ro/SSA autoantigen are conformation dependent. II: Antibodies to the denatured form of 52 kD Ro/SSA are a cross reacting subset of antibodies to the native 60 kD Ro/SSA molecule. *Autoimmunity*, 14:89–95, 1992.

Ivanov, P.; Emara, M.M.; Villen, J.; Gygi, S.P.; and Anderson, P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Molecular Cell*, 43:613–623, 2011a.

Ivanov, P.; Kedersha, N.; and Anderson, P. Stress puts TIA on TOP. *Genes & Development*, 25:2119–2124, 2011b.

Jacks, A.; Babon, J.; Kelly, G.; et al. Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure*, 11:833–843, 2003.

Jöchl, C.; Rederstorff, M.; Hertel, J.; et al. Small ncRNA transcriptome analysis from Aspergillus fumigatus suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Research*, 36:2677–2689, 2008.

Kapranov, P.; Cheng, J.; Dike, S.; et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316:1484–1488, 2007.

Katibah, G.E.; Lee, H.J.; Huizar, J.P.; et al. tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. *Molecular Cell*, 49:743–750, 2013.

Kato, N.; Hoshino, H.; and Harada, F. Nucleotide sequence of 4.5S RNA (C8 or hY5) from HeLa cells. *Biochemical and Biophysical Research Communications*, 108:363–370, 1982.

Kawaji, H.; Nakamura, M.; Takahashi, Y.; et al. Hidden layers of human small RNAs. *BMC Genomics*, 9:157, 2008.

Kedersha, N.L. and Rome, L.H. Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *The Journal of Cell Biology*, 103:699–709, 1986.

Kelekar, A.; Saitta, M.R.; and Keene, J.D. Molecular composition of Ro small ribonucleoprotein complexes in human cells. Intracellular localization of the 60- and 52-kD proteins. *Journal of Clinical Investigation*, 93:1637–1644, 1994.

Ketting, R.F.; Haverkamp, T.H.; van Luenen, H.G.; and Plasterk, R.H. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner Syndrome Helicase and RNaseD. *Cell*, 99:133–141, 1999.

Ketting, R.F.; Fischer, S.E.; Bernstein, E.; et al. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & Development*, 15:2654–2659, 2001.

Kirsch, D.G.; Doseff, A.; Nelson Chau, B.; Sim, D.-S.; and de Souza-Pinto, N.C. Caspase-3-dependent cleavage of Bcl-2 promotes release of cytochrome C. *The Journal of Biological Chemistry*, 274:21155–21161, 1999.

Kiss-Laszlo, Z.; Henry, Y.; Bachellerie, J.-P.; Caizergues-Ferrer, M.; and Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, 85:1077–1088, 1996.

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

Krude, T.; Christov, C.P.; Hyrien, O.; and Marheineke, K. Y RNA functions at the initiation step of mammalian chromosomal DNA replication. *Journal of Cell Science*, 122:2836–2845, 2009.

Kucera, N.J.; Hodsdon, M.E.; and Wolin, S.L. An intrinsically disordered C terminus allows the La protein to assist the biogenesis of diverse noncoding RNA precursors. *Proceedings of the National Academy of Sciences*, 108:1308–1313, 2011.

Labbé, J.-C.; Burgess, J.; Rokeach, L.A.; and Hekimi, S. ROP-1, an RNA quality-control pathway component, affects Caenorhabditis elegans dauer formation. *Proceedings of the National Academy of Sciences*, 97:13233–13238, 2000.

Lagos-Quintana, M.; Rauhut, R.; Lendeckel, W.; and Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–858, 2001.

Lagos-Quintana, M.; Rauhut, R.; Meyer, J.; Borkhardt, A.; and Tuschl, T. New microRNAs from mouse and human. *RNA*, 9:175–179, 2003.

Lai, E.C.; Tomancak, P.; Williams, R.W.; and Rubin, G.M. Computational identification of Drosophila microRNA genes. *Genome Biology*, 4:R42, 2003.

Langley, A.R.; Chambers, H.; Christov, C.P.; and Krude, T. 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, 2010.

Larson, V.M.; Clark, W.R.; and Hilleman, M.R. Influence of synthetic (poly I:C) and viral double-stranded ribonucleic acids on adenovirus 12 oncogenesis in hamsters. *Proceedings of the Society for Experimental Biology and Medicine*, 131:1002–1011, 1969.

Lau, N.C.; Lim, L.P.; Weinstein, E.G.; and Bartel, D.P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, 294: 858–862, 2001.

Lau, N.C.; Seto, A.G.; Kim, J.; et al. Characterization of the piRNA complex from rat testes. *Science*, 313:363–367, 2006.

Laurent, A.G.; Krust, B.; Svab, J.; and Hovanessian, A.G. Characterisation of the interferon-mediated protein kinase by polyclonal antibodies. *Biochemical and Biophysical Research Communications*, 125:1–7, 1984.

Lebreton, A.; Tomecki, R.; Dziembowski, A.; and Sèraphin, B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature*, 456:993–996, 2008.

Lee, J.T.; Davidow, L.S.; and Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics*, 21:400–404, 1999.

Lee, R.C. and Ambros, V. An extensive class of small RNAs in Caenorhabditis elegans. *Science*, 294:862–864, 2001.

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

Lee, S.R. and Collins, K. Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. *Journal of Biological Chemistry*, 280:42744–42749, 2005.

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

Lerner, M.R.; Andrews, N.C.; Miller, G.; and Steitz, J.A. Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences of the United States of America*, 78:805–809, 1981a.

Lerner, M.R.; Boyle, J.A.; Hardin, J.A.; and Steitz, J.A. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science*, 211:400–402, 1981b.

Li, C.; Vagin, V.V.; Lee, S.; et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, 137:509–521, 2009.

Li, Y.; Luo, J.; Zhou, H.; et al. Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote Giardia lamblia. *Nucleic Acids Research*, 36:6048–6055, 2008.

Liang, C.; Xiong, K.; Szulwach, K.E.; et al. Sjögren syndrome antigen B (SSB)/La promotes global microRNA expression by binding microRNA precursors through stem-loop recognition. *Journal of Biological Chemistry*, 288:723–736, 2013.

Lim, L.P.; Lau, N.C.; Weinstein, E.G.; et al. The microRNAs of Caenorhabditis elegans. *Genes & Development*, 17:991–1008, 2003.

Lim, L.P.; Lau, N.C.; Garrett-Engele, P.; et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433:769–773, 2005.

Liu, J.; Lee, P.; Galbiati, F.; Kitsis, R.N.; and Lisanti, M.P. Caveolin-1 expression sensitizes fibroblastic and epithelial cells to apoptosis stimulation. *American Journal of Physiology. Cell Physiology*, 280:C823–C835, 2001.

Lu, Y.; Li, C.; Zhang, K.; et al. Identification of piRNAs in Hela cells by massive parallel sequencing. *BMB Reports*, 43:635–641, 2010.

Malone, C.D.; Brennecke, J.; Dus, M.; et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell*, 137:522–535, 2009.

Maraia, R.; Sakulich, A.L.; Brinkmann, E.; and Green, E.D. Gene encoding human Ro-associated autoantigen Y5 RNA. *Nucleic Acids Research*, 24:3552–3559, 1996.

Maraia, R.J.; Kenan, D.J.; and Keene, J.D. Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III. *Molecular and Cellular Biology*, 14:2147–2158, 1994a.

Maraia, R.J.; Sasaki-Tozawa, N.; Driscoll, C.T.; Green, E.D.; and Darlington, G.J. 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, 1994b.

Martino, L.; Pennell, S.; Kelly, G.; et al. Analysis of the interaction with the hepatitis C virus mRNA reveals an alternative mode of RNA recognition by the human La protein. *Nucleic Acids Research*, 40:1381–1394, 2012.

Matera, A.G.; Frey, M.R.; Margelot, K.; and Wolin, S.L. A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *The Journal of Cell Biology*, 129:1181–1193, 1995.

Meerovitch, K.; Svitkin, Y.V.; Lee, H.S.; et al. La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. *Journal of Virology*, 67:3798–3807, 1993.

Mei, Y.; Yong, J.; Liu, H.; et al. tRNA binds to cytochrome C and inhibits caspase activation. *Molecular Cell*, 37:591–592, 2010.

Meselson, M. and Stahl, F.W. The replication of DNA in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 44: 671–682, 1958.

MiRBase, . URL `http://www.mirbase.org`.

Mitchell, P.; Petfalski, E.; Shevchenko, A.; Mann, M.; and Tollervey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-5' exoribonucleases. *Cell*, 91:457–466, 1997.

Moroianu, J. and Riordan, J.F. Nuclear translocation of angiogenin in proliferating endothelial cells is essential to its angiogenic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 91:1677–1681, 1994.

Mosig, A.; Guofeng, M.; Stadler, B.M.; and Stadler, P.F. Evolution of the vertebrate Y RNA cluster. *Theory in Biosciences*, 126:9–14, 2007.

Mossink, M.H.; van Zon, A.; Scheper, R.J.; Sonneveld, P.; and Wiemer, E.A. Vaults: a ribonucleoprotein particle involved in drug resistance? *Oncogene*, 22:7458–7467, 2003.

Mourtada-Maarabouni, M.; Pickard, M.R.; Hedge, V.L.; Farzaneh, F.; and Williams, G.T. GAS5 a non-protein-coding RNA, control apoptosis and is downregulated in breast cancer. *Oncogene*, 28:195–208, 2009.

Mouse Genome Sequencing Consortium, . Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420:520–562, 2002.

Mullen, T.E. and Marzluff, W.F. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes & Development*, 22:50–65, 2008.

Myers, J.W.; Jones, J.T.; Meyer, T.; and Ferrell, J.E. Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nature Biotechnology*, 21:324–328, 2003.

Nakamura, K.; Bossy-Wetzel, E.; Burns, K.; et al. Changes in endoplasmic reticulum luminal environment affect cell sensitivity to apoptosis. *Journal of Cell Biology*, 150:731–740, 2000.

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

Nicoloso, M.; Qu, L.-H.; Michot, B.; and Bachellerie, J.-P. Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *Journal of Molecular Biology*, 260:178–195, 1996.

O'Brien, C.A. and Harley, J.B. A subset of hY RNAs is associated with erythrocyte Ro ribonucleoproteins. *The EMBO Journal*, 9:3683–3689, 1990.

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

O'Brien, C.A.; Margelot, K.; and Wolin, S.L. Xenopus Ro ribonucleoproteins: members of an evolutionarily conserved class of cytoplasmic ribonucleoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, 90: 7250–7254, 1993.

Ogawa, T.; Tomita, K.; Ueda, T.; et al. A cytotoxic ribonuclease targeting specific transfer RNA anticodons. *Science*, 283:2097–2100, 1999.

Ogawa, Y.; Sun, B.K.; and Lee, J.T. Intersection of the RNA Interference and X-Inactivation pathways. *Science*, 320:1336–1341, 2008.

Ohndorf, U.-M.; Steegborn, C.; Knijff, R.; and Sondermann, P. Contributions of the individual domains in human La protein to Its RNA 3'-end binding activity. *Journal of Biological Chemistry*, 276:27188–27196, 2001.

Opas, M.; Dziak, E.; Fliegel, L.; and Michalak, M. Regulation of expression and intracellular distribution of calreticulin, a major calcium binding protein of non-muscle cells. *Journal of Cellular Physiology*, 149:160–171, 1991.

Orrick, L.R.; Olson, M.O.; and Busch, H. Comparison of nucleolar proteins of normal rat liver and Novikoff hepatoma ascites cells by two-dimensional polyacrylamide gel electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America*, 70:1316–1320, 1973.

Ostareck, D.H.; Ostareck-Lederer, A.; Wilm, M.; et al. mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell*, 89:597–606, 1997.

Page-McCaw, P.S.; Amonlirdviman, K.; and Sharp, P.A. PUF60: a novel U2AF65-related splicing activity. *RNA*, 5:1548–1560, 1999.

Pang, K.C.; Frith, M.C.; and Mattick, J.S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics*, 22:1–5, 2006.

Pasquinelli, A.E.; Reinhart, B.J.; Slack, F.; et al. Conservation of the sequence and temporal expression of let-7 heterochronic Regulatory RNA. *Nature*, 408:86–89, 2000.

Pedersen, I.M.; Cheng, G.; Wieland, S.; et al. Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature*, 449:919–922, 2007.

Peek, R.; Pruijn, G.J.; van der Kemp, A..J.; and van Venrooij, W.J. Subcellular distribution of Ro ribonucleoprotein complexes and thier constituents. *Journal of Cell Science*, 106:929–935, 1993.

Pellizzoni, L.; Lotti, F.; Rutjes, S.A.; and Pierandrei-Amaldi, P. Involvement of the Xenopus laevis Ro60 autoantigen in the alternative interaction of La and CNBP proteins with the 5'UTR of L4 ribosomal protein mRNA. *Journal of Molecular Biology*, 281:593–608, 1998.

Peng, X.; Gralinski, L.; Ferris, M.T.; Frieman, M.B.; and Thomas, M.J. Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. *MBio*, 2: e00198–00211, 2011.

Perreault, J.; Noël, J.-F.; Brière, F.; et al. Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. *Nucleic Acids Research*, 33: 2032–2041, 2005.

Perreault, J.; Perreault, J.-P.; and Boire, G. Ro-associated Y RNAs in metazoans: evolution and diversification. *Molecular Biology and Evolution*, 24:1678–1689, 2007.

Persson, H.; Kvist, A.; Vallon-Christersson, J; et al. The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nature Cell Biology*, 11:1268–1271, 2009.

Peterson, R.C.; Doering, J.L.; and Brown, D.D. Characterization of two Xenopus somatic 5S DNAs and one minor oocyte-specific 5S DNA. *Cell*, 20:131–141, 1980.

Pothof, J.; Verkaij, N.S.; van IJcken, W.; et al. MicroRNA-mediated gene silencing modulates the UV-induced DNA-damage response. *The EMBO Journal*, 28:2090–2099, 2009.

Provost, P.; Dishart, D.; Doucet, J.; et al. Ribonuclease activity and RNA binding of recombinant human Dicer. *The EMBO Journal*, 21:5864–5874, 2002.

Pruijn, G. J.; H., Simons F.; and van Venrooij, W. J. Intracellular localization and nucleocytoplasmic transport of Ro RNP components. *European Journal of Cell Biology*, 74:123–132, 1997.

Pruijn, G.J..; Slobbe, R.J.; and van Venrooij, W.J. Analysis of protein-RNA interactions within Ro ribonucleoprotein complexes. *Nucleic Acids Research*, 19: 5173–5180, 1991.

Query, C.C.; Bentley, R.C.; and Keene, J.D. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, 57:89–101, 1989.

Reddy, R.; Henning, D.; Tan, E.; and Busch, H. Identification of a La protein binding site in a RNA polymerase III transcript (4.5 I RNA). *The Journal of Biological Chemistry*, 258:8352–8356, 1983.

Reinhart, B.J.; Slack, F.J.; Basson, M.; et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403:901–6, 2000.

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

Rosa, M.D.; Gottlieb, E.; Lerner, M.R.; and Steitz, J.A. Striking similarities are exhibited by two small Epstein-Barr virus-encoded ribonucleic acids and the adenovirus-associated ribonucleic acids VAI and VAII. *Molecular and Cellular Biology*, 1:785–796, 1981.

Rutjes, S.A.; Utz, P.; van der Heijden, A.; et al. The La (SSB) autoantigen, a key protein in RNA biogenesis, is dephosphorylated and cleaved early during apoptosis. *Cell Death and Differentiation*, 6:976–986, 1999a.

Rutjes, S.A.; van der Heijden, A.; Utz, P.; Van Venrooij, W.J.; and Pruijn, G.J. Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. *Journal of Biological Chemistry*, 274:24799–24807, 1999b.

Rutjes, S.A.; Lund, E.; van der Heijden, A.; et al. Identification of a novel cis-acting RNA element involved in nuclear export of hY RNAs. *RNA*, 7:741–752, 2001.

Saxena, S.K.; Rybak, S.M.; Davey, R.T.; Youle, R.J.; and Ackerman, E.J. Angiogenin is a cytotoxic, tRNA-specific ribonuclease in the RNase A superfamily. *The Journal of Biological Chemistry*, 267:21982–21986, 1992.

Schaeffer, D.; Tsanova, B.; Barbas, A.; et al. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nature Structural & Molecular Biology*, 16:56–62, 2009.

Scott, M. S.; Avolio, F.; Ono, M.; Lamond, A. I.; and Barton, G. J. Human miRNA Precursors with Box H/ACA snoRNA Features. *PLoS Computational Biology*, 5: e1000507, 2009.

Scott, M.S.; Ono, M.; Yamada, K.; et al. Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic Acids Research*, 40:3676–3688, 2012.

Seila, A.C.; Calabrese, J.M.; Levine, S.S.; et al. Divergent transcription from active promoters. *Science*, 322:1849–1851, 2008.

Seinsoth, S.; Uhlmann-Schiffler, H.; and Stahl, H. Bidirectional DNA unwinding by a ternary complex of T antigen, nucleolin and topoisomerase I. *EMBO Reports*, 4:263–268, 2003.

Shapiro, R and Vallee, B.L. Human placental ribonuclease inhibitor abolishes both angiogenic and ribonucleolytic activities of angiogenin. *Proceedings of the National Academy of Sciences of the United States of America*, 84:2238–2241, 1987.

Sharma, V. and Misteli, T. Non-coding RNAs in DNA damage and repair. *FEBS Letters*, 587:1832–1839, 2013.

Shen, C.K. and Maniatis, T. The organization, structure, and in vitro transcription of Alu family RNA polymerase III transcription units in the human alpha-like globin gene cluster: precipitation of in vitro transcripts by lupus anti-La antibodies. *Journal of Molecular and Applied Genetics*, 1:343–360, 1982.

Shi, H.; O'Brien, C.A.; Van Horn, D.J.; and Wolin, S.L. A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens. *RNA*, 2:769–784, 1996.

Shiroki, K.; Isoyama, T.; Kuge, S.; et al. Intracellular redistribution of truncated La protein produced by poliovirus 3Cpro-mediated cleavage. *Journal of Virology*, 73:2193–2200, 1999.

Sim, S. and Wolin, S.L. Emerging roles for the Ro 60-kDa autoantigen in noncoding RNA metabolism. *Wiley Interdisciplinary Reviews: RNA*, 2:686–699, 2011. ISSN 1757-7012.

Sim, S.; Weinberg, D.E.; Fuchs, G.; et al. 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, 2009.

Sim, S.; Yao, J.; Weinberg, D.E.; et al. The zipcode-binding protein ZBP1 influences the subcellular location of the Ro 60-kDa autoantigen and the noncoding Y3 RNA. *RNA*, 18:100–110, 2011.

Simons, F. H. 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, 1994.

Simons, F.H.; Rutjes, S.A.; van Venrooij, W.J.; and Pruijn, G.J.M. The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. *RNA*, 2:264–273, 1996.

Singh, N.K.; Atreya, C.D.; and Nakhasi, H.L. Identification of calreticulin as a rubella virus RNA binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, 91:12770–12774, 1994.

Sobel, S.G. and Wolin, S.L. Two yeast La motif-containing proteins are RNA-binding proteins that associate with polyribosomes. *Molecular Biology of the Cell*, 10:3849–3862, 1999.

Sorefan, K.; Pais, H.; Hall, A.E.; et al. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence*, 3:4, 2012.

Staikou, E.V.; Routsias, J.G.; Makri, A.A.; et al. Calreticulin binds preferentially with B cell linear epitopes of Ro60 kD autoantigen, enhancing recognition by anti-Ro60 kD autoantibodies. *Clinical and Experimental Immunology*, 134:143–150, 2003.

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

Steitz, J.A.; Berg, C.; Hendrick, J.P.; et al. A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. *Journal of Cell Biology*, 106:545–556, 1988.

Sun, D.; Lee, Y.S.; Malhotra, A.; Kim, H.K.; and Matecic, M. miR-99 family of microRNAs suppresses the expression of prostate-specific antigen and prostate cancer cell proliferation. *Cancer Research*, 71:1313–1324, 2011.

Tabara, H.; Sarkissian, M.; Kelly, W.G.; et al. The rde-1 Gene, RNA Interference, and transposon silencing in C. elegans. *Cell*, 99:123–132, 1999.

Taft, R.J.; Glazov, E.A.; Cloonan, N.; et al. Tiny RNAs associated with transcription start sites in animals. *Nature Genetics*, 41:572–578, 2009a.

Taft, R.J.; Glazov, E.A.; Lassmann, T.; et al. Small RNAs derived from snoRNAs. *RNA*, 15:1233–1240, 2009b.

Taft, R.J.; Kaplan, C.D.; Simons, C.; and Mattick, J.S. Evolution, biogenesis and function of promoter-associated RNAs. *Cell Cycle*, 8:2332–2338, 2009c.

Takaku, H; Minagawa, A; Takagi, M; and Nashimoto, M. A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. *Nucleic Acids Research*, 31:2272–2278, 2003.

Teunissen, S.W.; Kruithof, M.J.; Farris, A.D.; et al. Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. *Nucleic Acids Research*, 28:610–619, 2000.

Thompson, D.M. and Parker, R. The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae. *The Journal of Cell Biology*, 185:43–50, 2009a.

Thompson, D.M. and Parker, R. Stressing out over tRNA Cleavage. *Cell*, 138: 215–219, 2009b.

Thompson, D.M.; Lu, C.; Green, P.J.; and Parker, R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA*, 14:2095–2103, 2008.

Tomida, M.; Yamamoto, Y.; and Hozumi, M. Inhibition of the leukemagenicity of myeloid leukemic cells in mice and in vivo induction of normal differentiation of the cells by poly(I). poly(C). *Gann*, 71:457–463, 1980.

Tomita, K.; Ogawa, T.; Uozumi, T.; Watanabe, K.; and Masaki, H. A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. *Proceedings of the National Academy of Sciences of the United States of America*, 97:8278–8283, 2000.

Topfer, F.; Gordon, T.; and McCluskey, J. 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. *The Journal of Immunology*, 150:3091–3100, 1993.

Tsuji, T.; Sun, Y.; Kishimoto, K.; et al. Angiogenin is translocated to the nucleus of HeLa cells and is involved in ribosomal RNA transcription and cell proliferation. *Cancer Research*, 65:1352–1360, 2005.

Tyc, K. and Steitz, J.A. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO Journal*, 8:3113–3119, 1989.

Tycowski, K.T; Smith, C.M.; Shu, M.D.; and Steitz, J.A. A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, 379:464–466, 1996.

Tycowski, K.T.; Aab, A.; and Steitz, J.A. Guide RNAs with 5' caps and novel box c/D snoRNA-like domains for modification of snRNAs in metazoa. *Current Biology*, 23:1985–1995, 2004.

UCSC Genome Bioinformatics, and International Human Genome Project, . URL `http://genome.ucsc.edu/index.html`.

Vagin, V.V.; Signova, A.; Li, C.; et al. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313:320–324, 2006.

van Gelder, C.W.; Thijssen, J.P.; Klaassen, E.C.; et al. Common structural features of the Ro RNP associated hY1 and hY5 RNAs. *Nucleic Acids Research*, 22: 2498–2506, 1994.

Van Horn, D.J.; Eisenberg, D.; O'Brien, C.A.; and Wolin, S.L. Caenorhabditis elegans embryos contain only one major species of Ro RNP. *RNA*, 1:293–303, 1995.

Vaňáčová, S.; Wolf, J.; Martin, G.; et al. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biology*, 3:e189, 2005.

Vazquez, J.J. The fluorescent antibody method in the study of immunopathologic conditions. *Canadian Medical Association Journal*, 88:483–487, 1963.

Wang, K.C. and Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Molecular Cell*, 43:904–914, 2011.

Watson, J.D. and Crick, F.H. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171:964–967, 1953.

Wightman, B; Ha, I.; and Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75:855–862, 1993.

Wolin, S.J. and Steitz, J.A. 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, 1984.

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

Wurtmann, E.J. and Wolin, S.L. 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, 2010.

Xia, P.Z.; Geoghegan, W.D.; and Jordan, R.E. The particle (speckled-like thread) nuclear staining pattern: species and cellular distribution of Ro/SSA antigen. *Journal of Clinical and Laboratory Immunology*, 22:101–105, 1987.

Xiao, Q.; Sharp, T.V.; Jeffrey, I.W.; et al. The La antigen inhibits the activation of the interferon-inducible protein kinase PKR by sequestering and unwinding double-stranded RNA. *Nucleic Acids Research*, 22:2512–2518, 1994.

Xiao, Z.; Ko, H.L.; Goh, E.H.; Wang, B.; and Ren, E.C. hnRNP K suppresses apoptosis independent of p53 status by maintaining high levels of endogenous caspase inhibitors. *Carcinogenesis*, 34:1458–1467, 2013.

Xue, D.; Shi, H.; Smith, J.D.; et al. 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, 2003a.

Xue, L.-Y.; Chiu, S.-M.; and Oleinick, N. Staurosporine-induced death of MCF-7 human breast cancer cells: a distinction between caspase-3-dependent steps of apoptosis and the critical lethal lesions. *Experimental Cell Research*, 283:135–145, 2003b.

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

Yang, C.; Kim, M.S.; Chakravarty, D; Indig, F.E.; and Carrier, F. Nucleolin binds to the proliferating cell nuclear antigen and inhibits nucleotide excision repair. *Molecular and Cellular Pharmacology*, 1:130–137, 2009.

Yang, Z.X.; Lu, C.Y.; Yang, Y.L.; Dou, K.F.; and Tao, K.S. MicroRNA-125b expression in gastric adenocarcinoma and its effect on the proliferation of gastric cancer cells. *Molecular Medicine Reports*, 7:229–232, 2013.

Yoo, Christopher J and Wolin, Sandra L. The Yeast La Protein Is Required for the 3' Endonucleolytic Cleavage That Matures tRNA Precursors. *Cell*, 89:393–402, 1997.

Yoo, C.J. and Wolin, S.L. 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, 1994.

Zamore, P.D.; Tuschl, T.; Sharp, P.A.; and Bartel, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101:25–33, 2000.

Zhang, A.T.; Langley, A.R.; Christov, C.P.; et al. Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *Journal of Cell Science*, 124:2058–2069, 2011a.

Zhang, Y.; Liao, J.-M.; Zeng, S.X.; and Lu, H. p53 downregulates Down syndrome-associated DYRK1A through miR-1246. *EMBO reports*, 12:811–817, 2011b.

Zhou, H.; Chen, Y.Q.; Du, Y.P.; and Qu, L.-H. The Schizosaccharomyces pombe mgU6-47 gene is required for 2'-O-methylation of U6 snRNA at A41. *Nucleic Acids Research*, 30:894–902, 2002.

Zhu, J.; Hayakawa, A.; Kakegawa, T.; and Kaspar, R.L. Binding of the La autoantigen to the 5' untranslated region of a chimeric human translation elongation factor 1A reporter mRNA inhibits translation in vitro. *Biochimica et Biophysica Acta*, 152:19–29, 2001.

Zilberman, D.; Cao, X.; and Jacobsen, S.E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, 299:716–719, 2003.