# Identification and characterisation of novel small RNAs from repetitive elements in mammals

A thesis submitted for the degree of Doctor of Philosophy at the University of Glasgow

By

Louise Esther Docherty

Division of Cancer Sciences and Molecular Pathology
Faculty of Medicine
University of Glasgow
Glasgow
G11 5EB

July 2007

The research reported within this thesis is my own work except where otherwise stated, and has not been submitted for any other degree

Louise Esther Docherty

# Abstract

Transposable elements account for the almost half of the sequence encoded by mammalian genomes, which become silenced during early embryonic development. This thesis sought to explore the hypothesis of the involvement of the RNAi pathway in the silencing of transposable elements in mammals, predominantly through the identification of transposon-associated RNA of ~20-25nt using a gel blotting technique.

Initially cell lines of embryonic and tumour origin were analysed. This lead to the identification of several previously unreported transposon-associated RNA ranging from 70-90nt. However, it was not until a more detailed analysis of the embryonic cell lines, with the induction of differentiation in cell culture that several discrete RNA of ~20nt were detected for the mouse transposons L1 and B2. The differentiation of embryonic cell lines in culture also serendipitously lead to the detection of several short 5S rRNA of ~22-26nt, these were also later detected in several human cell lines of breast cancer origin and healthy breast tissue. Intriguingly the ~20-26nt repeat-associated identified were predominantly observed after two-days of differentiation in cell culture in several cell lines and often coincided with an ethidium bromide stainable band of ~19nt. The latter may indicate a large population of these RNA.

Further analysis of the B2 and 5S rRNA repeat-associated short RNA revealed both to have reduce accumulation in Dicer-null embryonic stem cells, implicating a possible association with a known component of the RNAi pathway. Dicer was also observed to process the longer ~50-80nt 5S rRNA to ~20-26nt *in vitro*. BLAST was also used to identify possible mRNA targets for the short B2 and 5S rRNA. One of these, the mRNA encoding sialic acid acetylesterase (SIAE) was consistently observed to be reduced with the accumulation of the short RNA using end-point RT-PCR, consistent with targeting through the RNAi or a similar pathway. However, no further links with the RNAi pathway were established, with no targeting detected for the short B2 or 5S rRNA using dual luciferase sensor assays.

The repeat-associated RNA identified in this thesis are among the first of their type and further work will be required to establish what relevance they have to the RNAi pathway and transposon regulation.

# Acknowledgments

I would like to acknowledge the following people for their help and support during my PhD. I am grateful to my supervisor Dr. Andrew Hamilton for all of his guidance and encouragement throughout this project in addition to proof reading my thesis. I would also like to thank my advisor Dr. Katherine West who provided me with advice on several occasions.

I would also like to thank everyone in Dr. Andrew Hamilton's group past and present for all of their advice. I would particularly like to thank Dr. Gurman Pall for taking the time to introduce me to the experimental techniques required to carry out this project.

I am grateful to the The University of Glasgow medical faculty for their studentship with particular thanks to Prof. Barry Gusterson.

I would like to dedicate this thesis to my family with a special tribute to my grandmother Ruby Pettit. I would like to thank my long suffering boyfriend David Blackie for all of his love, support and distraction during the course of my PhD, you really helped keep the stress levels to a minimum.

# Table of Contents

# List of Tables

# List of Figures

# Abbreviations

| | |
|---|---|
| °C | Degrees Celsius |
| μg | Microgram |
| μl | Microlitre |
| AMP | Ammonium persulphate |
| BLAST | Basic local alignment search tool |
| DEPC | Diethyl pyrocarbonate |
| $dH_2O$ | Distilled $H_2O$ |
| DMEM | Dulbecco's modified eagle medium |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| dNTP | 2` deoxy (nucleotide) triphosphate |
| EC | Embryonic carcinoma |
| *E.coli* | *Escherichia coli* |
| ES | Embryonic stem |
| EST | Expressed sequence tags |
| EtBr | Ethidium bromide |
| HeLa | Human cervical carcinoma cell line |
| HERV-L | Human endogenous retrovirus-L |
| IAP | Intracisternal A-particle |
| kb | Kilobases |
| LIF | Leukaemia inhibitory factor |
| LINE | Long interspaced elements |
| LTR | Long terminal repeat |
| M | Molar |
| MCF-7 | Human breast cancer cell line |
| miRNA | Micro ribonucleic acid |
| Mm | Millimolar |
| MPSS | Massively parallel signature sequencing |
| mRNA | Messenger ribonucleic acid |
| MuERV-L | Murine endogenous retrovirus-L |
| NCBI | National centre for biotechnology information |
| ng | Nanograms |

| | |
|---|---|
| PCR | Polymerase chain reaction |
| piRNA | Piwi-interacting RNA |
| pmol | picomole |
| PTGS | Post-transcriptional gene silencing |
| rasiRNA | Repeat-associated short interfering ribonucleic acid |
| RdRP | RNA-dependant RNA polymerase |
| RISC | RNA-induced silencing complex |
| RNA | Ribonucleic acid |
| RNase | Ribonuclease |
| rRNA | Ribosomal ribonucleic acid |
| RT-PCR | reverse transcriptase-PCR |
| sd | Standard deviation |
| SDS | Sodium dodecyl sulphate |
| SIAE | Sialic acid acetylesterase |
| siRNA | Short interfering ribonucleic acid |
| TBS | Tris buffered saline |
| TGS | Transcription gene silencing |
| TPM | Transcripts per million |
| Tris | 2-amino-2-(hydroxymethyl)-1,3-propanediol |
| TRNA | Transfer ribonucleic acid |
| UTR | Untranslated region |
| UV | Ultraviolet |
| w/v | Weight per volume |

# 2

# Introduction

# 2.1 Transposable elements

## *2.1.1 Background of transposable elements*

Transposons are regions of DNA capable of moving within the genome of a cell. In the 1950's Barbara McClintock was the first to describe transposable elements as controlling elements (McClintock, 1956). This work involved the observation that mutations in maize relating to the colour of corn kernels could be spontaneously reversed producing maize with variegated kernel colour. She attributed the instability of these mutants to the movement of control elements in and out of regulatory regions near these genes, thus affecting the genes action. However, it was not until transposable elements were discovered in bacteria many years later that the mechanisms of transposition began to be detailed at the molecular level (MacHattie and Shapiro, 1978, Shapiro, 1979).

Transposons have since been identified in all major branches of life although they vary greatly in the percentage of the host genome they occupy (Abrusan and Krambeck, 2006). Transposon insertion has frequently associated with mutations of the host genome by insertion into functional genes, the aberrant expression of neighbouring genes, inversions, deletions and non-homologous recombination. Therefore they have been previously been regarded as selfish parasitic DNA elements. However, transposon movement also provides a valuable source of heritable variation within the genome.

## *2.1.2 Classes of transposable elements*

There are many different prokaryotic and eukaryotic transposable elements but they can be divided into two main classes according to their mechanism of transposition.

The first class of transposable elements are mobilised through an RNA intermediate and are known as retrotransposons. This method of mobilisation has similarity to retroviral replication (Boeke et al., 1985) and is exclusive to eukaryotic genomes. This highly replicative mechanism of transposition has allowed these sequences to accumulate to high copy numbers resulting in the domination of eukaryotic genomes by retrotransposons. In mammals

retrotransposons account for ~40% of the genome (Lander et al., 2001, Waterston et al., 2002). The class I transposons can be further sub-divided into the long terminal repeat (LTR) and the non-terminal repeat transposons. The LTR transposons encode the enzymes reverse transcriptase and integrase, which are required for retrotransposition and are similar to those encoded by retroviruses. These sequences are flanked by direct-repeats providing promoter activity and make up 8% and 10% of the human and mouse genomes respectively (McCarthy and McDonald, 2004). The non-LTR transposons lack these flanking sequence and can be divided into two groups the long interspersed elements (LINEs) and the short interspersed elements (SINEs). LINE-1 encode reverse transcriptase which is required for transposition of the non-LTR transposons (Dewannieux et al., 2003, Dewannieux and Heidmann, 2005), whereas the non-autonomous SINEs do not encode any retroviral related enzymes. The individual properties of the some transposons from these groups will be discussed further within the results chapters.

The class II transposable elements do not require an RNA intermediate for their mobilisation and are also known as DNA transposons. The DNA transposons are predominantly prokaryotic (Siguier et al., 2006) but also represent a small proportion of the eukaryotic transposon population (Lander et al., 2001, Waterston et al., 2002). There are two proposed mechanisms of DNA transposon mobilisation, conservative and replicative which both require transposase activity. Conservatively transposed elements are excised from their original location and inserted into a new location, a process commonly referred to as "cut and paste" transposition (Kleckner, 1977). This is a non-replicative method of transposition therefore these transposons do not accumulate in the genome of their host. Alternatively replicative transposition, commonly referred to as "copy and paste" transposition does result in the accumulation of these sequences in the host genome (Shapiro, 1979). During replicative transposition the transposon is retained at its original location where it is replicated prior to insertion into a new target site. DNA transposons can be sub-divided according to the mechanism by which they are mobilised. Within these groups there are many families of DNA transposons across many species. Some of these encode the viral related enzymes of transposase and resolvase required for DNA transposition. Non-autonomous DNA transposons can be mobilised by the recruitment of these enzymes provided by autonomous elements.

### *2.1.3 Models of transposon expansion*

There are often many transposable element families coexisting within a single genome. However, the transposition dynamics of these varies greatly between organisms with regard to the copy number and activity of individual transposons. For example mammals contain a few, highly abundant transposon families that are largely inactive, whilst other organisms such as *D.melanogaster* contain fewer transposons with high activity (Eickbush and Furano, 2002). These differences have lead to the assumption that transposition rates maybe dependant on the number of active elements within the family (Ohta, 1986). Two opposing models for the transposition of replicative elements have been described (Clough et al., 1996).

The first of these is the "master gene" model. This describes transposon families with few active elements, which dominate the propagation of a transposon family. The active transposons of this model are called master copies. Transposons expanding by the master gene model can be recognised as having multiple sub-families of which the majority are inactive. The generation of these sub-families is due to the sequential dominance of individual master copies during different periods of the transposons expansion. Therefore large sub-families indicate periods of rapid individual master copy expansion. When a transposable element family mobilised by this model are viewed as a phylogenetic tree a single lineage is observed. The majority of mammalian retrotransposons such as the human Alu or LINE-1 families replicate in this way with few active elements. Retrotransposons have few copies capable of transposition due to the accumulation of errors, truncations and deletions associated with their transposition and maintenance in the host genome.

The second is the "random template" model. This model predicts that every copy of a transposable element family has an equal probability of transposition. Therefore many active elements are capable of simultaneously transposing within the host genome. Transposon families mobilised through this model can be recognised by the polymorphic insertion of transposons within a population. Therefore when viewed as a phylogenetic tree numerous lineages will be observed. However, it would be difficult to imagine a transposon family in which

every copy was capable of transposition due the accumulation of mutations over time.

However, these models are not mutually exclusive and few transposon families strictly adhere to either. Instead the transposition of many transposon families is proposed to follow various intermediate models that fall between the extremes of the single master gene and random template models (Kass et al., 2000, Shen et al., 1991).

### *2.1.4 Transposable element expression*

The transcription of transposable elements is essential for both transposition and retrotransposition. The majority of transposable elements contain their own promoters to facilitate this. For example within the retrotransposons, SINEs have maintained the RNA polymerase III activity of their original source sequences (tRNA or 7SL) (Ferrigno et al., 2001, Vassetzky et al., 2003) whereas LINE and LTR transposons contain RNA polymerase II promoters (Maksakova and Mager, 2005, Svoboda et al., 2004). However, in order to protect the host genome from the detrimental effects of excessive transposition, many organisms (including mammals) have mechanisms to transcriptionally repress the promoters of transposable elements (discussed 2.1.5).

Upon fertilisation, the DNA of the oocyte and sperm, which are highly methylated undergo passive and active global de-methylation of their genomes respectively (with the exception of imprinted regions) (Allegrucci et al., 2005). This results in the de-repression of the parental genomes allowing the opportunistic expression of transposable elements on activation of the zygotic genome (see 2.1.4). Studies have shown that after this initial high level of transcriptional activity transposons become transcriptionally repressed to very low basal levels of expression (Evsikov et al., 2004, Kigami et al., 2003, Peaston et al., 2004). The global repression of transposable elements is achieved in the early implantation embryo, however the rate of its progression varies between transposable elements. For example with the murine MuERV-L transposon, its highest level of expression is observed at the two-cell stage, with very low levels observed by the blastocyst stage. However, another murine LTR transposon IAP

is expressed at low levels in the two-cell stage being up-regulated until the blastocyst where it is at its highest expression (Svoboda et al., 2004).

In mammals, the transcriptional activity of transposable elements is largely restricted to germ cells and pre-implantation embryos, during periods of global de-methylation of the genome. In somatic cell, transposons are widely transcriptionally repressed, although cellular stresses (Hagan et al., 2003, Li et al., 1999, Liu et al., 1995) or disease states such as cancer (Schulz, 2006) have been observed to relieve this repression. However, read-though transcription from neighbouring gene promoters can also provide an alternative source of low-level transposon transcription in somatic cells.

Using cDNA libraries prepared from mouse oocytes and different mouse pre-implantation embryo stages, it was demonstrated that transposon expression is highest during the two-cell stage (Ko et al., 2000). This peak in transposon expression is likely due to the activation of the zygotic genome, which also occurs at the two-cell stage. Therefore transposable element sequences are among the first sequences to be transcribed by the embryonic genome. The transposon-derived sequences represent a substantial proportion of the two-cell transcriptome (8-17%) (Evsikov et al., 2004, Ko et al., 2000), with ~80% of the two-cell transposon-derived sequences identified as retrotransposons (Evsikov et al., 2004). The translation of these sequences was also established by assessing the reverse transcriptase (RT) activity at different developmental stages. This indicated there to be a ten fold increase at the two-cell stage compared with RT levels in oocytes and other pre-implantation stages, and a ~1800 greater level compared to somatic cell RT levels (Beraldi et al., 2006). These results indicate that the two-cell stage may provide a developmental window for retrotransposition. Transposon activity in the germ line and pre-implantation embryo maybe tolerated because it provides a benefit to the host genome. This could be through providing a source of heritable variations. In addition the early embryonic expression of some transposable elements appear to have acquired functions making them indispensable during development (Beraldi et al., 2006).

In mammalian genomes there are a few currently active transposable elements that transpose at low frequency. Of these, a single retrotransposon family LINE-1 (L1) is responsible for the majority of new transposition events by mobilising

itself and some SINE elements (Dewannieux et al., 2003, Dewannieux and Heidmann, 2005, Ostertag et al., 2003).

### *2.1.5 Transposon repression*

In mammals, during gametogenesis and shortly after embryonic implantation a wave of *de novo* cytosine methylation occurs resulting in the re-methylation of genomic DNA (Yoder et al., 1997). This process is required for the re-programming of germ-line and embryonic cells and results in the extensive methylation of transposable elements. The DNA methyltransferases (Dnmt) 3a and 3b are responsible for this *de novo* methylation and appear to be essential for mammalian development (Lei et al., 1996, Okano et al., 1999). However, the maintenance of this methylation in subsequent cell divisions requires Dnmt1 (Okano et al., 1999).

The DNA methylation of transposon promoter regions acts to repress their transcription (Walsh et al., 1998, Woodcock et al., 1997). This repression is further reinforced by the acquisition of histone modifications to these regions, resulting in repressive chromatin remodelling. These histone modifications primarily include methylation of lysine 9 (H3mK9) of histone 3 (Martens et al., 2005), which directly recruits the heterochromatin protein complex Swi6/HP1. The assembly of transposable elements into repressed chromatin, in addition to repressing their transcription, also assists in providing genomic stability by preventing recombination between transposons. Thus, through DNA methylation and H3mK9 methylation transposable elements become unable to move around the genome (silenced). The inhibition of either of these modifications results in the de-repression of transposable elements in the genome (Li et al., 2000, Ma et al., 2001, Woodcock et al., 1997). However, despite knowledge of the epigenetic modifications required for the repression of transposable elements in mammals, little is known about what controls the specific targeting of these sequences.

The epigenetic modifications of cytosine methylation and H3mK9 are also extensively associated with transposon repression in many other organisms. In many species including plants both of these modifications are associated with repressed transposable elements (Gendrel et al., 2002, Lippman et al., 2003). In

some other species such as fission yeast and *D.melanogaster,* which largely lack DNA methylation, transposon repression is associated with H3mK9 modification alone (Pal-Bhadra et al., 2004, Volpe et al., 2002). It was through the investigation of transposon repression in other species that the RNA interference (RNAi) pathway (see 2.2) was first implicated to be involved in this process. The RNAi pathway has been proposed to achieve this transcriptionally, by the specific targeting of epigenetic modifications to transposon sequences in the genome and post-transcriptionally through the repression and turnover of transposon transcripts. This has been shown to be a conserved mechanism of transposon repression in several organisms and will be discussed in more detail later in this chapter.

## 2.2 RNA interference

### *2.2.1 Background to RNA interference*

RNA interference (RNAi) is the term used to describe the process of gene regulation through the base-pairing of short ~22nt RNA to complementary nucleic acid sequences. The RNAi pathway, which mediates this process is highly evolutionarily conserved, with no eukaryotic organism having yet been described that lack the mechanism (Rana, 2007).

The sequence-specific gene silencing of RNAi was first described in plants (Napoli et al., 1990, Smith et al., 1990, van der Krol et al., 1990). In petunia flowers this phenomena was termed co-suppression and was observed following attempts to over-express the chalcone synthase enzyme (Napoli et al., 1990, van der Krol et al., 1990), which is involved in biosynthesis of the purple pigmentation of petals. Over-expression of the enzyme was expected to result in the darkening of the flower petals. However, in some transgenic lines over-expression of the transgene resulted in the partial or complete loss of colour from sectors of the petal. This effect was attributed to the unanticipated under-expression of both the endogenous and transgene chalcone synthase caused by post-transcriptional targeting of the mRNA. Subsequently similar observations of post-transcriptional sequence-specific gene silencing were also made in fungi (Cogoni et al., 1994) and *C.elegans* (Fire et al., 1991), where they were termed quelling and RNAi respectively.

The link between these processes was not made until the triggers and mechanism underlying co-suppression (Napoli et al., 1990), quelling (Cogoni et al., 1994) and RNAi (Fire et al., 1991) were separately investigated to uncover an evolutionary conserved pathway. The observation that short ~22nt antisense RNA could be regulating transcripts with complementarity was first proposed in *C.elegans*. This hypothesis was based on the negative regulation of the non-coding RNA lin-4 on lin-14 mRNA, with which it shared complementary sequence in the 3´untranslated region (UTR) (Lee et al., 1993). It was not until later investigations in *C.elegans* that a simple antisense mechanism of RNAi was ruled out, with the observation that long antisense RNA triggers were substantially less efficient than dsRNA in inducing RNAi (Fire et al., 1998). Investigations in plants

undergoing RNAi then identified the natural trigger of RNAi to be dsRNA of ~25nt (Hamilton and Baulcombe, 1999), which after further investigation uncovered two short RNA species of ~22nt and ~26nt (Hamilton et al., 2002). DsRNA of ~22nt have subsequently been observed as the predominant trigger in all organisms undergoing RNAi (Elbashir et al., 2001, Elbashir et al., 2001). However, there are other examples of larger short RNA like those found in plants which function through the RNAi pathway (Aravin et al., 2006, Girard et al., 2006). In addition to the dsRNA trigger, several proteins are required for the functioning of the RNAi pathway. Many of these proteins were identified through mutational analysis and are conserved between organisms capable of RNAi. These proteins include the RNase III enzyme Dicer that is responsible for the processing of long dsRNA into the short dsRNA triggers (Bernstein et al., 2001) and the Argonaute family of proteins that form the key component of the targeting complex (Fagard et al., 2000, Hammond et al., 2001).

The conservation of the RNAi pathway across the eukaryotic kingdoms of life has lead to speculation as to the cause and significance of this striking evolutionary selection. In many organisms a major source of dsRNA is RNA viruses therefore this lead to speculation of the RNAi pathway acting as an antiviral defence, protecting the genome from invading viruses (Smith, 1999) and those, which become integrated into the genome as transposable elements (Flavell, 1994). This idea is supported with many viruses encoding suppressors of RNAi (Mette et al., 2001, Ye et al., 2003) and growing evidence of the pathways role in the suppression of transposable elements (Lippman et al., 2003, Svoboda et al., 2004, Tabara et al., 1999, Yang and Kazazian, 2006). However, irrespective of the original cause of selection, RNAi is now associated with several important processes in the cell, included the formation of heterochromatin at centromeric regions of the genome (Volpe et al., 2002) and the control of some developmental processes (Bernstein et al., 2003).

### *2.2.2 Classes of short RNA*

The short dsRNA that guide RNAi were initially divided into two classes based on their precursor dsRNA and their proposed action in post-transcriptional gene silencing. These classes were the small interfering RNA (siRNA) and micro RNA

(miRNA). More recently Piwi-interacting RNA (piRNA) that range from 26-31nt in length have emerged as a third class of short RNA in mammals and insects.

The siRNA were the original short RNA trigger associated with RNAi (Hamilton and Baulcombe, 1999). SiRNA are generated by the cleavage of long dsRNA into ~22nt dsRNA with a 5´ phosphate and a 2nt 3´ overhang (Zamore et al., 2000). This class of short RNA are far less abundant than the miRNA and have yet to be identified as endogenous to mammals, where long dsRNA induce an interferon response. Plants differ from other organisms by having two species of siRNA of ~22nt and ~26nt (Hamilton et al., 2002). These siRNA species are generated from the processing of long dsRNA by different members of the Dicer family of enzymes and are proposed to act in the RNAi pathway to mediate different types of gene suppression (Xie et al., 2004). However, a partial overlap in the generation and function of these RNA has been observed (Khvorova et al., 2003).

The miRNA are the most abundant endogenous class of short RNA in both plants and animals and have been associated with functions in development (Hornstein and Shomron, 2006, Krutzfeldt et al., 2006) and cellular differentiation (Bentwich, 2005). The first member of the class that have become known as miRNA was a 22nt RNA derived from the non-coding RNA *lin*-4, which suppresses *lin*-14 mRNA translation (Lee et al., 1993). However, the association of miRNA with the RNAi pathway did not come until after the discovery of siRNA (Hamilton and Baulcombe, 1999). MiRNAs are encoded mostly within introns in mammals and are generally transcribed by RNA polymerase II as part of long primary transcripts (pri-RNA) (Lee et al., 2004), however up to 20% of human miRNA are reported to be transcribed from the read-through transcription of repeat elements by RNA polymerase III (Borchert et al., 2006). These transcripts are often long and can encode several miRNA (Mineno et al., 2006), which fold to form the characteristic dsRNA hairpin structures. The dsRNA hairpin structures then undergo sequential cleavage to generate first a ~70nt precursor and then a ~22nt mature miRNA (Figure 1B). Like siRNA both of these dsRNA cleavage products have a 5´ phosphates and a 2nt 3´overhang characteristic of products of RNase III like enzymes. For a review of miRNA see (Bartel, 2004).

The piRNA are a class of single stranded ~26-30nt RNA that were originally found in the testis of mammals (Aravin et al., 2006, Girard et al., 2006). These were

originally observed through their association with PIWI proteins, which are members of the argonaute family. PiRNA are highly abundant in mammalian testis, thus can be detected by SYBR gold staining as a discrete band of ~30nt (Aravin et al., 2006, Girard et al., 2006). The generation of piRNA is much less well understood than that for siRNA and miRNA. However, they have been observed to cluster in the genome, which may suggest that they are processed from long primary transcripts (Aravin et al., 2006). Of the piRNA identified so far in mammals a number associating with MILI (a PIWI sub-family member in mouse) are derived from repeat-sequence including transposable elements (Aravin et al., 2007). This along with their size makes the mammalian piRNA, similar to the repeat-associated small interfering RNA (rasiRNA) identified in *D.melanogaster* (Aravin et al., 2001). The rasiRNA also, despite being defined as a sub-class of siRNA because they are processed from long dsRNA, function through PIWI in *D.melanogaster* (Vagin et al., 2006).

### *2.2.3 Processing of dsRNAs*

There are many potential sources of dsRNA within the cell, which can be derived from bi-directional transcription, repeat elements or hairpins. In addition to these, external sources of dsRNA can be introduced to the cell from viruses or short hairpin RNA (shRNA) expression plasmids.

The cleavage of long dsRNA into mature siRNA that guide RNAi was first demonstrated in *D.melanogaster* embryos and embryonic extracts (Yang et al., 2000, Zamore et al., 2000). These siRNA when analysed were observed to have the distinctive properties of RNase III–like cleavage, with a 2nt 3´overhang and a 5´phosphate (Elbashir et al., 2001). Several ribonuclease members of the RNase III family were assessed for their ability to generate RNA of ~22nt from long dsRNA because of their affinity for dsRNA substrates. These studies revealed an RNase III like enzyme, which has become known as Dicer to be responsible for catalysing the cleavage of long dsRNA into RNA of ~22nt (Figure 1A) (Bernstein et al., 2001). Subsequently Dicer was also associated with the cleavage of pre-miRNA into mature miRNA through mutational studies in *C.elegan* (Ketting et al., 2001).

The Dicer enzyme contains a dsRNA binding domain, PAZ domain, helicase domain and two RNase III domains. The first model for the cleavage of dsRNA by Dicer was proposed by Hannon indicating Dicer to function as dimer to mediate the cleavage of dsRNA into ~22nt RNA (Hannon, 2002). This mechanism was implicated because the active catalytic RNase III domains of a Dicer dimer would be positioned ~22nt apart. However, this model was subsequently dismissed in favour of a monomeric Dicer cleavage model. This proposes that Dicer forms a single processing centre with which both of the RNase III domains associate, each cleaving one of the two RNA strands to generate dsRNA of ~22nt (Zhang et al., 2004). The Dicer enzyme is highly conserved and can be found in worms, flies, fungi, plants and mammals (Bernstein et al., 2001). Several organisms encode more than one Dicer gene, which can be responsible for processing dsRNA from different sources or for different functions. For example in plants there are four Dicer genes, with Dicer1 processing miRNA, Dicer2 processing viral siRNA and Dicer3 processing endogenous siRNA (Xie et al., 2004). In mammals however there is a single Dicer gene, which is responsible for the processing of both siRNA and miRNA. Unsurprisingly this Dicer is critical in mammalian development, with Dicer-null mammals failing to progress beyond early development (Bernstein et al., 2003).

Prior to the action of Dicer, miRNA require processing of the pri-miRNA into pre-miRNA (Han et al., 2004) to allow export into the cytoplasm where Dicer is located. Analysis of the pre-miRNA identified a 2nt 3´overhang, a feature of RNase III processing. Because of the nuclear localisation of the pri-miRNA, the nuclear localised RNase III enzyme Drosha was investigated. Using Drosha purified from mammalian cell lysate it was shown to process pri-miRNA into RNA of pre-miRNA of ~70nt (Lee et al., 2003). Drosha like Dicer is highly conserved in several species and associates with another protein DGCR8 to bind and efficiently cleave pri-miRNA (Han et al., 2006). In other words pri-miRNA are sequentially processed by two RNase III enzymes Drosha and Dicer to generate mature miRNA (Figure 1B).

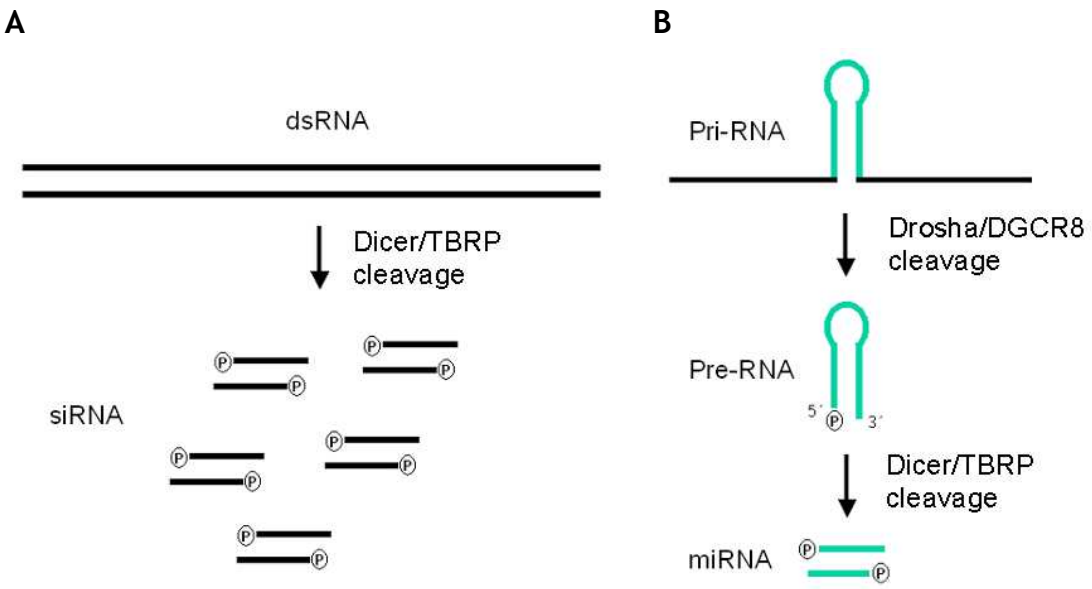


**Figure 1- dsRNA processing.** This figure illustrates the processing of long dsRNA into siRNA by the Dicer complex (A) and the sequential processing of pri-RNA by Drosha and Dicer complexes into miRNA (B). The mammalian Drosha and Dicer complex associating factors of DGCR8 and TRBP respectively have also been included, however these are not the only cofactors to associated with the Drosha and Dicer complexes. Both siRNA and miRNA processing by the Dicer complex have a 5´ phosphate and a 3´ 2nt overhang.

### *2.2.4 Post-transcriptional gene silencing*

Post-transcriptional gene silencing (PTGS) originally described the process of mRNA targeting by cleavage and/or translational suppression guided by siRNA or miRNA in plants and was analogous to RNAi in *C.elegans* (Hamilton and Baulcombe, 1999). However, since then, RNAi has also been associated with transcriptional gene silencing and heterochromatin formation (Martienssen, 2003, Onodera et al., 2005, Pal-Bhadra et al., 2004, Volpe et al., 2002). Therefore for the purpose of this thesis I will use the term PTGS to refer globally to RNA directed post-transcriptional repression (Figure 2).

To mediate PTGS, short RNA associates with an effecter complex of proteins conferring a sequence specific endonuclease activity, termed the RNA-induced silencing complex (RISC) (Hammond et al., 2000). All of the proteins comprising RISC have yet to be clearly established and may vary significantly between effecter complexes and species. However, the catalytic centre of RISC was the first component to be identified in RISC through a combination of biochemical and mutational studies (Hammond et al., 2001, Tabara et al., 1999). This protein

is a member of the argonaute family of proteins. Argonaute proteins are consistently found to be the key component of miRNA, siRNA and more recently piRNA directed RISC in various species (Aravin et al., 2006, Fagard et al., 2000, Girard et al., 2006). The phylogenetic analysis of several argonautes divided them into two main clades, the Ago-like and Piwi-like argonautes (Tolia and Joshua-Tor, 2007). The number of argonaute proteins varies greatly between different species. Mammals encode eight: four Ago-like and four Piwi-like. Various members of the argonaute protein family have been shown to be involved in the formation of different RISC complexes, typically with the Ago-like argonautes forming siRNA-RISC and miRNA-RISC (Okamura et al., 2004), and the Piwi-like argonautes forming piRNA-RISC (Aravin et al., 2006, Girard et al., 2006). Of the four mammalian Ago-like argonautes assessed, the Ago 2 RISC alone has been demonstrated to be capable of cleaving mRNA targets and with the short RNA can form the minimal active siRNA-RISC (Rivas et al., 2005). However, both siRNA and miRNA associated with RISC containing other Ago-like argonautes to exert translational suppression in addition to any potentially unknown function.

To mediate PTGS through RISC in mammals one strand of the mature siRNA or miRNA is unwound and loaded into RISC by the Dicer-TRBP complex (Chendrimada et al., 2005). The loaded strand is known as the guide strand and is selected based on the internal stability of the mature short RNA ends (Khvorova et al., 2003). The passenger strand of the mature short RNA is then often degraded (Matranga et al., 2005). The guide strand directs RISC to the target sequence through the binding of complementary sequence. Positions 2-8nt of the 5´guide miRNA "seed region" have been shown to be critical in the recognition of the target site (Doench and Sharp, 2004). Mismatches in this area often result in the loss of target binding and repression. The target mRNA is then either cleaved or translationally repressed dependant on the extent of complementarity and the argonaute component of RISC. Cleavage of the target RNA requires significant complementary to the short RNA and is most often associated with siRNA. The target mRNA is cleaved in the centre of the complementary region, between the nucleotides base-paired to nucleotides 10 and 11 of the guide siRNA strand (Elbashir et al., 2001). The guide strand is not cleaved in this process and after the release of the cleavage products does not dissociate from RISC. This allows multiple rounds of cleavage for each short RNA-

directed RISC formed (Haley and Zamore, 2004). When there is mismatch between the guide strand and the target outside the seed region as is often the case with miRNA, cleavage does not occur. The association of RISC, results in the translational repression of the target mRNA.

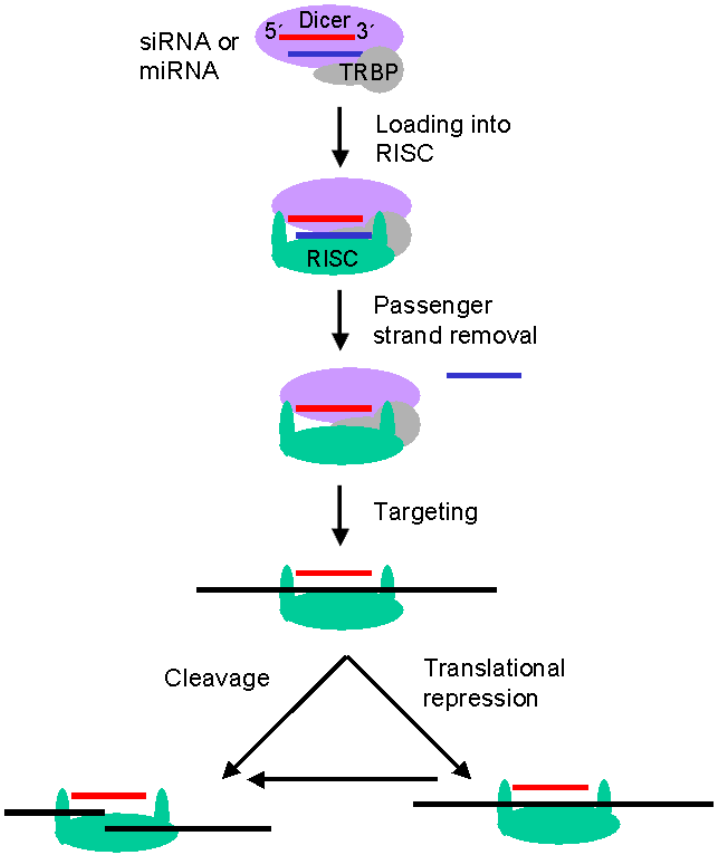


**Figure 2- Post-transcriptional gene silencing.** Both siRNA and miRNA are loaded into RISC by the Dicer-TRBP complex, the passenger strand then dissociates and becomes degraded. The guide strand then directs RISC to a region of complementarity in the target mRNA (black). Dependant on the extent of complementarity the target will either be cleaved or remain bound to the mRNA resulting in translational repression although some translationally repressed mRNA may later become cleaved.

### *2.2.5 Transcriptional gene silencing*

Transcriptional gene silencing (TGS) is a siRNA directed mechanism of gene repression through the establishment of DNA methylation and/or heterochromatin formation. TGS has been observed naturally in plants, animals and fission yeast (Morris et al., 2004). However, the establishment of TGS and the involvement of DNA methylation in this process vary between species. Mutational studies have identified mutations that affect both TGS and PTGS, suggesting some overlap between the pathways (Volpe et al., 2002). Genes that

are essential to both pathways include argonaute family members and Dicer (Volpe et al., 2002).

The phenomenon of RNA-dependant DNA methylation (RdDm) was originally reported in viroid infected plants where regions corresponding to the viroid RNA were specifically methylated (Wassenegger et al., 1994). Subsequently short RNA were observed to direct RdDm when siRNA acting in PTGS, also established DNA methylation of the homologous DNA region (Jones et al., 1999). When either endogenous or synthetic siRNA containing sequences specific to the promoter region of a gene induce RdDm, TGS of the gene is established (Mette et al., 2000). In plants the longer siRNA species of ~26nt were found to be exclusively involved (Hamilton et al., 2002). Despite the ~26nt siRNA being specific to plants, RdDm has been observed naturally in other organisms suggesting the targeting short RNA involved in this process may differ between species. In mammals the endogenous establishment of RdDm has yet to be confirmed, although synthetic siRNA have been reported to direct TGS (Kawasaki and Taira, 2004, Morris et al., 2004).

The establishment of RdDm in addition to the siRNA also requires several proteins, many of which are conserved between the mechanisms of various species. Generally these include a Dicer for the production of the siRNA, an argonaute protein for targeting, DNA methyltransferases and chromatin-modifying factors for the establishment of *de novo* methylation. The Dicer and argonautes required for TGS can vary from those required for PTGS. For example in mammals the argonaute Ago1 is required for the establishment of RdDm and TGS whereas Ago2 is required for PTGS (Kim et al., 2006).

Heterochromatin formation can also establish TGS either alone or in combination with RdDm. The majority of heterochromatin is usually found at the centromeres and telomeres (Lippman et al., 2004) of chromosomes and is established by modification of histone proteins. The establishment of heterochromatin by histone modification require the action of several groups of proteins including the histone deacetylases, methyltransferases and chromo-domain proteins. In addition, mutations in proteins of the RNAi pathway have been reported to relieve heterochromatin formation (Kanellopoulou et al., 2005, Pal-Bhadra et

al., 2004, Volpe et al., 2002). This observation may suggest the involvement of the RNAi pathway in the formation of heterochromatin.

RNA-dependent heterochromatin formation (Rdhf) is observed in various organisms and spreads beyond the region of RNA-DNA homology for several kilobases (Volpe et al., 2002). One proposed mechanism of general Rdhf is by the *de novo* DNA methylation of a genomic region through the recognition of the RNA-DNA hybrid by DNA methyltransferases. This methylated DNA is then proposed to recruit chromodomain proteins that establish lysine-9 methylation and heterochromatin formation (Figure 3). However, there is limited evidence as to the exact interactions involved with RNA-RNA interaction models (Grewal and Elgin, 2007) also being proposed in this process and many unanswered questions.

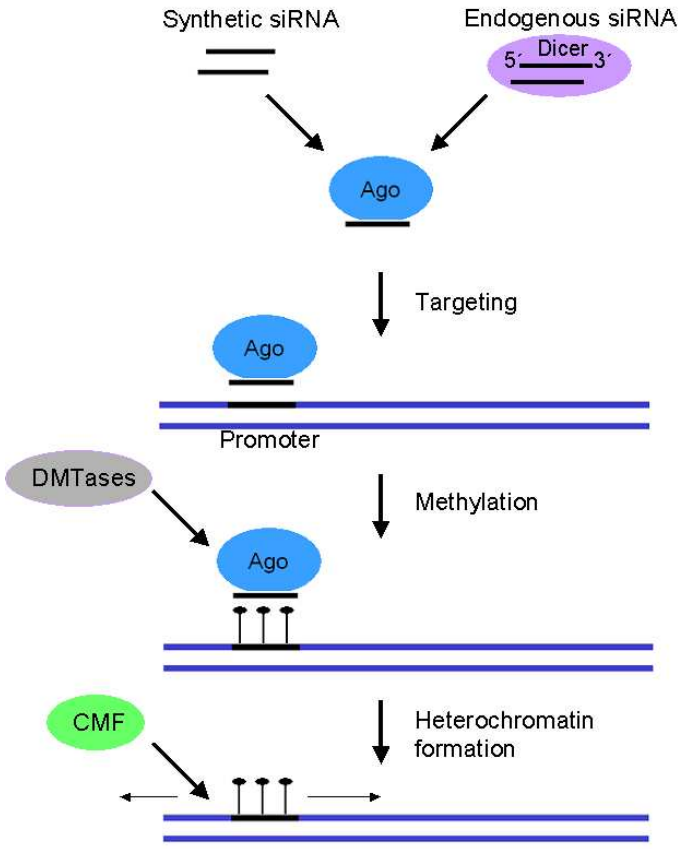


**Figure 3- Transcriptional gene silencing via DNA methylation and heterochromatin formation.** Synthetic and endogenous siRNA can both transcriptionally repress genes when there sequence is specific to the promoter region of a gene. The siRNA associate with an Ago protein, different to the PTGS pathway Ago (Ago4 in plants and Ago1/2 in mammals). The siRNA then targets the homologous gene promoter. This RNA-DNA hybrid is recognised recruiting DNA methyltransferases (DMTases) that establish *de novo* cytosine methylation. The DNA methylation can then recruit chromatin-modifying factors methylate lysine-9 of histone 3 and establish heterochromatin. The heterochromatin can then spread bi-directionally from the original site of siRNA-induced methylation.

However, some organisms such as *S.pombe* do not establish *de novo* methylation of the target DNA and therefore have a *de novo* methylation independent method of establishing Rdhf. The purification of chromodomain complexes associated with the centromeric regions in the fission yeast *S.pombe* lead to the establishment of an alternative model for siRNA induced Rdhf specific to *S.pombe*. This model proposes the siRNA direction of a RNA-induced TGS complex (RITS) to the complementary DNA region where it binds, resulting in the recruitment of chromatin modifying proteins that impose heterochromatin on the targeted region. The RITS complex contains siRNA specific to the region being silenced, an Ago family protein different to that found in RISC (Ago), a chomodomain protein (Chp1) and a novel protein (Tas3), which is not yet known to have any mammalian homolog (Verdel et al., 2004). The guide siRNA was shown to be required for the binding of the associated proteins to the DNA. In recent studies this has been proposed to be due to the siRNA targeting the chromosome by base-pairing with pre-mRNA as RNA polymerase II transcribes (Buhler et al., 2006). However, the RITS complex has only been identified in *S.pombe* and there is no evidence of a mammalian equivalent.

### *2.2.6 Adenosine deaminases acting on RNA and the RNAi pathway*

The presence of dsRNA within the cell often is an indication of viral infection. Therefore the cell has developed strategies to repress these transcripts. In addition to RNAi the hyper-editing of dsRNA >30nt by the adenosine deaminases acting on RNA (ADAR) family of enzymes is known to repress dsRNA in the cells. This process is proposed to be active in all metazoa and catalyses the conversion of adenosine residues to inosine through deamination (see Figure 4). The deamination of adenosine residues to inosine alters the coding capacity of the RNA because inosine is recognised as guanosine. This modification also results in the destabilisation of the dsRNA structure by the introduction of G:U wobble base pairing.

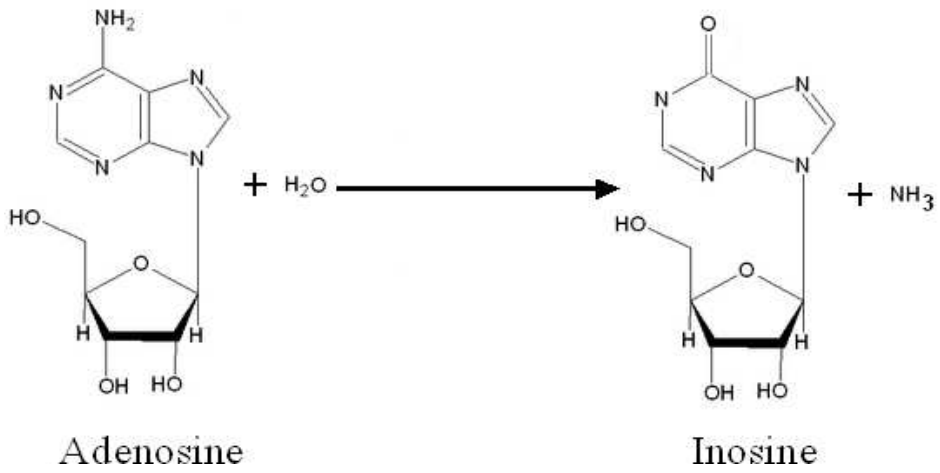


**Figure 4- The deamination of adenosine to inosine.** The hydrolytic deamination of adenosine is catalysed by ADAR proteins and results in the loss of and amine group from the nucleotide forming inosine.

The ADAR enzymes act in the nucleus and localise to sites of transcriptional activity. The hyper-editing of dsRNA requires at least ~25-30nt of dsRNA with optimal editing for dsRNA transcripts of >100nt. ADARs are essential for normal life in metazoa, however the effect of null-mutations vary greatly between species. For example *C.elegans* and *D.melanogaster* ADAR null-mutants are viable with behavioural defects (Palladino et al., 2000, Tonkin et al., 2002), whereas ADAR editing is essential in mammalian development with null-mutations being lethal (Higuchi et al., 2000). In humans the majority of the hyper-editing targets have been identified as transposable element RNAs from a single SINE family Alu (Levanon et al., 2004). This was achieved by comparing cDNA sequences back to the genome and recording observed guanosine/adenosine anomalies.

RNAi and hyper-editing by ADARs are often regarded as mutually antagonistic processing pathways that compete for dsRNA substrates (Scadden and Smith, 2001) as depicted in Figure 5. This theory is further supported by evidence that mutations of the ADAR pathway can be rescued by mutations of the RNAi pathway in *C.elegan* (Tonkin and Bass, 2003). The fate of a dsRNA is therefore dependant on which of these pathways it interacts with first. When the RNAi pathway is first to encounter dsRNA, the transcript is processed into dsRNA of ~22nt which is a very poor substrate for hyper-editing by ADARs. Conversely increasing levels of dsRNA ADAR modification progressively impair the ability of Dicer to cleave the RNA (Scadden and Smith, 2001). Hyper-edited dsRNA are also often retained in the nucleus of the cell through associations with specific

protein complexes (DeCerbo and Carmichael, 2005) or are targeted for cleavage (Scadden and Smith, 1997). This further represses the ability of Dicer to process these transcripts, which in turn suppresses the induction of RNAi.

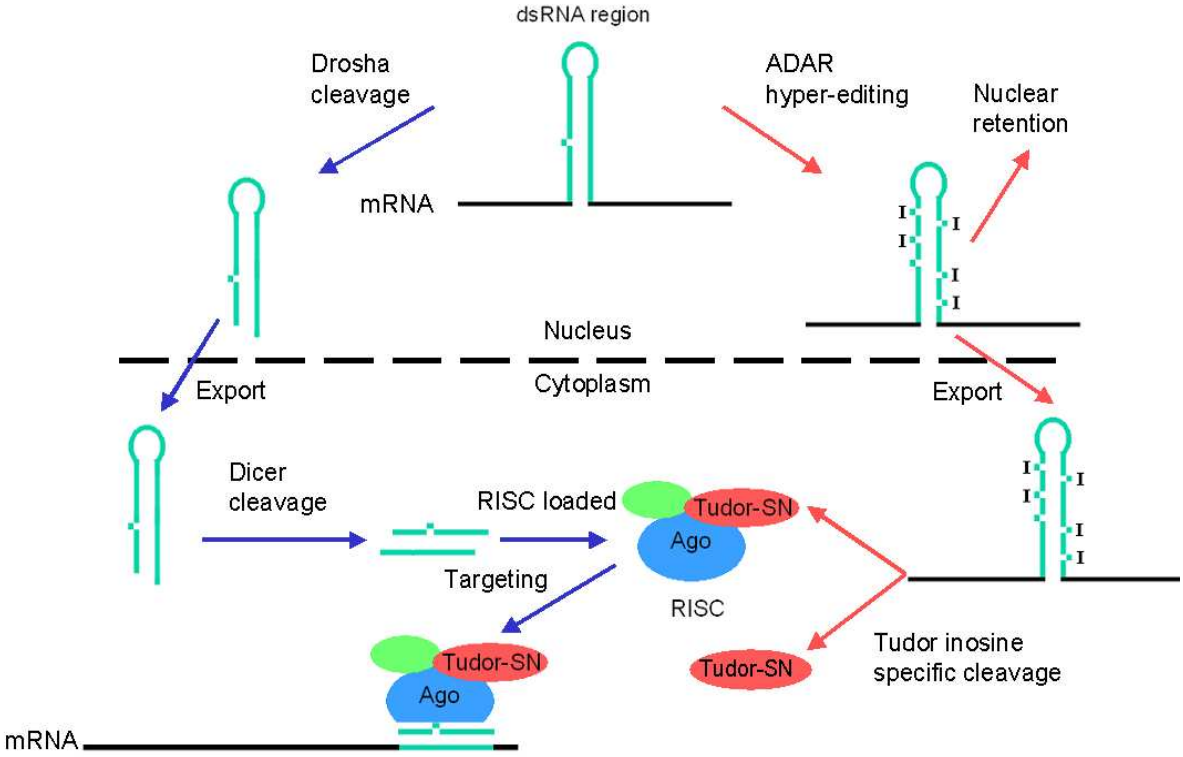


**Figure 5- The antagonistic interaction of the RNAi and ADAR pathways.** This image uses the example of a pri-RNA to depict the antagonistic interaction of the RNAi (blue arrows) and ADAR pathway (pink arrows). When dsRNA of >30nt is first encountered by the ADAR pathway hyper-editing of the double stranded region converting adenosine to inosine residues by deamination occurs. The hyper-editing of dsRNA reduces the secondary structure by the introduction of G:U wobble base pairs which makes the RNA a poor substrate processing in the RNAi pathway. The inosine modified RNA transcripts are then either retained in the nucleus or exported to the cytoplasm for specific degradation by Tudor-SN. However, it is unknown if RISC associated and free Tudor-SN are both able to cleave ADAR modified RNA. Alternatively if the RNAi pathway acts on the pri-RNA first the pre-miRNA is exported from the nucleus and progresses through the RNAi pathway. Cleavage of the dsRNA to <30nt generates poor substrates for ADAR modification and is therefore antagonistic to the ADAR pathway.

Despite the antagonistic properties of the RNAi and the ADAR hyper-editing pathways there is cooperation between the pathways in some aspects of dsRNA processing. For example the nuclease, which targets hyper-edited dsRNA for cleavage was identified to be Tudor-SN a component of the RISC complex of RNAi (Scadden, 2005). However, it is unknown whether the RISC associated Tudor-SN can cleave ADAR modified RNA. Also recently it has been recognised that the tissue specific ADAR modification of miRNA can generate functional

isoforms that target different sets of genes (Kawahara et al., 2007). The ADAR editing machinery is also implicated in heterochromatin formation in higher eukaryotes and mammals. Hyper-edited RNA are though to mediate heterochromatin formation through their binding of the protein vigilin in the nucleus (Wang et al., 2005). This may complement the previously proposed RNAi dependant process or act independently.

### *2.2.7 Transposon suppression*

It was first hypothesised that the process now referred to as RNAi may be involved in the repression of transposable elements in an early review of the co-suppression phenomenon in plants (Flavell, 1994). However, experimental evidence supporting this theory did not emerge until RNAi pathway mutants were studied in *C.elegans* (Tabara et al., 1999). This study identified several RNAi-deficient mutants, some of which resulted in the mobilisation of endogenous transposable elements. Further support for the RNAi mediated suppression of transposable elements in *C.elegans* then emerged from the identification of transposon specific siRNA (Sijen and Plasterk, 2003). Similar observations in *D.melanogaster* have also identified some RNAi pathway mutations to relieve transposon silencing (Aravin et al., 2001) and the accumulation of transposable element short RNA (rasiRNA) (Vagin et al., 2006). However, these rasiRNA did not require the Dicer proteins involved in miRNA and siRNA production and were observed to direct an alternative targeting complex containing a Piwi member of the argonaute family making them more similar to the recently identified piRNA (Aravin et al., 2006, Girard et al., 2006).

The suppression of transposable elements through the RNAi pathway has also been observed in plants. Originally mutations of the RNAi pathway were only found to relieve the repression of some transposable elements (Lippman et al., 2003, Zilberman et al., 2003). However, plant transposon repression via the RNAi pathway appears to be achieved through a slightly different mechanism to that of animals. This is because plant transposon suppression requires RNA polymerase IV, which is unique to plants for the accumulation of transposon specific siRNA (Pontier et al., 2005). The two isoforms of RNA polymerase IV are both required for transposon suppression acting in different stages of siRNA

accumulation for the initiation and maintenance of transposon silencing (Pontier et al., 2005).

Due to these observations in other species and the conservation of the RNAi pathway, it has been hypothesised that mammalian transposons may also be repressed by short RNA acting through the RNAi pathway. However, at the outset of this project there was little evidence supporting this hypothesis, with no repeat-associated short RNA identified in mammals. Therefore it was unknown if the RNAi pathway was involved in repressing mammalian transposable elements. However, more recently support for the susceptibility of mammalian transposons to RNAi mediated repression has been obtained. This includes some Dicer mutants relieving transposon repression in mammals (Svoboda et al., 2004), the repression of an active transposons using synthetic short RNA (Soifer et al., 2005) and the identification of a human L1 transposon short RNA from the promoter region which associates with the RNAi pathway (Yang and Kazazian, 2006). However, transposon short RNA have yet to be shown to repress transposons *in vivo*.

### *2.2.8 miRNA in cancer*

Mammalian cancers are recognised as having altered epigenetic states. These alterations often involve the hyper-methylation of promoter regions that result in the inappropriate transcriptional silencing of tumour suppressor genes and the global hypo-methylation of the genome (Jones and Baylin, 2002). In addition to these epigenetic alterations, miRNA expressed in adult tissues are deregulated in many cancers (Lu et al., 2005). Whilst this deregulation includes the up-regulation of some miRNA it generally results in a global down-regulation of miRNA in cancer (Iorio et al., 2005, Lu et al., 2005). The down-regulation of miRNA is also associated with undifferentiated cells, therefore may help maintain cancer cells. Reduced levels of miRNA are also reported to promote cell transformation (Kumar et al., 2007).

The first observation of miRNA deregulation in cancer was made in chronic lymphocytic leukaemia where hsa-mir-15a and hsa-mir-16-1 were mapped within a minimal deletion associated with >50% of all cases (Calin et al., 2002). Subsequently the tissue specific deregulation of various miRNA has been

observed in many cancers and can be used to classify cancers of unknown origin (Iorio et al., 2005, Lu et al., 2005). The deregulation of miRNA as a genuine feature of cancer is further supported by the mapping of >50% of known miRNA to fragile sites in the genome frequently altered in cancer (Calin et al., 2004). In addition to miRNA deregulation, mutations of the argonaute proteins, which associated with miRNA targeting have also been associated with cancer (Carmell et al., 2002).

### *2.2.9 Strategies to identify novel short RNA associated with RNA interference*

There are a number of experimental methods that can be employed to identify novel short RNA. The most extensively used of these to date have been northern hybridisation, large-scale short RNA sequencing and bioinformatic prediction.

Northern hybridisation was the first method used to identify short RNA in plants (Hamilton and Baulcombe, 1999). This method as well as being simple and reliable is very informative, providing sequence specific information on the abundance and the size of the short RNA and related RNAs. This ability to simultaneously detect related RNA also allows the discrimination of different classes of short RNA. In addition this method allows the quality of the RNA used and the discrete properties of the short RNA band to be easily assessed. Therefore this method has become regarded as the gold standard of short RNA verification. However, the efficient detection of short RNA by northern blot hybridisation does require a large amount of the short RNA. In some cases the abundance of short RNA under study is sufficient to compensate for the low sensitivity of this method.

Methods of large-scale short RNA sequencing in contrast are highly sensitive in detecting short RNA and yield a large number of short sequences. Traditionally cloning of the short RNA was used to generate hundreds of sequences. This was a useful tool for the identification of many of the first miRNA identified. However, the number of sequences generated made this an inefficient method of detecting low abundance short RNA. The deep sequencing of ~18-25nt size fractionated RNA by amplified tag sequencing (454) and massively parallel signature sequencing (MPSS) has overcome this hurdle, yielding many thousands

of sequences. However, this technology has only recently become accessible to general research. Various protocols have been used for both of these methods differing in the individual steps involved. These approaches often exploit the 5´ phosphate and 3´ hydroxyl properties of RNase III processing to enrich for Dicer products. However, within the large data sets of sequences generated many RNA degradation products are also sequenced. Therefore before analyses of the published data all repeat-associated RNA including rRNA, tRNA and transposons are removed under the assumption that they represent non-regulatory degradation products. Known short RNA sequences are then removed allowing the identification of potential novel short RNA sequences. Estimates of short RNA abundance can also be made from the frequency of individual sequences, under the assumption that all RNA are processed equally in preparation for sequencing and should be proportional to cellular abundance. The sequence information gained can also be used to identify short RNA isoforms by the identification of post-transcriptional modifications such as A-to-I editing (previously discussed). However, novel short RNA sequences identified by these methods require validation by other methods (e.g. northern blot) to ensure the discrete size fractionation of a single band and estimate their abundance.

The bioinfomatic approach to short RNA detection is predominantly used in the identification of miRNA. This method uses algorithms created using a number of characteristic features of known miRNA to predict new miRNA. These algorithms can be used to screen whole genomes or can be combined with the results of large-scale sequencing to identify sequences within regions of DNA capable of generating miRNA precursors. Various algorithms have been used to successfully identify many of the miRNA currently known in human. However, these algorithms produce many false positives making their subsequent verification labour intensive. The principles of this approach can be very useful when investigating a specific region of the genome for potential miRNA but are problematic for use in whole genome screens.

### *2.2.10 Uses of RNA interference in science and medicine*

Since its discovery, RNAi has been recognised as a powerful tool for genetic manipulation and has been exploited for both basic and clinical research purposes.

In basic research the ability to selectively repress specific genes either singly or in combination through the RNAi pathway has been used extensively (Agrawal et al., 2003, Elbashir et al., 2001, Elbashir et al., 2001). These uses include the introduction of siRNA or shRNA to investigate genes and their function, characterise processing pathways and the identification of potential drug targets. The used of the RNAi pathway for these investigations is attractive because of the simplicity of the targeting siRNA or shRNA design, the ease of application and efficiency of suppression. Therefore together these have made the RNAi pathway the quickest, cheapest and most flexible method of genetic manipulation of mammalian cells and organisms.

In addition to the research applications RNAi also offers potential therapeutic applications for the treatment of human disease. The first proof of principle for the therapeutic use of mammalian gene silencing by siRNA and shRNA came from an *in vitro* mouse study (McCaffrey et al., 2002). This showed the targeted suppression of a luciferase reporter RNA and hepatitis C pathogenic RNA in the liver. These sequences were introduced by hydrodynamic injection along with either synthetic siRNA or shRNA plasmid. However, despite the early promise the safe and effective systemic delivery of the siRNA and shRNA into higher mammals has encountered many obstacles. Despite these problems localised delivery has proved an effective method of siRNA delivery when the disease is restricted to a specific area of the body. Areas successfully targeted by localised delivery include: the lungs (intranasally); the eye (intraocularlly) and solid tumours (intertumoral).

Other obstacles to the therapeutic use of siRNA and shRNA therapy include: avoiding alteration of naturally occurring short RNA activity by competition with the therapeutic short RNA; off site targeting of the therapeutic short RNA and induction of the interferon response by the therapeutic short RNA. However, the ability to overcome these obstacles (although currently unknown) would result in a medical treatment able to targeting every gene in the genome, including those that cause disease through non-coding RNA, which cannot be targeted by drugs.

## 2.3 Experimental approach for the identification of novel repetitive short RNA in mammals

### *2.3.1 Why investigate short RNA specific to repetitive elements*

Transposable elements account for almost half of mammalian genomes. Despite the abundance of transposon sequence almost all become transcriptionally silenced and condensed into regions of heterochromatin (Martienssen, 2003). This makes the silencing of transposons one of the largest epigenetic events to occur in the cell. However, the process by which transposon silencing is mediated in mammals is poorly understood.

At the outset of this project short regulatory RNA acting through the RNAi pathway had been implicated in the silencing of repetitive elements including transposons in several organisms (2.2.7) and were hypothesised to also be involved in the silencing of mammalian transposons. However, no transposon-associated short RNA had been verified in mammals. Because the short transposon-associated RNA were essential to this hypothesis, this project sought to investigate if short transposon-associated RNA were expressed in mammalian cells. For this a candidate-lead approach was adopted in which attempts were made to detect small RNA from high abundance transposable elements by gel blotting. This was done to maximise the detection and relevance of any short RNA observed.

### *2.3.2 Method of detecting short transposon derived RNA and EDC cross-linking*

Short regulatory RNA have the common feature of their sizes falling within a narrow, defined range of 19-31nt. The precise processing of these RNA and related longer RNA is a valuable identifying feature and can provide information as to the class and function of the regulatory RNA.

Northern hybridisation provides the most convincing way to accurately display the size distribution of small RNA and is often employed to identify and validate short regulatory RNA. In addition to this, the method also has the benefits of allowing the identification of regulatory RNA that cannot be currently predicted

by computer-based approaches (siRNA and piRNA), being technically straight forward and relatively inexpensive. Despite these properties, this technique has not been extensively used in mammals. This is because the low sensitivity of the traditional method, makes the low levels of endogenously expressed mammalian small regulatory RNA very difficult to detect. However, the northern hybridisation method has been modified in the laboratory to increase the sensitivity of detection by 25-50 fold (Pall et al., 2007). Therefore this new method was selected to assess if short transposon derived RNA were expressed in mammals.

The increase in sensitivity of the northern hybridisation method was achieved by altering the RNA cross-linking step to the nylon membrane. This was done because traditional UV cross-linking was thought to limit the availability of short RNA for hybridisation, likely by the formation of covalent bonds between nucleotide bases and free amine groups at the nylon membrane surface. Instead RNA was chemically cross-linked, exploiting the 5` phosphate generated during Dicer cleavage (Zhang et al., 2004). This method uses the water-soluble carbodiimide, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), which requires a terminal phosphate to cross-link the RNA to the nylon membrane and results in the immobilisation of the RNA with its full-length assessable for hybridisation. A draw back of this method is that only RNA with a terminal phosphate will be immobilised on the nylon membrane. However, the detection of siRNA, miRNA and piRNA have all been shown to be greatly improved by this method (Pall et al., 2007).

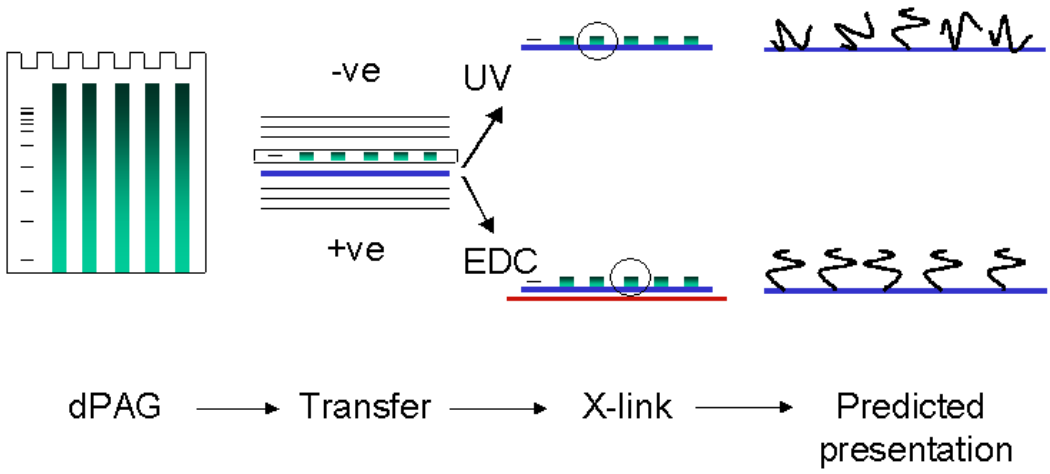


**Figure 6-Northern hybridisation method using UV or EDC cross-linking.** Image taken from (Pall *et al*., 2007 in press NAR). This shows the denaturing polyacrylamide gel (dPAGE) electrophoresis followed by the electro transfer of the RNA onto RNA membrane. Cross-linking (X-link) of the RNA can then be achieved using the UV or EDC method. With UV the covalent cross-linking of RNA residues to the nylon results in the reduction in the length of uninterrupted sequence available for hybridisation. Where as the chemical cross-linking of the RNA to the nylon membrane, with EDC uses a terminal phosphate and results in the maximum length of RNA available for hybridisation.

### *2.3.3 The selection of cell lines*

Various retrotransposons are reported to have a developmentally regulated expression pattern (Peaston et al., 2004), with high expression in the early embryo (Evsikov et al., 2004) followed by dramatic reduction later in development. Thus it was reasoned that any short RNA regulating transposon silencing would likely be expressed during this time. However, the quantity of RNA required for analysis by gel blotting could not be obtained from an embryonic source. As an alternative embryonic cell lines, which may mimic this period of mammalian development were used as these could be grown clonally in bulk and also made to differentiate by simple alterations of culture conditions.

Transposable elements often become transcriptionally active in cancer. Therefore to investigate if transcriptionally active transposons were being processed to generate short regulatory RNA, cancer cell lines were also assessed.

# 2

# Materials and methods

## 2.4 Cell culture

### *2.4.1 Culture of embryonic cells*

Mouse embryonic stem cells (ES XX pgk121, ES XY Efc/1), embryonic carcinoma cells (P19) and human embryonic carcinoma cells (NT2) were grown under standard conditions of 5% $CO_2$ at 37ºC. Undifferentiated ES cells were grown on gelatinised tissue culture plastic ware in high glucose DMEM (GIBCO) supplemented with 20% foetal bovine serum (Autogen Bioclear), 0.1mM non-essential amino acids, 2mM L-glutamine, 0.06mM β-mercaptoethanol and $10^3$ units/ml of leukaemia inhibitory factor (LIF) (Chemicon). The spontaneous differentiation of mouse ES cell lines was achieved by passaging the cells into ES culturing medium without LIF. The LIF minus medium was then refreshed every second day for the duration of the differentiation.

Undifferentiated mouse P19 and human NT2 embryonic carcinoma cells were grown on untreated tissue culture plastic ware in the absence of feeder cells in high glucose DMEM (GIBCO). The medium was supplemented with 10% foetal bovine serum (Autogen Bioclear), 0.1mM non-essential amino acids, 2mM L-glutamine and 0.06mM β-mercaptoethanol. P19 and NT2 embryonic carcinoma cells lines were differentiated by the addition of all-trans retinoic acid to a final concentration of 1μM and 10μM respectively.

The ES and P19 EC cell lines were passaged every second day when ~80-90% confluent and the NT2 cells were grown at high density >50% and regularly passaged every two to three days.

### *2.4.2 Culture of cancer cell lines*

The human cancer cell lines MDA-MB-435, MDA-MB-456, Hela (ECCAC-sigma) and MCF-7 were grown under standard conditions of 5% $CO_2$ at 37ºC. The Hela and MCF-7 cells were grown on tissue culture plastic ware in high glucose DMEM (GIBCO) supplemented with 10% foetal bovine serum (Autogen Bioclear), 0.1mM non-essential amino acids and 2mM L-glutamine. The cells were passaged every three to four days.

The human cancer cell line CaCo-2 was also grown under standard conditions of 5% $CO_2$ at 37ºC. These cells were grown on tissue culture plastic ware in high glucose DMEM (GIBCO) supplemented with 10% foetal bovine serum (Autogen Bioclear) and 0.1mM non-essential amino acids. The cells were maintained at high confluency (>50%) passaging every 3-4 days. The CaCo-2 cells lines were differentiated by the addition of sodium butyrate at final concentration of 2μM or 5μM. The media was replaced every 24 hours and the cells harvested after three days of differentiation.

## 2.5 Tissue collection

### *2.5.1 Mouse tissue*

The mouse samples were obtained from 3 month old male mice (C57BL/6) culled using a schedule 1 method in accordance with the UK Home Office regulations and frozen immediately in liquid nitrogen.

### *2.5.2 Human tissue*

The human breast material was obtained from the pathology department at the Western Infirmary, Glasgow and was anonymous and randomly selected with approval of the Local Research Ethics Committee having been stored in liquid nitrogen.

## 2.6 Transfection

### *2.6.1 Plasmid transfection*

Plasmid transfections were used to assay any potential targeting of the endogenous short repeat-associated RNA identified in chapters 5 and 6. This used the Firefly/Renilla Dual-Luciferase Reporter™ assay system with modified pGL3 (Firefly) containing hypothetical target sites to the short RNA and unmodified pHRL (Renilla) as an internal control. Typically the co-transfection of the plasmids was carried out in 6 well tissue culture plastic ware using the Lipofectamine™2000 (invitrogen) transfection agent. Cells were seeded the day before the experiment in their normal growth medium in sufficient number to

achieve a confluence of ~50% at the time of transfection ($10^3$-$10^5$ cells dependant on cell type). With the embryonic stem cells, differentiating culture medium was used. The medium was replaced before the transfection of the plasmid constructs (combination of pHRL and modified pGL3 plasmid typically at a 1 to 10 ratio) according to the manufacturers instructions. All experiments were carried out alongside the combined transfection of pHRL and unmodified pGL3, the individual transfection of each and a no transfection control. The cells were then harvested 24h post-transfection for analysis.

### *2.6.2 SiRNA transfection*

SiRNA transfections were typically carried out in 6 well tissue culture plastic ware. Healthy growing adherent cells were trypsinized (standard protocol) less than one hour before transfection and diluted in their normal growth medium to $10^5$ cells/ml and maintained at 37ºC. The siRNA transfection agent was then prepared by diluting 5µl of siPORT™ *NeoFX*™ (Ambion) into 95µl of Opti-MEM® (GIBCO) for each 6 well transfection. This mixture was allowed to stand for 10 minutes at room temperature. The siRNA were then diluted to the required concentration in 100µl Opti-MEM® (GIBCO) for each transfection. The transfection agent and siRNA dilutions were then combined for each transfection, mixed well and allowed to stand for 10 minutes to allow the transfection complexes to form. After this time the newly formed transfection complexes were dispensed into the empty 6 well plates. 2.3ml of diluted cells was then added to each well. The cells were then incubated for the required length of time before sampling at various time intervals post-transfection, to assess the siRNA-induced target gene knockdown. All experiments were carried out alongside the transfection of an siRNA with no target sequence in the genome such as GFP and a no transfection control.

| SiRNA name | Strand | Company | Sequence |
|---|---|---|---|
| Drosha 1 | Sense | Dharmacon | GGAAUUAUAUGACUGGAAUuu |
| | Antisense | Dharmacon | AUUCCAGUCAUAUAAUUCCuu |
| Drosha 2 | Sense | Dharmacon | CAACAUAGACUACACGAUUuu |
| | Antisense | Dharmacon | AAUCGUGUAGUCUAUGUUGuu |
| Dicer 1 | Sense | Ambion | GGUACUUAGGAAAUUUCGAtt |
| | Antisense | Ambion | UCGAAAUUUCCUAAGUACCtc |
| Dicer 2 | Sense | Ambion | GGUUCUCAUUAUGACUUGCtt |
| | Antisense | Ambion | GCAAGUCAUAAUGAGAACCtg |
| GFP | Sense | Qiagen | GCAAGCUGACCCUGAAGUUCau |
| | Antisense | Qiagen | GAACUUCAGGGUCAGCUUGCcg |

**Table 1-Pre-annealed commercial siRNA.** The above table contains the sequences of the pre-annealed siRNA used and the companies purchased from. The upper case nucleotides represent the annealed sequence with the lower case representing the 2nt 3` overhang.

# 2.7 Transformation

## *2.7.1 Transformation of competent E.coli*

Competent bacteria (subcloning efficiency DH5α competent E.coli™) were purchased commercially (Invitrogen). Competent cells were thawed on ice, then incubated with approximately 1ng of DNA, mixed gently and incubated on ice for 30 minutes. The cells were then subjected to heat-shock for 20 seconds at 37ºC and placed on ice for two minutes. The transformation volume was then made up to 1ml with pre-warmed LB medium and shaken for 1 hour at 37ºC. Typically 200μl of this medium was spread on LB agar plates containing a selective antibiotic (100μg/ml) and incubated overnight at 37ºC. Additionally, negative (no plasmid DNA) and positive plasmid control transformations were carried out. Individual colonies of interest were picked for further analysis.

## *2.7.2 Bacterial culture*

Generating large quantities of a plasmid generally involved the inoculation of an individual colony of *E.coli* into 50ml of antibiotic selective LB media overnight 37ºC. However, smaller volumes of selective LB medium and shorter incubation times were used to generate smaller plasmid yields.

## 2.8 Nucleic acid isolation and quantification

### *2.8.1 Extraction of genomic DNA from cultured cells*

Cultured mammalian cells were trypsinised and washed in PBS at room temperature and centrifuged (5 minutes at 1000xg room temperature) to yield a cellular pellet. The DNeasy kit™ (Qiagen) was then used to harvest genomic DNA from the cell pellet, according to the manufacturer's recommended instructions. The genomic DNA was always re-suspended in $dH_2O$ to allow for quantification.

### *2.8.2 Extraction of plasmid DNA from bacterial culture*

Transformed bacterial cultures were centrifuged to yield a pellet of bacteria (5 minutes at 6000xg and 4ºC) from which the plasmid DNA was to be extracted. A Maxi plasmid kit™ (Qiagen) was typically used for the extraction of plasmid DNA from 50ml bacterial cultures, according to the manufacturer's recommended instructions. However, other Qiagen plasmid kits were occasionally used for larger and smaller bacterial cultures. The plasmid DNA was always re-suspended in $dH_2O$ to allow for quantification.

### *2.8.3 Extraction of total RNA from cultured cells*

Cultured mammalian cells had their medium removed prior to the cells being lysed and homogenised by the addition of TRI-Reagent® (Sigma) directly into the culturing plastic ware on ice. Total RNA was then extracted according to the manufacturers instructions with the following additional steps. Samples were extracted with an equal volume of Tris/HCl buffered phenol/chloroform (pH 4.5) (Sigma) after the standard chloroform extraction and prior to precipitation with an equal volume of isopropanol. The nucleic acid pellets were re-dissolved in DEPC-treated $H_20$ and treated with TurboDNase® (Ambion) in the manufacturers buffer at approximately 1 unit per 100μg of RNA at 37ºC for one hour. The RNA was then extracted with an equal volume of Tris/HCl-buffered phenol/chloroform (pH4.5) (Sigma) prior to precipitation with isopropanol and one-tenth volume of 3M sodium acetate (pH6). The RNA was re-dissolved in DEPC-treated water (double distilled $H_2O$ containing 0.1% diethylpyrocarbonate

(DEPC) allowed to stand for 1 hour at room temperature prior to autoclaving at 15psi (1.05kg/$cm^2$) for 15 minutes).

### *2.8.4 Extraction of RNA from tissue*

Animal tissues frozen in liquid nitrogen were ground to a fine powder prior to homogenisation in Tri-Reagent® (Sigma). Total RNA was then extracted according to the manufacturers instructions including the additional steps described for cell culture RNA collection and DNA removal (2.8.3).

### *2.8.5 Nucleic acid quantification*

DNA and RNA dissolved in distilled or DEPC-treated $H_2O$ was quantified by measuring the absorbance at 260nm using a Nanodrop® UV spectrometer (Labtech). After quantification the DNA was stored at 4-8ºC and RNA was stored at -20ºC with the addition of deionised formamide (Sigma) to a final concentration of 50% for long-term storage.

## 2.9 Polymerase chain reaction

### *2.9.1 Standard PCR*

Standard PCR was typically carried out using *Taq* DNA polymerase (Promega) for the routine amplification of cDNAs from genomic DNA. The manufacturer's recommended protocol was applied and optimised as appropriate. Each reaction contained 1X PCR buffer (minus $MgCl_2$), 1.5mM $MgCl_2$, 0.2 mM each dNTP, 0.5μM forward and reverse primers and 1.25 units of *Taq* DNA polymerase in a final concentration of 25μl. Cycling was carried out in a thin-walled 0.2ml PCR tubes (ABgene) in a Biometra T3 thermocycler. Annealing temperatures were dependant on the primers used, ranging from 40-60ºC. Filter tips were used (ABgene).

Cycling conditions were typically:
94ºC for one minute to ensure template denaturation.
25-30 cycles; denaturing 94ºC,15 s/ annealing 40-60ºC, 1-15 s/ extension 72ºC
Followed by 5 minutes at 72ºC

### *2.9.2 Reverse transcriptase RT-PCR*

RT-PCR analysis of RNA requires an initial step of reverse transcription of the RNA followed by amplification of the resultant cDNA using standard PCR methods. The reverse transcription of total or gel purified RNA was generally carried out using 50-200 units Superscript™ II (Invitrogen) reverse transcriptase with 100ng random hexamers primers/25μl reaction (Roche). Typically 1-5μg of total RNA was used for the reverse transcription following the manufacturers instructions. The reverse transcription reaction was terminated by a 15 minute incubation at 70ºC and the RNA of the RNA/DNA duplexes degraded by the addition of 2 units of RNase H (Invitrogen) for 20 minutes at 37ºC. The synthesised cDNA was then used with specific primers for each PCR reaction.

## 2.10 Rapid amplification of cDNA ends (RACE)

### *2.10.1 5`RACE*

RNA of the required size was gel purified from denaturing-polyacrylamide gel (2.11.4). 200ng of the gel purified RNA was then ligated with 20 units of T4 RNA ligase (New England biolabs) to 200pmols of the 5` adaptor 5`OH-AGGGAGGACGAUGCGG-OH3`, with the addition of Super RNase inhibitor (Ambion) in the manufacturers buffer at 16ºC for ~16 hours. After the ligation the ligated RNA was separated from the non-ligated and adaptor RNA by denaturing-polyacrylamide gel electrophoresis (2.11.3) and purified (2.11.4).

RT-PCR was carried out using the gel purified, ligated RNA using random hexamers (Roche) for reverse transcription and DNA polymerase PFU (Promega) with primer pairs specific to the adaptor and regions of the 5S rRNA transcript in the PCR (Table 2). The Various PCR products were then separated by denaturing-polyacrylamide gel electrophoresis and size fractions of the desired size purified. The gel purified PCR products were then phosphorylated prior to ligating into plasmid the T-easy plasmid following the manufactures instructions (Promega).

| Primer name | Primer sequence |
|---|---|
| 5` adapter | AGGGAGGACGCTGCGG |
| 5S 80nt 1 | GCC$^{G/}{}_{A}$ACCCTGCTTAGCTT |
| 5S 80nt 2 | TCTCCCATCCAAGTACTAACC |

**Table 2- 5` RACE primers.** PCR reactions were carried out using 5`adaptor/5S 80nt 1 and 5`adaptor/5S 80nt 2 primer pairs. The sequences are shown above.

## 2.11 Gel electrophoresis and purification DNA and RNA

### *2.11.1 Agarose gel electrophoresis*

RNA and DNA were separated by electrophoresis in agarose gel (0.8-4%) buffered with 0.5X TBE (90mM Tris, 90mM boric acid and 2mM EDTA). Prior to loading, 6X loading dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/w) glycerol in $H_2O$) was added to the samples to a final concentration of 1X. After gel electrophoresis, gels were stained with ethidium bromide (EthBr) and imaged on a FLA-5000 system (Fuji) with Aida Imaging Analyser software to record the amount and distribution of RNA or DNA. The 1kb+ DNA marker (Invitrogen) prepared according to the manufacturer's instructions was run alongside samples to allow the estimation of size.

### *2.11.2 Purification of nucleic acids from agarose gels*

DNA fragments were excised from the gel using a clean scalpel blade and the DNA extracted using a Gel Extraction kit (Qiagen) according to the manufacturer's instructions. Alternatively electrophoresis was used to elute the DNA or RNA into 0.5 TBE using dialysis membrane (Medicell international) and subsequently precipitated with an equal volume of isopropanol and one tenth volume of 3M sodium acetate (pH 6) before being re-dissolved in $dH_2O$.

### *2.11.3 Denaturing-polyacrylamide gel electrophoresis*

RNA samples dissolved in DEPC-treated $H_2O$ prior to the addition of deionised formamide to a final concentration of 50% (Sigma) and 1X loading dye were heat denatured at 95ºC for 2 minutes before snap cooling on ice. Alternatively RNA dissolved in DEPC-treated $H_2O$ were denatured with the addition of an equal volume of glyoxal solution (Ambion) at 50ºC for 1 hour where described. The RNA

samples were then separated by electrophoresis using the Protean II rig (Bio-Rad) on either 10% or 15% polyacrylamide (19:1) gels cast in 7M urea and buffered with 20mM MOPS/NaOH (pH7). The electrophoresis buffer was 20mM MOPS/NaOH (pH7). RNA markers were generated from the $\gamma^{32}$ P-ATP end labelling of Decade® RNA markers (Ambion) prepared according to the manufacturers instructions. After gel electrophoresis, gels were stained using ethidium bromide (EtBr) and imaged on a FLA-500 system (Fuji) with Aida image analyser software to visualise and record the amount and distribution of RNA.

### *2.11.4 Purification of nucleic acids from polyacrylamide gels*

Regions containing the RNA or DNA of interest were excised from the gel using a clean scalpel blade. The DNA or RNA was then eluted overnight in a buffer of 20mM Tris pH 8.5, 300mM NaCl and subsequently precipitated with an equal volume of isopropanol and one tenth volume of 3M sodium acetate (pH 6) before being re-dissolved in $dH_2O$.

## *2.12 In vitro Dicer digestion*

### *2.12.1 In vitro Dicer digestion*

RNA of the required size when compared with the parallel size mark Decade (Ambion) was gel purified from a denaturing-polyacrylamide gel (2.11.4). This was then incubated with recombinant human Dicer (1 unit/500ng) (Stratagene) at 37ºC for 16 hours in the manufactures buffer. Routinely an equal amount of RNA was incubated in the reaction buffer minus the recombinant Dicer for the same length of time and at the same temperature as a negative control. Subsequent to Dicer treatment the samples were extracted with an equal volume of Tris/HCl buffered (pH4.5) Phenol/chloroform prior to precipitation with an equal volume of isopropanol and one-tenth volume of 3M sodium acetate (pH6). The resulting pellets were then re-dissolved in DEPC-treated $H_2O$.

## 2.13 Northern blotting of small RNA

### *2.13.1 Northern blotting*

After dPAGE, gels were placed on top of a sheet of nylon hybridisation membrane (Hybond-NX® Amersham/Pharmacia). This was then sandwiched between pieces of 3MM® Whatman filter paper wetted in distilled $H_2O$ (three layers each side) and placed in a "semi-dry" electroblotter (SciPlas). Excess liquid and air bubbles were removed from the sandwich to ensure even conductivity across the gel by rolling a clean pipette on the surface. The electophoretic transfer of the RNA from the gel onto the nylon membrane was carried out at 20V for 60 minutes/4ºC.

### *2.13.2 EDC cross-linking*

Immediately prior to use a solution of 0.16M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)(Sigma) was prepared in 0.13M 1-methylimidazole at pH 8 (adjusted with HCl). 3MM® Whatman filter paper cut just larger than the size of the nylon to be cross-linked was placed on a 20 X 20cm glass plate and was then saturated in the cross-linking solution. Excess solution was removed by draining. The nylon membrane containing the transferred RNA was then placed face up on the 3MM® and then wrapped in SARAN wrap and incubated at 60ºC for 60 minutes unless otherwise stated. The membrane was then rinsed in an excess of distilled $H_2O$ to remove any residual EDC prior to pre-hybridisation.

### *2.13.3 UV cross-linking*

Following the blotting of RNA onto the nylon membrane the membrane was immediately placed in a UV irradiation cabinet (Stratalinker) and 0.24J of UV was used to cross-link the RNA.

### *2.13.4 Probe preparation*

End-labelling of DNA oligonucleotides (Sigma) was carried out (1pmol per labelling reaction) with 20 units of T4 Polynucleotide Kinase (New England Biolabs) and $\gamma^{32}$ P ATP (6000 Ci.m.mol$^{-1}$; 10 mCi.ml$^{-1}$) in a 10μl reaction volume

containing the manufacturer's buffer at 37ºC for one hour. Table 3 displays the probe sequences (Sigma).

**A**

| Alu antisense probes | Alignment to sense sequence | Probe sequence |
|---|---|---|
| 1 | 1-30 | GGCCGGGCGCGGTGGCTCACGCCTGTAATC |
| 2 | 21-50 | GCCTGTAATCCCAGCACTTTGGGAGGCCGA |
| 3 | 41-70 | GGGAGGCCGAGGCGGGCGGATCACGAGGTC |
| 4 | 61-90 | GTCACGAGGTCAGGAGATCGAGACCATCCTG |
| 5 | 81-110 | GACCATCCTGGCTAACACGGTGAAACCCCG |
| 6 | 101-130 | GTGAAACCCCGTCTCTACTAAAAATACAAAA |
| 7 | 121-150 | GAAATACAAAAAATTAGCCGGGCGTGGTGGC |
| 8 | 141-170 | GCGTGGTGGCGGGCGCCTGTAGTCCCAGCT |
| 9 | 161-190 | GAGTCCCAGCTACTCGGGAGGCTGAGGCAGG |
| 10 | 181-210 | GCTGAGGCAGGAGAATGGCGTGAACCCGGGA |
| 11 | 201-230 | GAACCCGGGAGGCGGAGCTTGCAGTGAGCC |
| 12 | 221-250 | GCAGTGAGCCGAGATCGCGCCACTGCACTC |
| 13 | 241-270 | GCACTGCACTCCAGCCTGGGCGACAGAGCGA |
| 14 | 261-281 | GACAGAGCGAGACTCCGTCTC |

**B**

| 5S rRNA oligonucleotide probes | Alignment to sense sequence | Probe sequence |
|---|---|---|
| Sense specific 1 | 1-16 | GTGGTATGGCCGTAGAC |
| Sense specific 2 | 6-31 | GCGGGCGCGTTCAGGGTGGTATGGCC |
| Sense specific 3 | 21-46 | GAGATCAGACGAGATCGGGCGCGTT |
| Sense specific 4 | 36-61 | GCCCTGCTTAGCTTCCGAGATCAGAC |
| Sense specific 5 | 51-76 | GACTAACCAGGCCCGACCCTGCTTAG |
| Sense specific 6 | 66-91 | GTCTCCCATCCAAGTACTAACCAGG |
| Sense specific 7 | 81-106 | GCGGTATTCCCAGGCGGTCTCCCATC |
| Sense specific 8 | 96-121 | GAAAGCCTACAGCACCCGGTATTCCC |
| Antisense specific 1 | 1-25 | GTCTACGGCCATACCACCCTGAACG |
| Antisense specific 2 | 15-40 | GACCCTGAACGCGCCCGATCTCGTCT |
| Antisense specific 3 | 30-55 | GAACGCGCCCGATCTCGTCTGATCT |
| Antisense specific 4 | 45-70 | GCGGAAGCTAAGCAGGGTCGGGCCTG |
| Antisense specific 5 | 60-85 | GTCGGGCCTGGTTAGTACTTGGATG |
| Antisense specific 6 | 75-100 | GTACTTGGATGGGAGACCGCCTGGGA |
| Antisense specific 7 | 90-115 | GCCGCCTGGGAATACCGGGTGCTGTA |
| Antisense specific 8 | 105-121 | GGGTGCTGTAGGCTTT |

**Table 3- DNA oligonucleotide probe sequences.** The sequences in the tables were radioactively end-labelled for use as probes. Table A contains the overlapping oligonucleotide sequence used to map the ~70nt antisense RNA, table B the overlapping sense and antisense specific 5S rRNA oligonucleotide sequences used to map the short ~22nt and ~26nt RNA.

To generate probes of longer than >100nt PCR (2.9.1) was used to amplify DNA from genomic DNA or overlapping DNA oligonucleotides. The primers used typically contained 5` specific sequence corresponding to the T7 or T3 bactereophage prompter sequences to generate amplified DNA with a 5` T7 (blue) and 3` T3 (green) RNA polymerase recognition sites. The DNA were then purified from agarose gels (2.11.2) before use as a template for *in vitro*

transcription. Table 4 shows the sequences of the PCR primers used to generate the probe templates (Sigma).

The probes were generated by *in vitro* transcription, using $\alpha$-$^{32}$P UTP (800Ci.mmol$^{-1}$; 10 mCi.ml$^{-1}$) (Amersham) and 20 units of either T7 RNA polymerase (Promega) or T3 RNA polymerase (Promega) in a 20µl reaction volume containing the manufacturer's buffer "hot started" at 37ºC to generate individual strands of RNA complementary to sequences of interest. To create an easily strippable probe based on the protocol of the strip-EZ kit (Ambion) US patent 6891032, CTP was replaced with iodine sensitive cytidine-5'- (1-thiotriphosphate)(Ambion).

**A**

| Probe name | DNA oligonucleotide probe, primer sequence |
|---|---|
| Alu Y consensus | TAATACGACTCACTATAGGGCCGGGCGCGGTGGCTCAC |
| | AATTAACCCTCACTAAAGGAGACGGAGTCTCGCTCTGT |
| B1 consensus | TAATACGACTCACTATAGGCYRGGYRTGGTGGYRCAYR |
| | AATTAACCCTCACTAAAGYRAGACAGGGTTTCTCTGTR |
| 7SL | TAATACGACTCACTATAGGCTATTCTGCGCCGGTATCC |
| | AATTAACCCTCACTAAAGGTAACGTATTTATTGTTCAAA |
| L1A | TAATACGACTCACTATAGGCGGATCTGGGGCACAAGTCC |
| | AATTAACCCTCACTAAAGGAGGCCCGGGTAGCCTGCTTC |
| $L1_{GF}$ | TAATACGACTCACTATAGGTGAGAGCACGGGGGTCTGCC |
| | AATTAACCCTCACTAAAGGCCAGGAAGGTGGCCGGCTGT |
| $L1_{TF}$ | TAATACGACTCACTATAGGCCAGAGGAGAGGTGTCTGCC |
| | AATTAACCCTCACTAAAGGCAGGGAAGGTGCCCGGATGT |
| IAP | TAATACGACTCACTATAGGTGTTATTARCGCGTTCTCAC |
| | AATTAACCCTCACTAAAGGTGTTGGGAGCCGCGCCCACA |
| MuERV-L | TAATACGACTCACTATAGGTGTAGTGGTTATTCCTGGTTGT |
| | AATTAACCCTCACTAAAGGTGTATTAGTCAGGGTTCTCTTG |
| B2 | TAATACGACTCACTATAGGGGCTGGAGAGATGGCTCAG |
| | AATTAACCCTCACTAAAGGATTTATTTATTATATGTAAG |

**B**

| Probe name | Overlapping DNA oligonucleotide PCR primers |
|---|---|
| 5S rRNA T7 | TAATACGACTCACTATAGGTCTACGGCCATACCACCCTGAACGCGCCCG ATCTCGTCTGATCTCGGAAGCTAAGCAGGGTCGGGCCTGGTTAGTACTT |
| 5S rRNA T3 | AATTAACCCTCACTAAAGAAAGCCTACAGCACCCGGTATTCCCAGGCGG TCTCCCATCCAAGTACTAACCAGGCCCGAC |

**Table 4- PCR primers for probe templates.** Table A contains the primer sequences used to generate probe templates from genomic DNA with B containing the overlapping primer sequences used as the PCR template for the 5S rRNA (overlapping region highlighted in red). In both tables the forward primer contains sequence specific to the T7 RNA polymerase promoter (blue type), with the reverse primer having sequence specific to the T3 RNA polymerase promoter (green type).

The "T7 probe templates" referred to throughout the thesis were generated by annealing DNA oligonucleotides to the sequence to be hybridised (see Figure 7 for illustration) with an equal concentration of the DNA oligonucleotide TAATACGACTCACTATAGG (Sigma) to create a dsDNA region specific for T7 RNA

polymerase. The annealed oligonucleotides were then used at 250nM as templates for *in vitro* transcription with $\alpha$-$^{32}$P UTP (800Ci.mmol$^{-1}$; 10 mCi.ml$^{-1}$) (Amersham) and 20 units of T7 RNA polymerase (Promega) in a 20µl reaction volume containing the manufacturer's buffer at 22°C. As with the probes generated from the PCR templates, CTP was replaced with iodine sensitive cytidine-5'- (1-thiotriphosphate)(Ambion) to create an easily strippable probe.

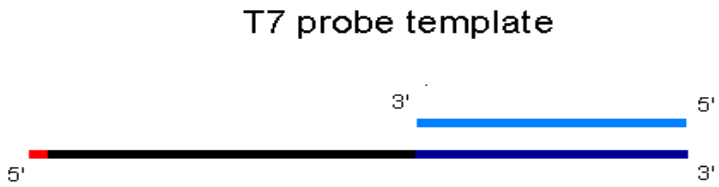


**Figure 7-T7 template structure.** The above image represents a T7 probe template. The light blue upper strand represents the T7 promoter sequence, which hybridises to the dark blue region of the lower strand (reverse complement of the T7 promoter). The black region of the lower strand indicates the sequence to be detected with the red region representing optional adenosine residues to increase the activity to the probe.

The sequences used to generate the lower strand oligonucleotide are shown in Table 5.

| Probe name | T7 DNA oligonucleotide probe sequence |
|---|---|
| B1 sense 26-44nt | AAAATCCCAGCACTTGGGAGGCCTATAGTGAGTCGTATTA |
| B1 sense 103-122nt | AAGAGGCCAGCCTGGGCTACACATGCCTATAGTGAGTCGTATTA |
| $L1_{TF}$ sense 41-62nt | AATGAGCATCTGCGGGAGACATCTGCCTATAGTGAGTCGTATTA |
| $L1_{TF}$ antisense 75-96nt | AAGAAGCAGTGTAGCTTCTGTAGTGCCTATAGTGAGTCGTATTA |
| IAP2 sense | AATGGCUGGUCGCGAGAACGCGGGCCTATAGTGAGTCGTATTA |
| IAP antisense | AACCCGCGTTCTCGCGACCAGCCACCTATAGTGAGTCGTATTA |
| MuERV | AATCATCTGGGATTAAAGGTGTGGCCTATAGTGAGTCGTATTA |
| B2 sense 1-20nt | AAGGGCTGGAGAGATGGCTCAGCCGCCTATAGTGAGTCGTATTA |
| B2 antisense 1-20nt | AATTGAGCCATCTCTCCAGCCCTCGCCTATAGTGAGTCGTATTA |
| B2 sense 35-58nt | AAACTGCTCTTCCAGAGGTCCTGCCTATAGTGAGTCGTATTA |
| 5S 122nt | AAGTCTACGGCCATACCACCCTACCTATAGTGAGTCGTATTA |
| 5S 77-98nt | AAGACTTGGATGGGAGACCGCCTGGCCTATAGTGAGTCGTATTA |
| 5S 92-113nt | AACGCCTGGGAATACCGGGTGCTGCCTATAGTGAGTCGTATTA |
| 28S (1) | AATCAACAAGTACCGTAAGGGAAACCTATAGTGAGTCGTATTA |
| 28S (2) | AATCCCCGCGGGGCCCCGTCGTCCCTATAGTGAGTCGTATTA |
| 18S | AACGAAAGCATTTGCCAAGAATGTCCTATAGTGAGTCGTATTA |
| Mir 15 | AATAGCAGCACATAATGGTTTGTGCCTATAGTGAGTCGTATTA |
| Mir 16 | AATAGCAGCACGTAAATATTGGCGCCTATAGTGAGTCGTATTA |
| Mir 21 | AATAGCTTATCAGACTGATGTTGACCTATAGTGAGTCGTATTA |
| Mir 292as | AAAAGTGCCGCCAGGTTTTGAGTGTCCTATAGTGAGTCGTATTA |
| Mir 98 | AATGAGGTAGTAAGTTGTATTGTTCCTATAGTGAGTCGTATTA |

**Table 5- T7 DNA oligonucleotide probes.** This table contains T7 template sequences with the sequence complementary to the T7 RNA polymerase promoter in blue text the probe sequence in black text and additional adenosine residues used to increase the specific activity of the probe in red text.

### 2.13.5 Probe Hybridisation

Pre-hybridisation and hybridisation of both EDC and UV cross-linked membranes was carried out in 2X SSC, 1% SDS and 100μg/ml sheared, denatured salmon sperm DNA (sonicated (misonex sonicator 3000) until an average length of ~500nt was achieved). Hybridisation temperatures and washes varied between 30-60ºC dependant on the experiment. The concentration of SSC used also varied between 0.1 and 0.2X in 0.2% SDS dependant of the stringency of wash required.

### 2.13.6 Probe removal

Method 1- Applicable to all probes, membranes were stripped by briefly placing them in a solution of 10mM Tris/HCl (pH 8.5 at 25ºC) at 100ºC for one minute.

Method 2- This allows for the specific degradation of RNA containing the modified iodine sensitive nucleotide cytidine-5'- (1-thiotriphosphate) (Ambion). The 200X degradation solution (500mM Iodine dissolved in 100% ethanol) based on the strip-EZ kit (Ambion) US patent 6891032 was diluted to 1X working solution in 0.1% SDS. This solution was mixed well before pouring over the membrane in a container, using no less than 10ml/$cm^2$ of membrane to be treated. This was incubated for 2 minutes at room temperature ensuring even coverage of membrane before incubating at 68°C for 10min. Excess liquid was then removed form the container. The 100X reconstitution solution (1M sodium thiosulfate diluted in $dH_2O$) was then diluted to a 1X working solution in 0.1% SDS, again 10ml/$cm^2$ membrane and add to the container. The membrane was incubated in this solution at 68°C for 10min. This solution was then discarded prior to a final wash of membrane with 0.1% SDS for 10min.

In both stripping methods probe removal was assessed by re-exposure of the membrane to a phosphor-imaging plate for at least as long as any previous exposure of the membrane. These images were recorded using a FLA-5000 system (Fuji) with Aida Image Analyser software.

## 2.14 Plasmid construction and validation

### *2.14.1 Sensor plasmid construction*

The pGL3 luciferase reporter vector (Promega) was digested using the XbaI restriction endonuclease (Invitrogen) in accordance with the manufacturers instructions to generate a vector linearised in the 3`UTR of the firefly luciferase gene. The linearised plasmid was then ligated with the sensor inserts, which were formed by annealing complementary DNA oligonucleotides with a four nucleotides (CTAG) 5´overhang at each end to generate complementary sticky ends (Table 6). The complementarity of the sticky ends did not lead to the reestablishment of an XbaI restriction site. Therefore the ligations were heated to denature the ligase prior to repeated treatment with the XbaI restriction endonuclease to reduce background from the re-ligation of plasmids.

| Sensor name | DNA oligonucleotide sequence |
|---|---|
| B2 1-23nt XbaI | CTAGGGGCTGGAGAGATGGCTCAGTGG |
| | CTAGCCACTGAGCCATCTCTCCAGCCC |
| B2 1-23nt XbaI control | CTAGGAGCTCGAGAGCTGGCTCCGTAG |
| | CTAGCTACGGAGCCAGCTCTCGAGCTC |
| 5S 1-25nt XbaI | CTAGGTCTACGGCCATACCACCCTGAACG |
| | CTAGCGTTCAGGGTGGTATGGCCGTAGAC |
| 5S 1-25nt XbaI control | CTAGGTCTACCGCCATTCCACCCGGAAAG |
| | CTAGCTTTCCGGGTGGAATGGCGGTAGAC |
| 5S 77-98nt XbaI | CTAGGACTTGGATGGGAGACCGCCTGG |
| | CTAGCCAGGCGGTCTCCCATCCAAGTC |
| 5S 77-98nt XbaI control | CTAGGAGTTTGATGGCAGACAGCCTGG |
| | CTAGCCAGGCTGTCTGCCATCAAACTC |
| 5S 92-121nt XbaI | CTAGCCTGGGAATACCGGGTGCTGTAGGCTTT |
| | CTAGAAAGCCTACAGCACCCGGTATTCCCAGG |
| 5S 92-121nt XbaI control | CTAGCCTGGGAGTACCGCGTGCTGTATGCTGT |
| | CTAGACAGCATACAGCACGCGGTACTCCCAGG |
| 5S SIAE XbaI | CTAGTTGTAAAGACGTCTACGGCCATATCACCCTGAACGCGCCTGATCTCATCTGA |
| | CTAGTCAGATGAGATCAGGCGCGTTCAGGGTGATATGGCCGTAGACGTCTTTACAA |
| Mir 15 | CTAGTAGCAGCACATAATGGTTTGTG |
| | CTAGCACAAACCATTATGTGCTGCTA |
| Mir 16 | CTAGTAGCAGCACGTAAATATTGGCG |
| | CTAGCGCCAATATTTACGTGCTGCTA |
| Mir 21 | CTAGTAGCTTATCAGACTGATGTTGA |
| | CTAGTCAACATCAGTCTGATAAGCTA |

**Table 6-Sensor construct DNA Oligonucleotides.** The DNA oligonucleotide sequences above were annealed in their pairs to form the insert to be ligated into the XbaI site of the pGL3 luciferase reporter vector. The nucleotides highlighted in yellow indicate polymorphisms in the control construct sequence.

### *2.14.2 Screening and sequencing*

The sensor ligations were transformed into competent *E.coli* (DH5α) and plated onto agar as described (2.7.1) to generate colonies. The individual colonies were then screened to test the direction of insertion of each sequence using PCR (2.9.1) with each of the DNA oligonucleotides used originally to generate the sensor construct with the plasmid specific primer ACGCAAGAAAAATCAGAGAGAT. For selected colonies identified to have inserts in the desired orientation, bacterial culture (2.7.2) was carried out along with plasmid purification (2.8.2). The resulting plasmids were then sequenced to confirm insertion (GATC-Biotech).

## 2.15 Protein extraction and analysis

### *2.15.1 Extraction of protein from cultured cells*

The luciferase sensor knockdown experiments were assayed using the Dual-Luciferase Reporter™ assay system (Promega). Protein was extracted from these samples using the passive lysis buffer contained within the Luciferase Reporter™ assay system (Promega) kit, *in situ* (in each well) or after trypsinisation and centrifugation to yield a cellular pellet in accordance with the manufacturer's recommended instructions.

### *2.15.2 Quantification of Firefly and Renilla luciferase expression by luminescence measurement*

The luciferase sensor knockdown experiments were assayed using the Dual-Luciferase Reporter™ assay system (Promega). This system provided an efficient means of performing two reporter assays in parallel, namely the activities of Firefly (*Photinus pyralis*) and Renilla (*Renilla reniformis*) luciferase. These are measured sequentially from a single sample obtained by the passive lysis of cultured cells (2.15.1). The Firefly luciferase reporter is measured first by adding the Luciferase Assay™ Reagent II to generate a luminescent signal after which the reaction is quenched and the Renilla luciferase reaction is initiated simultaneously by adding Stop and Glo™ reagent to the same sample. The luminescent signal was measured on a Luminskan Ascent (Thermo Labsystems)

luminometery system. Negligible background activity readings were obtained from non-transfected cell lysate controls.

# 3

# Novel Alu and B1 transposon derived small RNA (~70nt) detected in mammalian cell lines

## Results 1

## 3.1 Summary

This chapter outlines the initial investigation of transposon-associated short RNA for the SINEs Alu (human) and B1 (mouse). Using the MCF-7 (Human breast adenocarcinoma cell line), Hela (Human cervical carcinoma cell line) and P19EC (mouse embryonic carcinoma) cell lines as sources of RNA. A gel blotting approach was used in an aim to identify short RNA (~20-25nt) and was selected because a chemical cross-linking method developed within the laboratory (personal communication G, Pall) greatly improved the sensitivity of gel blotting for short RNA detection. It was reasoned that this increased the possibility of detecting low abundance transposon-associated short RNA, that may have previously gone undetected in mammals by previous gel blotting techniques.

Using this method a novel ~70nt antisense RNA was observed by hybridisation to the Alu and B1 full-length transposon consensus probes. However, no shorter RNA relating to these transposon sequences were observed. The ~70nt Alu RNA was partially mapped to a region of the Alu transcript using short overlapping DNA oligonucleotide probes of 30nt. Further analysis of the Alu ~70nt antisense RNA in respect to the RNAi pathway was then undertaken by the siRNA targeted knockdown of Drosha and Dicer RNase III enzymes. However, no link between Dicer or Drosha in the production or processing of the ~70nt antisense RNA was observed.

## 3.2 Introduction

### *3.2.1 7SL RNA derived SINE transposon*

Short Interspaced Elements (SINEs) is the term used to describe repeat sequences ~80-400nt propagated by retrotransposition within the genome. The majority of SINEs are derived from tRNA sequence, however two SINE families were found to originate from 7SL RNA (Vassetzky et al., 2003). In eukaryotes 7SL RNA (~300nt) associates with a number of proteins to form the signal recognition particle (SRP). The SRP is involved in translocation of newly synthesised proteins from the ribosome to the endoplasmic reticulum (Bovia and Strub, 1996). However, 7SL RNA-derived SINEs have only been identified in primates (Alu) and rodents (B1) suggesting recent evolutionary origins (Vassetzky et al., 2003).

As can be seen from Figure 8, Alu (~300nt) and B1 (~150nt) transposable elements differ considerably in size, but share much sequence similarity. The modern Alu dimer is a head to tail fusion of two similar but non-identical 7SL SINE monomers (left and right) joined by an A-rich linker. This Alu dimer subsequently underwent expansion in primates, predominantly in humans where it represents ~10% of the genome (Gibbons et al., 2004). The rodent specific B1 element also underwent expansion as a monomer and is most similar to the left Alu monomer, varying by a 29nt internal tandem repeat and 9nt deletion (Quentin, 1994). Both transposons contain RNA polymerase III transcription sites like the 7SL RNA gene, however few are actively transcribing and capable of transposition. Alu and B1 transposons are separated into several subfamilies, these represent different periods of master gene expansion during evolution through transposition (Kass et al., 2000, Shen et al., 1991).

The appearance of the initial 7SL RNA SINE master gene is thought to have occurred in an ancestor common to primates and rodents. Subsequent transposition and divergence of this original monomeric 7SL SINE is thought to have generated the monomers of the modern Alu and B1 transposons. This evolutionary route is supported by the identification of intermediate sequences in both human and mouse genomes (Quentin, 1994).

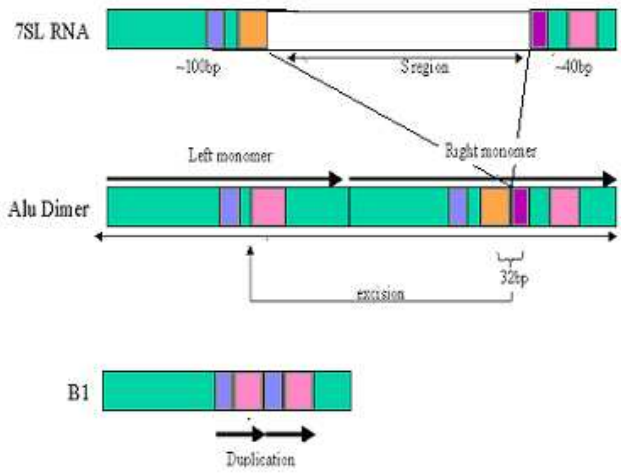


**Figure 8-Origins of B1 and Alu SINEs from 7SL RNA.** The above image represents the regions of 7SL RNA from which the B1 and Alu transposons are derived. The central (S region) of 7SL is absent, with Alu and B1 monomers containing sequence from terminal end fusion of the 7SL sequence. The Alu transposon is a fusion of two non-identical 7SL derived monomers the coloured sections highlight regions in association with the 7SL sequence (green= areas largely unchanged between the Alu and B1monomers with other colours representing regions of variation). The left monomer of the Alu dimmer differs from the right by a 32nt deletion (represented by orange and purple regions). The B1 transposon is also shown to resemble the left monomer of Alu, the most significant variation is the 29nt internal duplication (blue and pink regions).

Both Alu and B1 represent non-autonomous transposons and therefore are not capable of propagating their own transposition. Instead their retrotransposition is mediated by the autonomous long interspersed element L1 (LINE-1). Initially this was inferred by the identification of inverted repeats similar to those found in LINE-1 retrotransposition, flanking Alu insertions and similar sites of integration (Hagan et al., 2003). This hypothesis was further supported by the requirement of the LINE-1 ORF2-encoded endonuclease/reverse transcriptase in Alu transposition (Dewannieux et al., 2003). B1 mobilisation has also been linked with LINE-1, although at lower frequencies of retrotransposition than Alu. This lower transposition frequency is thought to be due to a single nucleotide mutation, at a site required for interaction with the SRP 9/14 proteins (Dewannieux and Heidmann, 2005). The SRP9/14 proteins therefore may be involved in positioning the SINE transcripts for use of the LINE-1 transposition machinery.

In addition to the Alu and B1 SINEs there is a third highly abundant 7SL RNA-like sequence: 4.5S RNA, thought to originate from mouse B1 element (Gogolevskaya

et al., 2005). The mouse 4.5S RNA is encoded by a 4.2kb repeating unit, containing a 90-94nt RNA polymerase III transcribed region (Gogolevskaya et al., 2005, Leinwand et al., 1982). There are ~800 copies of the 4.5S repeating unit arranged in tandem on chromosome 6, representing ~0.1% of the genome (Schoeniger and Jelinek, 1986). The function of 4.5S RNA is unknown, however it has been shown to form hydrogen bonds with poly (A) terminated RNA suggesting a possible function in association with mRNA (Harada and Kato, 1980, Leinwand et al., 1982, Schoeniger and Jelinek, 1986). The 4.5S RNA are found in both the nucleus and cytoplasm having a short half live of ~30min. The reason for this high turnover is unknown. In addition to the tandemly repeated copies an equal number of single copy 4.5S "pseudogenes" are scattered throughout the genome. These may have arisen from transposition, because inverted repeats have been identified flanking several dispersed 4.5S RNA copies (Kraft et al., 1992).

Alu and B1 transposons are known sites of *de novo* methylation, which have been shown to act as methylation centres spreading methylation to adjacent promoters in some cancers (Graff et al., 1997, Yates et al., 1999). The cause of *de novo* methylation at these and other transposable elements is unknown, however the RNAi pathway has been shown to methylate transposon sequences in other species as discussed in the introduction (2.2.5). The hypothesis was that if this was also the case in mammals, discrete short RNA of ~21nt would be expected to be detected with probes corresponding to Alu and B1 transposons.

## 3.3 Results

### *3.3.1 The differential detection of RNA <100nt using glyoxal and formamide denaturation*

Transposons are silenced early in mammalian development to protect the genome from the deleterious insertion of elements. However, prior to silencing, transposable elements are highly expressed in the pre-implantation embryo (Evsikov et al., 2004, Svoboda et al., 2004). Transposable elements can also be transcriptionally expressed in some cancers due to the reactivation of somatic cell transposons (Schulz, 2006). Therefore if short RNA were driving the silencing of transposons, the expression of short transposon-associated RNA may be restricted to the early embryo and some cancers because RNA-mediated silencing requires some transcription. The gel blotting method selected to identify any transposon-specific short RNA, requires large quantities of RNA not readily provided by whole embryo or tumour samples. Therefore cell lines of embryonic and tumour origin, were chosen for the analysis of short transposon-associated RNA.

Prior to the investigation of short transposon-associated RNA, it was assessed if the quantity of RNA being used would be sufficient for the detection of short RNA by the gel blotting method. Formamide is frequently used to denature RNA in gel blotting methods. However, formamide dissociates from the RNA during gel electrophoresis allowing some re-association of complementary RNA. In the case of siRNA, often typified as dsRNA of ~22nt, the re-association of dsRNA could reduce or prevent their detection. Therefore it was thought that if any transposon-associated RNA were double stranded like the siRNA, their detection may be improved by preventing re-association of the RNA after denaturation. Based on this, glyoxal RNA denaturation was selected for comparison to formamide, because it remains attached to the RNA by forming stable adducts with guanosine residues (Broude and Budowsky, 1971). Thus it prevents the re-association of the denatured RNA.

To assess the different methods of denaturation equal volumes and concentrations of complete RNA extracted from Hela cells were denatured using glyoxal for various lengths of time at 50ºC to create a time course. RNA

denatured using the standard formamide protocol was used for comparison. Because the glyoxal denatured RNA is retarded relative to formamide denatured RNA during gel electrophoresis (due to the association of glyoxal with guanosine residues), both glyoxal and formamide denatured size markers were generated to estimate RNA size. The denatured RNA and size markers were then separated by denaturing-polyacrylamide gel electrophoresis before electro-blotting onto nylon membrane and cross-linking using EDC. The RNA was then hybridised to probes specific to the micro RNA hsa-mir-21 and hsa-mir-98 consecutively. The probes were removed between hybridisations (2.13.6 method 1) and removal confirmed by phosphor-imaging. Micro RNA were selected for this analysis because they provided a validated source of human short RNA to assess the quantity of RNA used and any differential detection efficiency between the RNA denaturation methods.

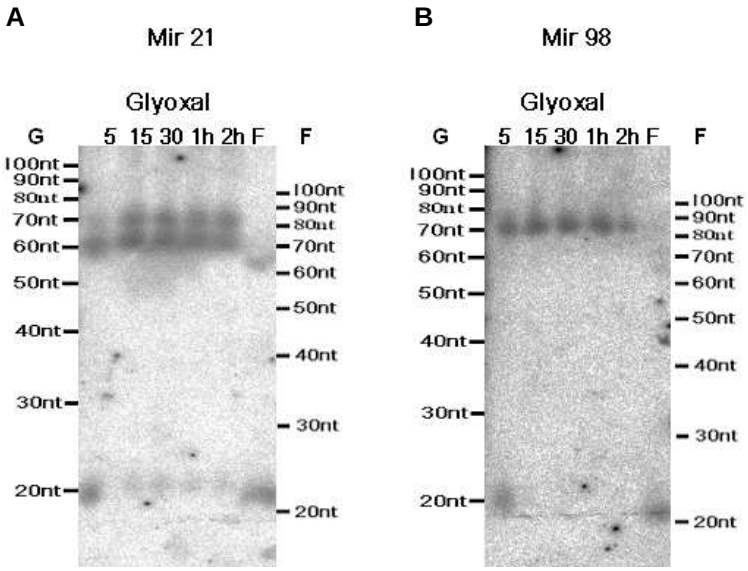


**Figure 9- Detection of miRNA in Hela RNA denatured with glyoxal or formamide.** Total RNA was extracted from Hela cells. ~10µg of RNA was denatured by glyoxal incubation for 5 minutes (5), 15 minutes (15), 30 minutes (30), 1 hour (1h), 2 hours (2h) or denaturation with formamide (F). The RNA samples were separated by 10% denaturing-polyacrylamide gel by electrophoresis before being electro-blotted and cross-linked to nylon membrane using EDC. Hela RNA was then hybridised in 2X SSC and 1% SDS buffer to the hsa-mir-21 (A) and hsa-mir-98 (B) probes at 40ºC, wash 0.2X SSC 0.2% SDS 40ºC. The radioactive signal was recorded by phosphor-imaging with 16 hour exposures are shown for both with marker sizes annotated on the left of the images for glyoxal and the right for formamide.

The hsa-mir-21 probe hybridisation, detected retardation and reduced abundance of mature hsa-mir-21 RNA for the glyoxal-denatured RNA (expected size of 22nt). The five-minute glyoxal-denaturation was the only exception to this, being more comparable in mobility and abundance to the formamide-

denatured sample. There was also variation in the appearance of the hsa-mir-21 precursor between glyoxal-denatured and formamide-denatured samples. The hsa-mir-21 precursor RNA is predicted to be 60nt. Both glyoxal and formamide-denatured RNA samples displayed a RNA band of the expected size, when compared to the relevant size marker. However, the glyoxal-denatured samples, with the exception of the five-minute treatment also detected an RNA of ~70nt.

The hsa-mir-98 probe hybridisation did not detect the mature hsa-mir-98 RNA (expected size 22nt) in the glyoxal-denatured RNA. The five-minute glyoxal-denaturation was the only exception to this detecting a RNA of ~20nt. The RNA detected in the five-minute glyoxal sample was comparable in abundance to that of the formamide-denatured sample. The hsa-mir-98 precursor RNA is predicted to be 80nt. The precursor again differs between samples with a predicted ~70nt RNA only being observed in samples denatured using glyoxal (Figure 9B).

From this analysis it was concluded that ~10μg of total RNA was sufficient for the detection of known short RNA when using formamide denatured RNA on a denaturing polyacrylamide gel. However, the detection of these RNA was substantially reduced when glyoxal denaturation was used. It is likely that the reduction in the detection of the mature miRNA was as a result of the incomplete removal of glyoxal from the RNA after EDC cross-linking step, resulting in a reduction of the sequence length available for hybridisation. However, with longer RNA this was not observed to present a problem, with apparent increased detection of the miRNA precursor RNA.

### *3.3.2 The unexpected detection of Alu and B1 antisense ~70nt RNA*

The transposable elements Alu and B1 were selected for the identification of transposon-specific short RNA. These transposons were selected because both Alu and B1 are known sites of de novo methylation (Graff et al., 1997, Yates et al., 1999). Also their high abundance in the human and mouse genomes was thought likely to be advantageous in the detection of any short transposon-specific RNA. Full-length probe templates for Alu and B1 transposons were generated by PCR specific to the consensus sequences of these elements (see materials and methods). They also included RNA polymerase T7 and T3 promoter

sequence at the 5` and 3` of the PCR probe templates respectively (see materials and methods). Sense-specific (hybridising to the sense transcript) and antisense-specific (hybridising to the antisense transcript) probes were then generated by T7 or T3 *in vitro* transcription of the PCR product. Probes specific to both strands were generated because dependant on their primary transcript and RISC loading short RNA can be detected for single or both strands.

Because any transposon-associated short RNA may have similarity to miRNA or siRNA the ability to detect RNA of miRNA precursor and mature short RNA size was thought to be important. For this reason the blot from Figure 9 was stripped (method 1) with probe removal confirmed by phosphor-imaging, prior to hybridisation with the sense and antisense-specific full-length Alu probe, allowing us to test both formamide and glyoxal denatured RNA.

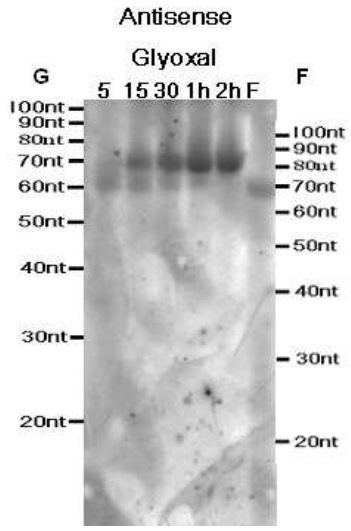


**Figure 10- Antisense specific Alu 70nt RNA Glyoxal Vs Formamide.** Total RNA was extracted from Hela cells, with ~10µg of RNA denatured by glyoxal incubation for 5 minutes (5), 15 minutes (15), 30 minutes (30), 1 hour (1h), 2 hours (2h) or denaturation with formamide (F). The RNA samples were separated by 10% denaturing-polyacrylamide gel by electrophoresis before being electro-blotted and cross-linked to nylon membrane using EDC. Hela RNA was then hybridised in 2X SSC and 1% SDS buffer to the full-length antisense specific Alu probe at 60℃, wash 0.2X SSC 0.2% SDS 60ºC. The radioactive signal was recorded by phosphor-imaging with a 1 hour exposure is shown with marker sizes annotated on the left of the image for glyoxal and the right for formamide.

The sense-specific full-length Alu probe did not hybridise any RNA of <100nt (not shown). However, the antisense-specific probe unexpectedly detected an RNA of ~70nt in both glyoxal and formamide denatured RNA. For the glyoxal denatured RNA a ~60nt RNA was observed to diminish with the increased length of glyoxal

incubation and appearance of the ~70nt RNA (Figure 10). This observation could be interpreted as the altered mobilisation of a single RNA from ~60nt to ~70nt, with the progressive glyoxal binding. The ~70nt antisense Alu RNA was also detected when this experiment was repeated using total RNA extracted from MCF-7 cells (data not shown).

The presence of B1 derived short RNA (~20-25nt) was also examined by gel blotting, using the mouse P19 embryonic carcinoma (EC) cell line. Total RNA was extracted from undifferentiated P19 EC cells. A dilution of the RNA was separated by agarose gel electrophoresis and stained with ethidium bromide to assess the quality of the RNA (data not shown). Equal amounts of RNA were then denatured using glyoxal or formamide, prior to separation by denaturing-polyacrylamide gel electrophoresis. The RNA was subsequently electro-blotted onto nylon membrane and cross-linked using EDC. Full-length sense and antisense-specific B1 probes were then hybridised to the RNA. Hybridisation of both formamide and glyoxal denatured P19EC RNA, with the full-length sense-specific B1 probe detected an RNA of ~90nt (Figure 11). However, differential hybridisation of the formamide and glyoxal denatured RNA was observed with the full-length antisense-specific B1 probe. This resulted in no RNA of <100nt being hybridised in the formamide denatured RNA, with a ~90nt and ~70nt RNA detected with glyoxal denaturation (Figure 11).

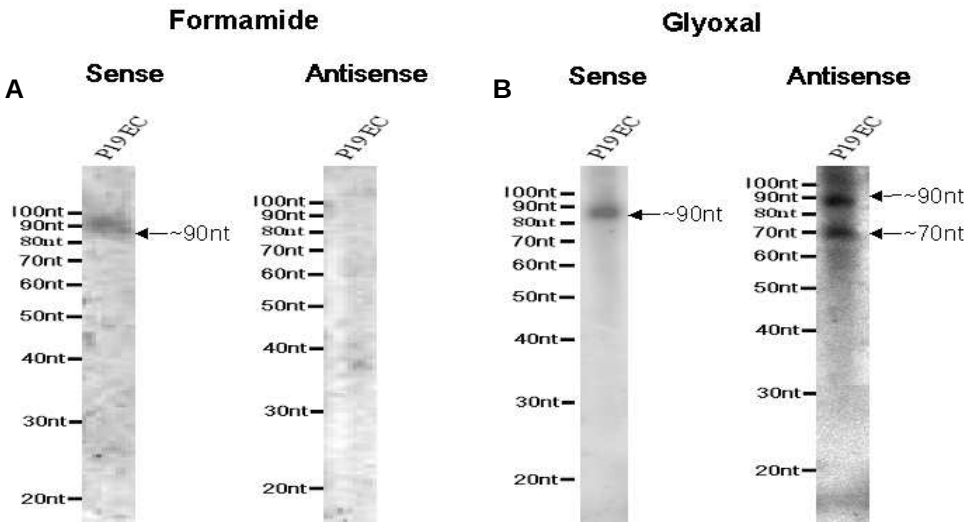


**Figure 11- Identification of a novel B1 like ~70nt antisense specific RNA.** RNA was extracted from mouse undifferentiated P19EC cells (P19EC). ~10µg of RNA was denatured using formamide (A) or glyoxal (B) and separated in adjacent lanes by 10% denaturing-polyacrylamide gel electrophoresis, before being electro-blotted and EDC cross-linked to nylon membrane using EDC. The membrane was then cut to separate the lanes of RNA, which were hybridised in 2X SSC and 1% SDS buffer to the full-length B1 sense-specific and antisense-specific probes at 50℃, wash 0.2X SSC 0.2% SDS 50ºC. Bands of ~70nt and ~90nt are indicated with arrows (right) with marker size annotated on the left of images.

The ~90nt sense RNA identified were assumed to be 4.5S RNA due to its size and shared sequence with the B1 element, however the ~90nt and ~70nt antisense RNA did not correspond to any known B1 like transcript. Because of the 7SL origins of Alu and B1 transposable elements, a 7SL specific probe was generated by PCR (2.13.4). Hybridisation of the human 7SL antisense-specific probe to MCF-7, Hela and P19 EC RNA did not identify a ~70nt RNA (data not shown), indicating the highly abundant 7SL RNA is not the source of the ~70nt RNA.

The apparent conservation of ~70nt antisense RNA transcripts between these related sequences of different species was of interest. This was because the size of these RNA was characteristic of miRNA precursor RNA and so may have indicated the potential for Alu and B1 short RNA from the processing of these transcripts. Therefore the ~70nt antisense-specific RNA were further investigated.

### *3.3.3 The Alu specific ~70nt can be observed with UV and EDC cross-linking methods.*

To confirm the novel Alu and B1 ~70nt antisense RNA detected were not an artefact of the EDC treatment, which was still at an early stage of development, the experiment was repeated using UV to cross-linked the RNA. The UV and EDC cross-linking methods can be compared in this situation because for RNA of this length the methods have similar detection efficiencies (personal communication G, Pall).

Total RNA extracted from MCF-7 cells was denatured using glyoxal with several adjacent lanes loaded with equal amounts of RNA and separated by denaturing-polyacrylamide gel electrophoresis. The RNA was electro-blotted onto nylon membrane and cut into strips, each containing one lane of MCF-7 RNA. These separate strips were cross-linked using standard UV or EDC (15 minutes and two hour incubations). The RNA was then hybridised to the full-length antisense-specific Alu probe. The ~70nt antisense RNA was detected at similar levels for both the UV and 15 minute EDC cross-linked RNA, with a slight increase in detection observed for the longer EDC cross-link of two hours. Therefore it was concluded that the ~70nt RNA was not an artefact of the EDC cross-linking method (Figure 12).

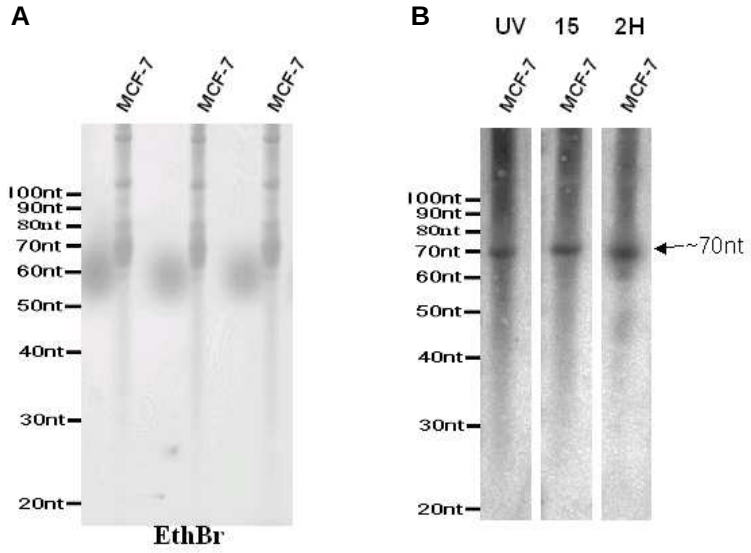


**Figure 12- Comparison of UV to EDC cross-linking for the visualisation of the ~70nt antisense specific Alu RNA.** RNA was extracted from human MCF-7 cells and denatured using glyoxal. Several lanes were loaded with 10µg of total RNA and separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A). The RNA was then electro-blotted onto nylon membrane, which was cut into strips to containing a single lane of MCF-7 RNA. The strips were cross-linked by standard UV (UV), EDC for 15 minutes (15) or EDC for two hours (2H). The RNA was then hybridised in 2X SSC and 1% SDS buffer to the full-length antisense-specific Alu probe at 40℃, wash 0.2X SSC 0 .2% SDS 40ºC. The radioactive signal was recorded by phosphor-imaging with a 1hour exposure is shown (B) with the ~70nt antisense RNA indicated with an arrow (right) and marker size annotated on the left of the image.

### *3.3.4 Determining the region of the Alu transcript specific to the ~70nt antisense RNA by DNA oligonucleotide mapping.*

Knowledge of where the ~70nt antisense Alu RNA maps in relation to the Alu transcript could indicate whether the RNA is derived from generalised or specific cleavage. In addition the identification of a specific region could provide secondary structure information on the ~70nt RNA.

Mapping of the ~70nt antisense Alu RNA was carried out using DNA oligonucleotides as probes. For this, 14 oligonucleotide probes (30nt in length with the exception of the first oligonucleotide) were designed to hybridise across the entire 251nt antisense Alu RNA consensus transcript. A 10nt overlap was introduced between adjacent oligonucleotides to reduce the effect of partial probe hybridisation on the hybridisation efficiency of the ~70nt antisense Alu RNA. Figure 13 indicates the position of hybridisation to the antisense transcript of each oligonucleotide probe. The exact sequence of probes used can be found in (Table 3A)

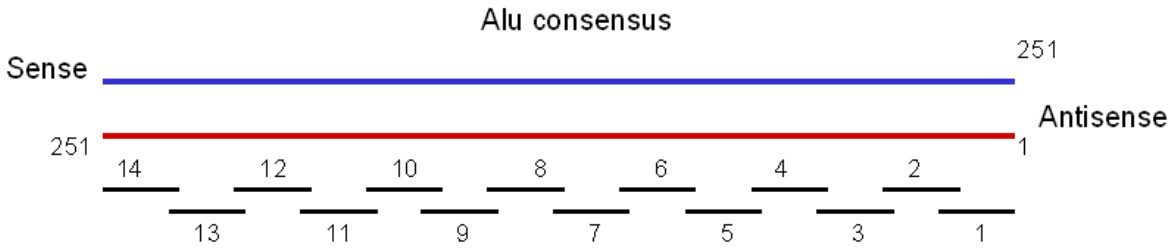


**Figure 13- Alu antisense-specific overlapping oligonucleotides for mapping.** The Alu consensus transcript is represented above with blue (sense) and red (antisense) strands. The oligonucleotide probes numbered 1-14, the black lines indicate their region of hybridisation relative to the antisense consensus sequence.

Equal amounts of total RNA extracted from MCF-7 cells, loaded into multiple adjacent lanes was separated by denaturing-polyacrylamide gel electrophoresis. The RNA was stained with ethidium bromide to demonstrate equal loading (Figure 14A) prior to electro-blotting and cross-linking to nylon membrane using EDC. The filter was then cut into four, each piece containing a single lane of MCF-7 RNA. The antisense-specific oligonucleotide probes were initially split into four pools 1-3, 4-7, 8-11 and 12-14 (see Figure 13 for positioning) and radioactively end labelled using $^{32}$P-γATP as described in the materials and methods. Hybridisation of the four oligonucleotide probes to MCF-7 RNA at 40°C

A B

did not detect the ~70nt antisense Alu RNA (data not shown). Subsequent hybridisation of the RNA at 30°C, found the ~70nt band to be detected with pool 4-7 (Figure 14B). Probe hybridised to the MCF-7 RNA was removed (method 1) with the effective removal of the probe confirmed by phosphor-imaging (Figure 14C). The MCF-7 RNA was then hybridised to the individual oligonucleotides 4,5,6 and 7. This identified the ~70nt antisense RNA to predominantly hybridise oligonucleotide probe 4, with a faint hybridisation of the ~70nt RNA with oligonucleotide probes 5 and 6 (Figure 14D).

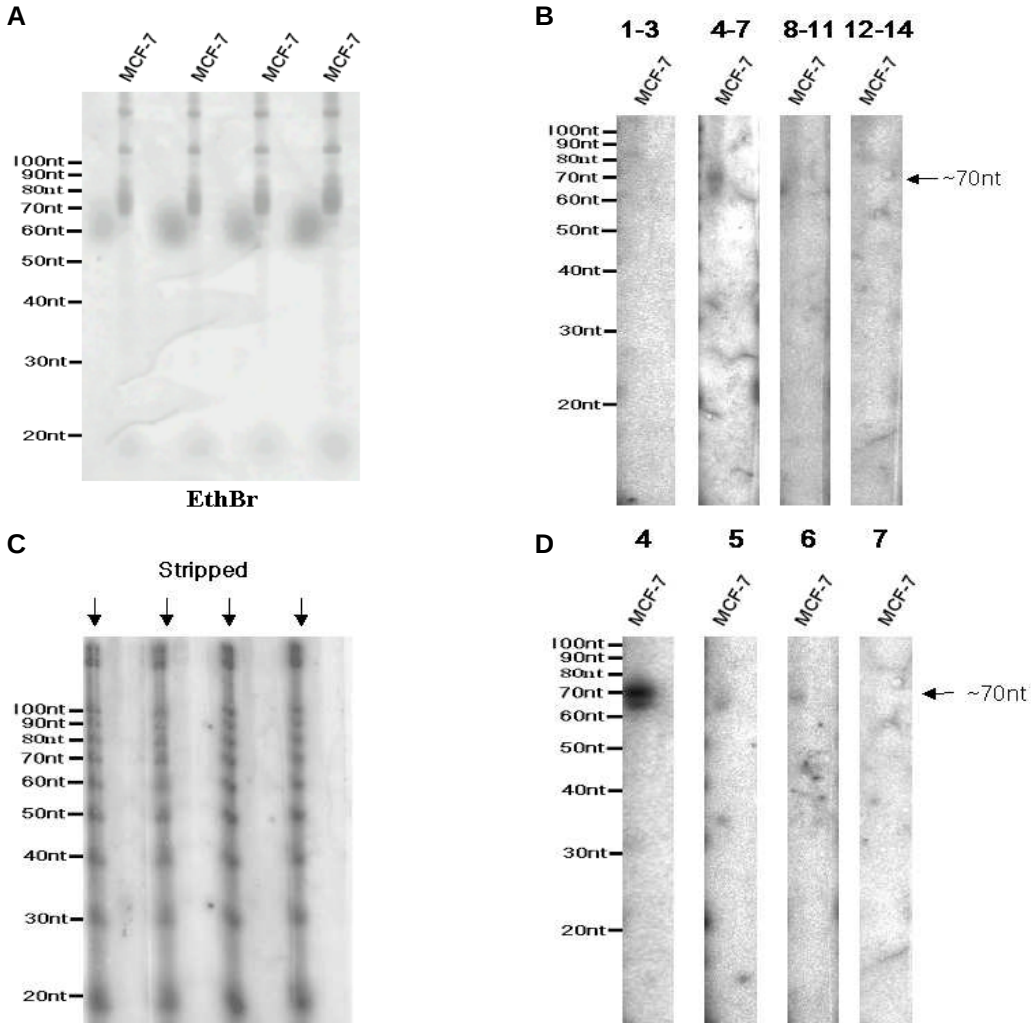


**Figure 14- Mapping of the ~70nt antisense Alu.** RNA was extracted from the human MCF-7 cells. Several lanes containing 10µg of MCF-7 RNA were separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide (A) prior to electro-blotting and cross-linking to nylon membrane using EDC. The RNA was then hybridised in 2X SSC 1%SDS buffer to the four (1-3, 4-7, 8-11 and 12-14) Alu antisense-specific oligonucleotide probe pools (B) and the individual oligonucleotide probes 4,5,6 and 7 (D) at 30℃, wash 0.2X SSC 0.2%SDS 30℃. The filter was stripped between hybrid isations with the phosphor-image shown (C). The radioactive signal was recorded by phosphor-imaging with all of the exposures being for 1 hour. RNA of ~70nt are indicated on the right with the marker on the left.

To assess the strength of hybridisation between oligonucleotide probe 4 and the ~70nt antisense Alu RNA, increased stringency wash steps were introduced. After the initial wash of 30°C and exposure the ~70nt band was easily observed (Figure 15A). The filter was then washed at 35°C, this showed a reduction in ~70nt RNA detected (Figure 15B). A final wash at 40°C subsequently removed the majority of the hybridised probe (Figure 15C). The weak hybridisation of oligonucleotide probes 5 and 6 was removed with a 35°C wash (data not shown).

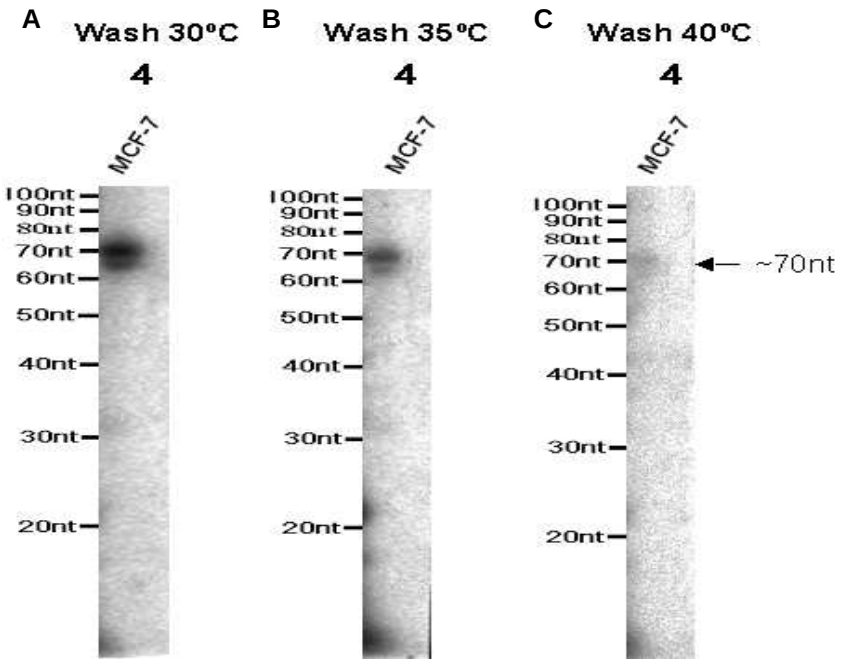


**Figure 15- Increased stringency washes of oligonucleotide 4 hybridisation.** The hybridisation of MCF-7 RNA to antisense-specific Alu oligonucleotide 4 probe shown in Figure 14D has been reproduced (A). Subsequent washing of this filter with wash 0.2X SSC 0.2%SDS at 35℃ (B) and 40℃ (C). The radioactive signal was recorded by phosph or-imaging with 1hour exposures shown above with the ~70nt band highlighted by an arrow on the right and the size marker indicated on the left of each image.

The weak hybridisation observed with the oligonucleotide probes may have been because of mismatch to the ~70nt antisense Alu RNA. A degenerate oligonucleotide 4 probe was created to assess if a higher stringency hybridisation could be achieved. The consensus sequences for the Alu young family members: Ya1, Ya4, Ya5, Ya8, Yb8 and Yb9 were used to generate the degenerate probe. These were selected as they represent the youngest and most transcriptionally active Alu elements. Figure 16 shows the alignment of each of the young family sequences. A single polymorphism within the region of oligonucleotide 4, at position 211nt of the sense-specific consensus from a guanosine (G) to an adenosine (A) was identified from this analysis.

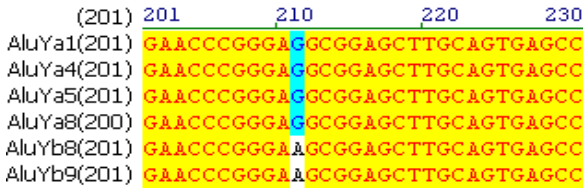


**Figure 16- Alignment of Alu Y consensus sequences.** Six consensus sequences obtained form Repbase (Jurka et al., 2005) representing the AluY a1, a4, a5, a9, b8 and b9 subfamilies were aligned. The sense specific Alu sequences corresponding to the oligonucleotide 4 probe are shown above with the numbering above the alignment indicating position relative to the sense specific consensus sequence. Regions of 100% identity are shown in yellow with polymorphic nucleotides not coloured.

The degenerate DNA oligonucleotide generated was expected to contain 50% guanosine and 50% adenosine at the sense-specific position 211nt. The degenerate probe was radioactively end labelled using $^{32}$P-γATP before hybridisation to the same MCF-7 RNA filter as in Figure 14. Like the original oligonucleotide probe 4, the ~70nt antisense Alu RNA was detected with 30˚C (Figure 17A) but not 40˚C hybridisation (data not shown). Increasing the stringency of washes for the 30˚C hybridisation also removed the ~70nt antisense RNA signal (Figure 17B). Therefore the polymorphic oligonucleotide probe does not appear more complementary to the ~70nt RNA than the original oligonucleotide 4 probe. A more inclusive degenerate could be made, however as there are no associated small RNA the precise sequence of the ~70nt RNA may not be the most relevant question.

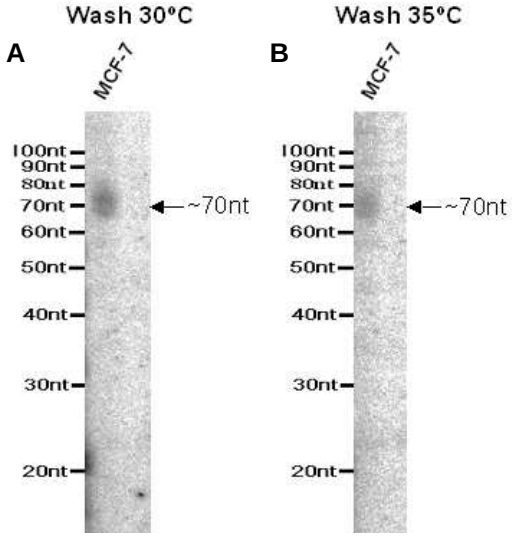


**Figure 17- Hybridisation of the polymorphic oligonucleotide 4 to MCF-7 RNA.** The MCF-7 RNA form Figure 14 was hybridised to the antisense-specific Alu degenerate oligonucleotide 4 probe at 30℃ with subsequent washes wash 0.2X SSC 0.2%SDS at 30℃ (A) and 35℃ (B). Radioactive signal was recorded by phosphor-imaging, 1hour exposures shown above. The ~70nt band highlighted by an arrow on the right and the size marker indicated on the left of each image.

### *3.3.5 SiRNA knockdown of RNAi components*

The known short RNA, miRNA and siRNA are both processed by overlapping components within the RNAi pathway. To assess whether the ~70nt antisense-specific Alu RNA may like a miRNA precursor be processed in association with the RNAi pathway, two RNase III enzymes Drosha and Dicer involved in miRNA processing were targeted using siRNA knockdown. Pre-designed Drosha and Dicer-specific siRNA were commercially obtained for this purpose (Table 1).

Hela cells were grown on six-well tissue culture plates prior to the separate transfection of Drosha and Dicer-specific siRNA. After seven days total RNA was extracted from the separate transfections. The RNA samples were then separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (Figure 18A). The ethidium stain showed the RNA to be equally loaded by comparison of the 5S rRNA bands. The RNA was then electro-blotted and cross-linked to nylon membrane using EDC.

To test whether the siRNA were effective at knocking-down Drosha and Dicer the RNA was hybridised to probes specific to hsa-mir-16 and hsa-mir-21. Hybridisation of the RNA to the hsa-mir-16 specific probe showed both Drosha specific siRNA transfections to have slightly reduced levels of the mature miRNA, compared with the eGFP and non-transfected controls. However, the hsa-mir-16 precursor estimated to be of 67nt was not detected, so could not be compared to assess knockdown. For the Dicer specific siRNA transfections, the Dicer 1 siRNA transfection showed an accumulation of higher bands. These included a ~70nt RNA of the expected precursor size, which is consistent with Dicer knockdown, however a coinciding reduction in the mature RNA was less clear. The Dicer 2 siRNA failed to cause the accumulation of precursor RNA or a reduction in mature miRNA levels. Hybridisation of the hsa-mir-21 specific probe more convincingly shows the reduction of mature miRNA for both Drosha specific siRNA transfections and the Dicer 1 siRNA transfection. However, no precursor RNA were observed for any of the samples. The hsa-mir-21 probe was removed (method 1) with removal confirmed by phosphor-imaging.

From the miRNA hybridisations it was concluded that Drosha 1, Drosha 2 and Dicer 1 siRNA had caused a detectable knockdown of their targets. Therefore the

RNA was then hybridised to the full-length antisense-specific Alu probe. This showed the ~70nt antisense RNA to be unaffected by the partial knockdown of Drosha and Dicer in Drosha 1, 2 and Dicer 1 samples, suggesting that the ~70nt antisense RNA is neither a processing product of Drosha nor a substrate of Dicer.

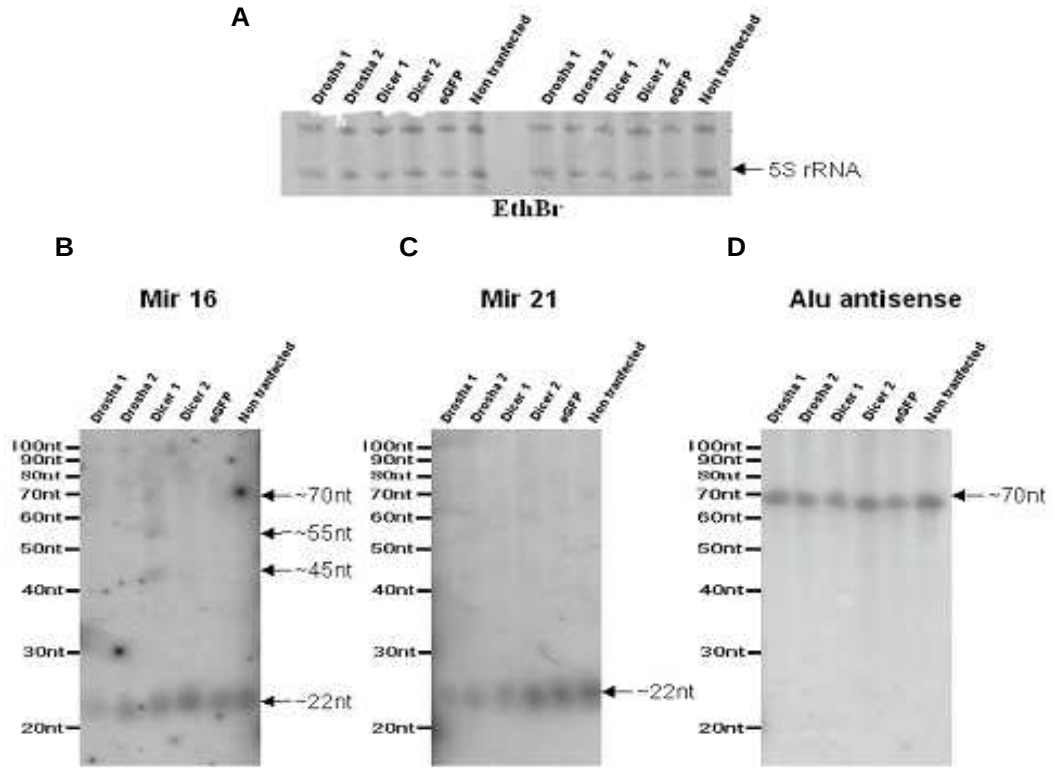


**Figure 18- The siRNA targeted knockdown of Drosha and Dicer in Hela cells.** Hela cells were grown on six-well plates and transfected with separate siRNA targeted to Drosha (Drosha 1 and Drosha 2) Dicer (Dicer 1 and Dicer 2) and green fluorescence protein (eGFP) as a negative control. The various transfections were incubated for seven days with refreshment of the media every second day with a non-transfected sample carried through in parallel. After seven days total RNA was harvested from each of the transfections and separated by 10% denaturing-polyacrylamide gel electrophoresis. The RNA was stained with ethidium bromide (A) prior to electro-blotting and cross-linking to nylon membrane using EDC. The RNA was then hybridised in 2XSSC and 1% SDS buffer to the hsa-mir-16 (B), hsa-mir-21 (C) probes at 40℃ and the full-length antisense-specific probe at 50℃ (D) washes with wash 0.2X SSC 0.2%SDS at 40℃ and 50℃ respectively. The radioactive signal was recorded by phosphor-imaging with 16 hour exposures are shown for all.

## 3.4 Discussion

### *3.4.1 SINE Alu and B1 antisense ~70nt RNA identified in cancer cell lines.*

The initial objective of the analysis described in this chapter was to identify any Alu or B1 specific short (20-25nt) RNA in the cancer cell lines examined. Despite no short RNA being identified, a ~70nt antisense RNA was serendipitously observed for both Alu and B1 (Figure 10 and Figure 11). These were previously unreported in the literature and their identification using UV cross-linked RNA (Figure 12) confirmed them not to be an artefact of the EDC cross-linking method. The ~70nt antisense transposon-specific RNA were of particular interest because of their similarity in size to known miRNA precursors processed by the Drosha-DGCR8 complex (Han et al., 2006). In addition the antisense specificity and size similarity of two RNA derived from evolutionarily-related transposable elements (Vassetzky et al., 2003) could indicate a conserved function of the RNA.

Antisense Alu and B1 transcripts are known to be highly abundant in human and mouse respectively (Lander et al., 2001, Waterston et al., 2002). However, the internal RNA polymerase III promoters of Alu and B1 have not been implicated in the generation of antisense transcripts. Despite this, global transciptome analysis suggests that a large proportion of the mammalian genome (including transposon sequences) can produce both sense and antisense specific transcripts from read-through transcription of transcriptional units (Katayama et al., 2005). In addition, Alu elements are reported to exist in the introns and UTRs of ~75% of human protein encoding RNA (Kim et al., 2004). Therefore during RNA polymerase II transcription of mRNA there is significant potential for the generation of antisense Alu transcripts.

Mapping of the ~70nt antisense Alu RNA partly defined a region of the full-length transcript generating the RNA. This was by the hybridisation of a single 30nt DNA oligonucleotide probe at low stringency (Figure 14). The region of hybridisation was located to a region with no conservation to the human 7SL RNA sequence, near the 5` of the antisense Alu transcript. This supports the result obtained with the 7SL specific probe hybridisation (data not shown), indicating the ~70nt

RNA not to be derived from the processing of 7SL RNA. However, the low stringency of the oligonucleotide probe 4 hybridisation indicated there to be some mismatch to the ~70nt Alu RNA (Figure 15). In an aim to generate a more complementary probe, a polymorphic nucleotide present in the Alu Y sub family (Jurka et al., 2005) was introduced to create a degenerate probe. However, no improvement in the hybridisation efficiency was observed (Figure 17). Therefore there appears to be an unknown source of mismatch between the DNA oligonucleotide probe sequences used and the ~70nt RNA sequence. The inability to observe the ~70nt RNA with the adjacent DNA oligonucleotide probes may also indicate more extensive mismatch across the ~70nt sequence. The full-length PCR generated probe was however able to compensate for any polymorphisms from the consensus sequence, hybridised at very high stringency to the ~70nt RNA.

Alternatively, the weak hybridisation of the oligonucleotide probes may be caused by the post-transcriptional ADAR modification of the sequence. This hypothesis is supported by reports that RNA polymerase II derived Alu transcripts (sense and antisense) comprise for the majority of targets of human adenosine deaminases that act on RNA (ADARs) modifications (Kawahara and Nishikura, 2006, Kim et al., 2004, Levanon et al., 2004). Post-transcriptional modification by ADARs as discussed in the introduction (2.2.6) recognise dsRNA and deaminate adenosine residues into inosine. Inosine is then recognised as guanosine resulting in adenosine to guanosine conversion. The alteration of Alu transcripts by these modifications could result in extensive change of the sequence from the consensus. Thus could provide an explanation for the mismatch observed between the consensus designed DNA oligonucleotide probes and the ~70nt RNA (Figure 15 and Figure 17). The PCR derived full-length antisense Alu probe may be able to tolerate this mismatch because of the increased length of hybridising region.

To assess if the ~70nt antisense Alu RNA were processed like miRNA precursors, Drosha and Dicer RNase III enzymes were knocked down using siRNA in Hela cells. The extent of knockdown achieved was then assessed by comparing the accumulation of mature and precursor miRNA between samples. This method of assessment was selected because it indicated the actual reduction of enzyme activity rather than the percentage of protein or mRNA knockdown as detected

by western and RT-PCR analysis respectively. From this analysis some Drosha and to a lesser extent Dicer knockdown was observed to effect miRNA accumulation, however there was no change in the accumulation of the ~70nt RNA (Figure 18). Thus indicating Drosha and Dicer not to be involved in the processing of the ~70nt antisense Alu RNA.

An alternative to Drosha processing of the ~70nt RNA may also involve the deamination of adenosine residues to inosine. Inosine modified dsRNA are reported to be cleaved by Tudor-SN (Scadden, 2005), which associates with the RISC complex. The efficiency of Tudor-SN cleavage increases proportionally with the number of consecutive I:U pairs in a sequence. Within the 5` of the antisense Alu transcript there are no optimal (three consecutive I:U pairs) cleavage sites. Therefore relatively long cleavage products (~70nt) of ADAR modified antisense Alu transcripts may be produced. The ADAR and RNAi pathways have been reported to be antagonistic to each other with ADAR modified dsRNA are not good substrates for Dicer. Therefore ADAR modification could also explain the absence of any short (~20-25nt) Alu RNA despite the confirmation of extensive Alu dsRNA within the cell (Kawahara and Nishikura, 2006).

Attempts were made to establish if the ~70nt antisense Alu RNA contained inosine modifications using a published method (Morse and Bass, 1997), however no data was obtained. Future work on the ~70nt antisense Alu RNA would focus on establishing the direct sequence of the RNA and any post-transcriptional ADAR modification. The identification of inosine modification to the Alu RNA in addition to explaining some of the results obtained in this chapter could implicate the ~70nt RNA in an antagonistic pathway to the RNAi pathway. This pathway has been implicated in heterochromatin formation (Wang et al., 2005) and could provide an alternative to the RNAi suppression of Alu.

# 4

# Investigation of murine organ and differentiated embryonic cell line RNA for the detection of transposon-associated siRNA

## Results 2

## 4.1 Summary

This chapter focuses on extending the investigation of transposon-associated short RNA (~20-25nt) in mouse to include SINE (B1, B2), LINE (L1) and LTR type-elements (IAP and MuERV). This involved examining RNA extracted from adult mouse tissues (testis, lung and kidney) and embryonic cell lines (P19EC, male ES and female ES) in undifferentiated and differentiated states, by gel blotting.

Using this method to investigate the mouse L1 transposon promoter regions revealed a novel $T_F$-type promoter short antisense RNA of ~19nt. The ~19nt RNA was specifically observed in differentiated embryonic cell RNA (ES and P19EC). Coincidently a ~19nt band was also frequently observed by ethidium staining of the RNA prior to transfer in these samples. However, no short RNA relating to the A-type and $G_F$-type promoter sequences were observed.

RNA extracted from mouse ES and P19 EC cells after two, four, six and ten days of differentiation were then examined for short RNA specific to L1, B1, IAP and MuERV-L transposons. These results indicated the induction of short RNA with several intermediate bands after two-days of differentiation for the $T_F$-type promoter. However, despite longer RNA (100-40nt) and indications of short RNA bands being detected, defined short RNAs for the other transposons investigated were not observed.

# 4.2 Introduction

## *4.2.1 The Long Interspaced Element LINE-1.*

Long Interspaced Element (LINE) is the term used to describe non-long terminal repeat sequences >500nt in length, propagated by retrotransposition within the genome. LINE-1 (L1) is the most abundant of these and is the only reported autonomously active LINE retrotransposon in mammals. L1 is estimated to account for 10-18% (Waterston et al., 2002) and 15-17% of the mouse and human genomes respectively (Lander et al., 2001, Smit, 1996).

The functional full-length L1 transposon is ~6-7kb in size and is comprised of the 5´ UTR, two open reading frames (ORF1 and ORF2) and the 3´UTR terminating in a polyadenylated sequence (Figure 19). The 5´ UTR containing an internal RNA polymerase II promoter, ORF1 encoding a non-specific RNA binding protein and ORF2 encoding a reverse transcriptase were all found to be essential for transposition by targeted point mutation (Moran et al., 1996). The majority of L1 transposons are truncated at the 5´. This is thought to be the result of incomplete reverse transcription during retrotransposition (Voliva et al., 1983). Many other L1 elements are inactive through various internal rearrangements, resulting in few elements capable of retrotransposition. It is estimated that a greater proportion of murine L1 transposons are active than their human counterparts (DeBerardinis et al., 1998).


**Figure 19- The functional full-length L1 transposon structure.** Diagram of the essential regions of the L1 transposon. From left to right the 5´ UTR containing the RNA polymerase II recognition site, ORF 1 encoding a non-specific RNA binding protein, ORF2 encoding the reverse transcriptase and the 3´ UTR terminating in a polyadenylated sequence are shown.

Despite the cross-species similarities, mouse L1 transposons differ considerably from human L1 copies in their 5´ UTR. The mouse L1 transposon is characterised by the presence of a variable number of repeats (~200nt) termed monomers in the 5' UTR region. It has been shown that these monomers possess promoter activity and that increasing their number increases the level of transcription

(Adey et al., 1994, DeBerardinis and Kazazian, 1999, Severynse et al., 1992). There are five families of mouse L1 transposons currently known, defined by their monomer sequences (V, F, A, $T_F$ and $G_F$). The different monomers are thought to have arisen from the occasional capture of sequences with promoter activity and or recombination with existing L1 monomers. The creation of a new monomer promoter allows the expression and expansion of the new progenitor elements (Adey et al., 1994, DeBerardinis and Kazazian, 1999, Goodier et al., 2001, Loeb et al., 1986, Mears and Hutchison, 2001).

Of these monomer families three are recognised as active; $T_F$, $G_F$ and A whereas V and F subfamilies are considered extinct through the accumulation of mutations resulting in their inability to transpose (Adey et al., 1994, Adey et al., 1994, Goodier et al., 2001, Mears and Hutchison, 2001). The $T_F$ and $G_F$ subfamilies are the youngest families and are currently expanding with an estimated ~3000 active $T_F$ (DeBerardinis et al., 1998) and ~400 active $G_F$ members (Goodier et al., 2001). In addition to this a small proportion of A-type L1 transposons are also active (DeBerardinis et al., 1998). All three active families can be evolutionary linked to the extinct F-type L1 family (Mears and Hutchison, 2001) although sequence alignment of the monomers shows there to be significant divergence.

The L1 transposon is highly expressed along with other transposable elements during embryonic genome activation (Evsikov et al., 2004). This early expression of L1 transposons and the encoded reverse transcriptase has been proposed to be essential to the pre-implantation embryo (Beraldi et al., 2006).

### *4.2.2 The long terminal repeat (LTR) elements, intracisternal A-particle (IAP) and murine endogenous retrovirus-L (MuERV-L).*

Long terminal repeats (LTR) elements are a third class of retrotransposons accounting for ~10% of the mouse genome (Waterston et al., 2002). LTR transposons are typified by 5´ and 3´ LTR sequence (~300-600nt), which provide the promoter activity (Christy and Huang, 1988). The LTRs flank protein-coding sequence which resemble a retroviral provirus, often including *gag* (group-specific-antigen genes) *pol* (reverse transcriptase) and in some instances *env*

(envelope) genes. Due to their structure, recently transposed LTRs (LTRs with high identity) are susceptible to internal homologous recombination. This results in deletion of most of the sequence, leaving solitary LTR sequences in the genome. Therefore LTR transposon families can be defined by >90% identity of reverse transcriptase genes for autonomous or >60% LTR similarity for non-autonomous elements (McCarthy and McDonald, 2004).

IAP is an autonomous LTR transposon that is currently active in the mouse genome. Its transposition rate is reported to be similar to that of the L1 transposon as defined by mobilisation assay (Dewannieux et al., 2004). The full-length IAP element encodes *gag* and *pol* genes in addition to a protease (*prt*) gene (Figure 20). All were found to be essential for autonomous transposition by mutational analysis (Dewannieux et al., 2004). There are ~1000 copies of the IAP LTR in the mouse genome. Internal deletions of these elements cause their size vary from ~3000-7000nt, resulting in few containing the ORFs required for transposition. There are thought to be up to 300 IAP elements currently active in the mouse genome (Dewannieux et al., 2004).

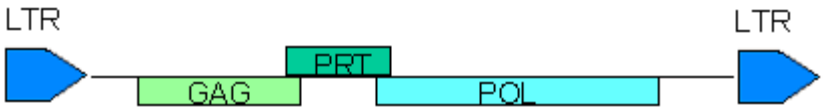


**Figure 20- The functional full-length IAP transposon structure.** The diagram indicates the essential regions of the IAP LTR transposon. Two LTR sequences flanking *gag*, *prt* and *pol* genes required for transposition. The arrowhead of the LTR regions indicates the orientation of the promoter.

MuERV-L is a second autonomous LTR retrotransposon thought to be currently active in the mouse genome due to the high identity between the 5´and 3´ LTR's (Costas, 2003). There are <200 copies of the MuERV-L transposon in mouse. Of theses ~20 full-length copies are predicted to be capable of retrotransposition (Costas, 2003). The full-length MuERV-L transposon encodes *gag* and *pol* genes (containing a small deletion) with no *env* region (Figure 21). MuERV-L retrotransposon is related to the ERV-L family of retrotransposons present in all placental mammals and is >70% identical to its inactive human equivalent HERV-L.

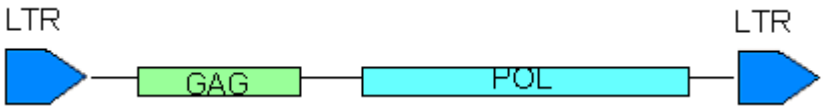


**Figure 21- The functional full-length MuERV-L transposon structure.** The diagram indicates the essential regions of the MuERV-L LTR transposon. Two LTR sequences flanking *gag* and *pol* genes required for transposition. The arrowhead of the LTR regions indicates the orientation of the promoter.

Both IAP and MuERV-L are highly expressed in mouse early embryo (Kigami et al., 2003, Svoboda et al., 2004) prior to transcriptional suppression (Ma et al., 2001, Walsh et al., 1998). Knockdown of the RNase III enzyme Dicer by siRNA injection of the one-day embryo has been reported to result in a 50% increase in both IAP and MuERV-L transposon mRNA as analysed by RT-PCR (Svoboda et al., 2004). In addition to this sense and antisense transcripts have been detected for both elements in the pre-implantation embryo (Svoboda et al., 2004).

## 4.3 Results

### *4.3.1 Short antisense-specific L1 $T_F$ monomer RNA identified in differentiated embryonic stem cells.*

The analysis of cancer cell line RNA did not identify any short RNA of 20-26nt with homology to B1 mouse or Alu human transposable elements (Chapter 3). However, this investigation was narrow, examining a single transposon type, in few cell lines. To expand the investigation of transposon-associated short RNA in mammals, mouse transposable elements were focused on. This was because mouse genomes have more active transposon populations than humans (Waterston et al., 2002), which may increase the detection efficiency of any short transposon-associated RNA. Also a wider range of tissue and embryonic cell line RNA was available for analysis.

The high abundance and transcriptional activity of transposable elements were deemed to be important criteria in the identification of short transposon-associated RNA. This was because transposon-associated short RNA can only be generated from transcribed transposon sequences (specific or read-through). Thus, the higher the transcriptional activity and transposon abundance, the greater the specific and read-through transcription potential respectively, increasing the ability to detect any transposon-associated short RNA generated. In addition to facilitating RNA identification, the more abundant transposons would have a greater significance when investigating the general RNAi mediated transposon-silencing hypothesis. The L1 element was therefore the first mouse transposon investigated for the presence of transposon-associated short RNA, because it is the most abundant transposon in the mouse genome (Waterston et al., 2002).

It was not possible to synthesise a full-length L1 for use as a probe template due to the inability to PCR a single band of the correct size by PCR with the primer sequences used. Instead the tandemly arranged monomer subunits in the 5´ UTR were analysed for the presence of short RNA. The 5´ UTR sequences were targeted because they contain the promoter function of the L1 elements, which may be targeted by short RNA to impose TGS. Probe templates representing the

consensus sequence (Jurka et al., 2005) for the three active L1 subgroups A, $G_F$ and $T_F$ were generated by PCR.

To assess the presence of L1 monomer derived short RNA (~20-25nt) in mouse, total RNA was extracted from various embryonic cell lines (male ES, female ES and P19EC) and adult mouse tissues (testis, lung and kidney). The male and female ES cells were analysed in both undifferentiated and differentiated states. Differentiation of the ES cells was induced by the removal of leukaemia inhibitory factor (LIF). A dilution of each of the RNA samples was then separated by agarose gel electrophoresis and stained with ethidium bromide. This was done to assess the quality and estimate the concentration of the RNA samples (data not shown). Quantification by light spectrophotometry was avoided because of the large quantity of the source RNA required to obtain meaningful values. Also light spectrophotometry provides no information on RNA integrity, which is critical to the analysis of short RNA. Despite analysis by ethidium staining providing a less accurate quantification of samples, this was used because the integrity of the RNA could be assessed with little source RNA required for quantification.

Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (Figure 22A). Staining with ethidium bromide allowed the equal loading of the samples to be assessed retrospectively by comparison of the 5S rRNA band. This showed not all samples to be equally loaded. The differentiated male and female ES samples had a more smeared appearance and had a distinct ~19nt band. The RNA was subsequently electro-blotted on to nylon membrane before cross-linking using EDC. Probes specific to the sense and antisense strands of the consensus sequences for A-type, $G_F$-type and $T_F$-type monomers were then hybridised to the RNA. Probe removal was carried out between hybridisations (method 1) and the removal verified by phosphor-imaging. The A-type monomer probes revealed a novel sense-specific RNA of ~65nt (Figure 22B) and an antisense-specific RNA of ~70nt (Figure 22C) to be present in all samples. Neither of these RNA corresponded to any known L1 like transcript. However, no RNA of ~20-25nt were identified.

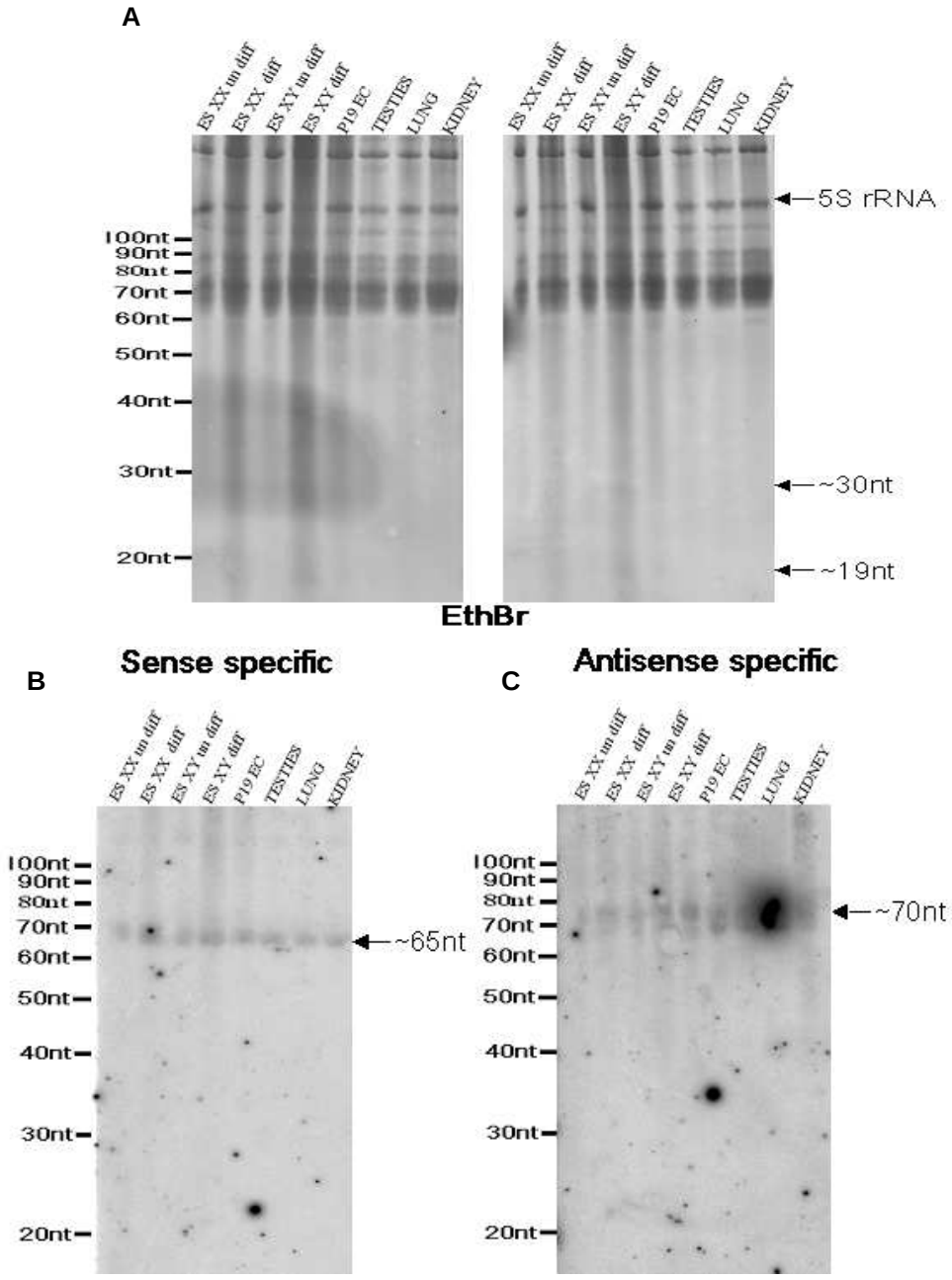


**Figure 22- Hybridisation of A-type L1 monomer to a range of mouse cell lines and adult tissue RNA.** RNA was extracted from female undifferentiated ES (ES XX un diff), female differentiated ES (ES XX diff), male undifferentiated ES (ES XY un diff), male differentiated ES (ES XY diff), P19 EC (P19 EC) cell lines and adult male mouse testis (Testis), lung (Lung) and kidney (Kidney) organs. ~20µg of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The various RNA samples were then hybridised in 2X SSC and 1% SDS buffer to the A-Type L1 monomer sense (B) and antisense-specific (C) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was r ecorded by phosphor-imaging with 16 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

Subsequent hybridisation of the $G_F$-type monomer sense-specific probe revealed a defined RNA band just above the 100nt marker in all samples (Figure 23A). However, this band was obscured by a smear in the differentiated female and

male ES samples. Given the appearance of the ethidium stained RNA, this was not an unexpected observation. The antisense-specific $G_F$-type monomer probe hybridisation identified an RNA doublet of ~70nt in all samples with the exception of undifferentiated and differentiated male ES samples (Figure 23B). The male undifferentiated and differentiated samples instead displayed a RNA of ~80nt. As the male and female cell lines have different origins this observation may be as a result of polymorphism. Irrespective of the cause the observation of this polymorphic band at ~80nt, does argue against cross-hybridisation of the tRNA. This is because the purifying selection imposed upon RNA with function, like the tRNA would act against polymorphism formation, whereas the transposon RNA with no cellular function would likely not be under purifying selection and therefore be more likely to have polymorphisms between different strains.

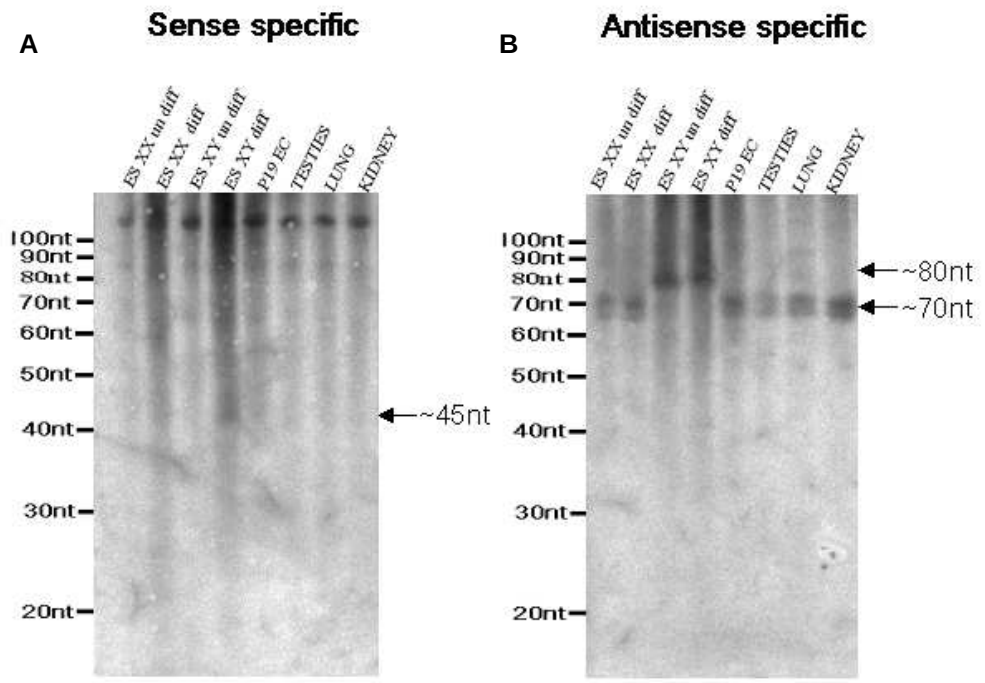


**Figure 23- Hybridisation of $G_F$-type L1 monomer to a range of mouse cell lines and adult tissue RNA.** The RNA samples from Figure 22 had probe removed (method 1) before hybridisation to the $G_F$-Type L1 monomer sense (A) and antisense-specific (B) probes in 2X SSC and 1% SDS buffer at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The ra dioactive signal was recorded by phosphor-imaging with 1 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

Hybridisation of the RNA with the $T_F$-type monomer probes identified a ~70nt sense-specific RNA present in all samples, although this appears smeared and possibly diminished in the differentiated male and female ES samples (Figure 24A). The antisense-specific $T_F$-type monomer probe identified several bands,

with RNA of ~70nt and ~80nt in the majority of samples (Figure 24B). In addition to these, discrete RNA of ~35nt (lung sample) and ~19nt (differentiated male and female ES samples) were also detected. The antisense-specific RNA of ~19nt was of particular interest because it was within the size range of known functional short RNA. The ~19nt RNA was observed to corresponded to an ethidium stained band observed in Figure 22A which is reminiscent of the piRNA class of short RNA which can be observed as a ~30nt ethidium stained band (Aravin et al., 2006, Girard et al., 2006). The sense and antisense-specific RNA of ~70nt being of similar in size to known miRNA precursors were also of interest. The observed reduced antisense-specific ~70nt RNA in the differentiated ES samples supports the idea that a similar processing pathway may be involved in the generation of the ~19nt RNA.

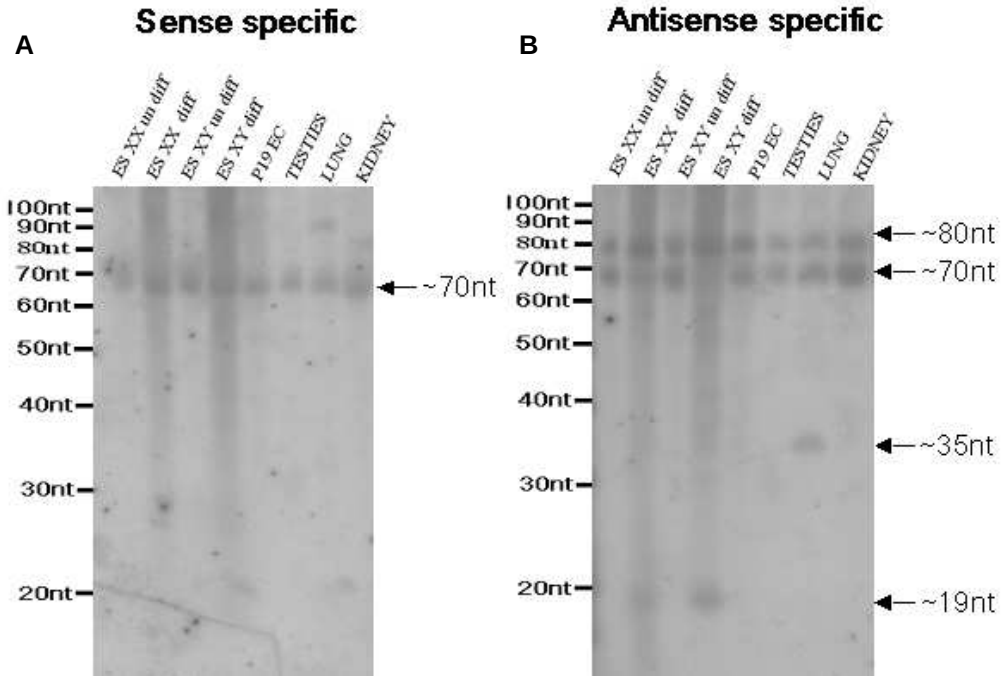


**Figure 24- Hybridisation of $T_F$-type L1 monomer to a range of mouse cell lines and adult tissue RNA.** The RNA samples from Figure 22 had probe removed before hybridisation to the $T_F$-Type L1 monomer sense-specific (A) and antisense-specific (B) probes in 2X SSC and 1% SDS buffer at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The ra dioactive signal was recorded by phosphor-imaging with 16 hour exposures shown for both hybridisation. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

Therefore RNA with the characteristic size of short interfering RNA, were only observed with the $T_F$-type antisense monomer probe and only in differentiated ES cell RNA. However, other RNA of characteristic size to miRNA precursors were observed with sense and antisense-specific probes to the A-type and $T_F$-type L1 monomers. None of these RNA have been previously described.

### *4.3.2 Non-discrete short antisense RNA identified in differentiated embryonic stem cells for B1 transposon.*

Male and female differentiated ES cells were both observed to accumulate at least one type of short transposon-associated RNA. Since a ~19nt band could be observed with ethidium bromide staining it was thought that this may contain other transposon-associated short RNA, processed in parallel to the L1 short RNA identified.

To investigate this the L1 $T_F$-type probes were removed (method 1) and the removal verified by phosphor-imaging. The RNA was then hybridised to the B1 full-length sense and antisense-specific probes. The sense-specific probe detected an RNA of ~90nt in all samples (likely to be 4.5S RNA), although this was less clear in the differentiated samples (Figure 25A). Hybridisation to the antisense-specific probe detected the previously identified ~90nt RNA and ~70nt RNA in all samples (Figure 25B). Again smearing was observed for the differentiate ES samples, with an indication of lower RNA bands of ~45nt, ~30nt and ~22nt present within the smear. However, none of these bands was of the discrete nature expected for functional short RNA.

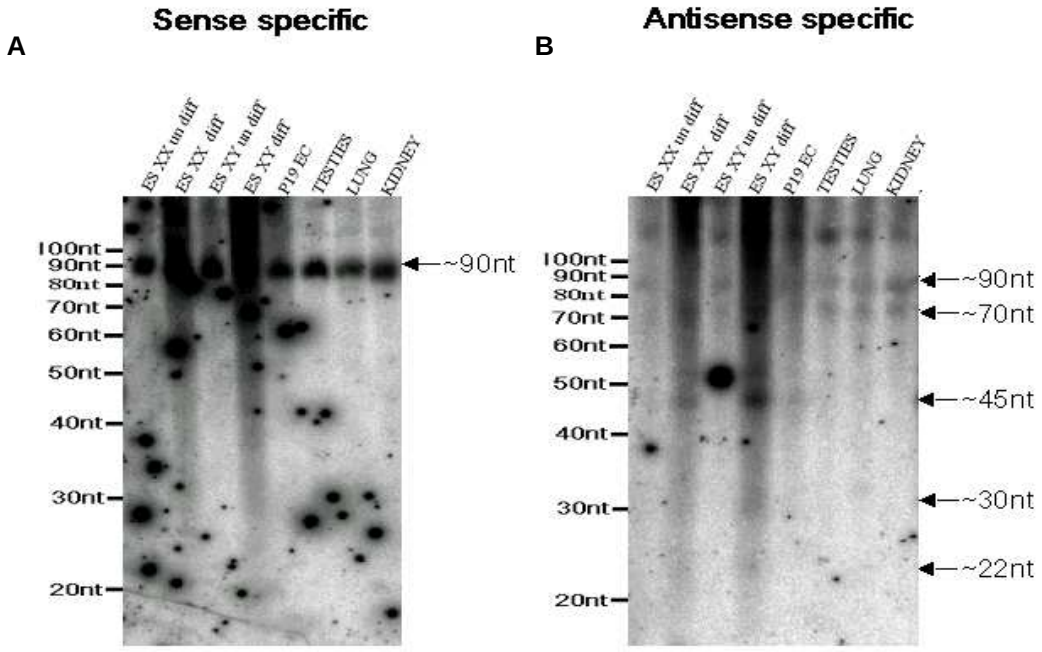


**Figure 25- Hybridisation of the full-length B1 probe to a range of mouse cell lines and adult tissue RNA.** The stripped RNA samples from Figure 22 were hybridisation in 2X SSC and 1% SDS buffer to the full-length B1 sense-specific (A) and antisense-specific (B) probes at 50℃, wash 0.2X SSC 0.2% SDS 50℃. The radioactive signal was recorde d by phosphor-imaging with 16 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

### 4.3.3 The $T_F$-type L1 monomer associated ~19nt RNA is identified in P19 EC differentiated cells.

The $T_F$-type ~19nt antisense RNA was identified in two differentiated ES cell lines. To assess if the accumulation of the short RNA was specific to differentiated ES cells or general to differentiated embryonic cell lines the P19 EC cell line was analysed. The P19 EC cell line was selected for this analysis, because it was another well-characterised cell line of embryonic origins and was available in the laboratory.

Mouse P19 EC cells were grown as standard and differentiation was induced by the addition of retinoic acid. Undifferentiated cells were grown in parallel without the addition of retinoic acid for comparison. Total RNA was extracted from undifferentiated and two-day differentiated P19 EC cells and concentration measured by $UV^{260}$ light absorbance. Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (Figure 26A). This showed the differentiated RNA sample to appear smeared in comparison to the undifferentiated RNA, with discrete ~19nt and ~35nt bands clearly visible.

The RNA was hybridised to $T_F$-type monomer sense and antisense-specific probes. This identified a sense-specific RNA of ~70nt in both samples (Figure 26B). Hybridisation to the antisense-specific probe detected several RNA, with RNA of ~70nt and ~80nt observed in both undifferentiated and differentiated samples. In addition, RNA of ~19nt and ~37nt were observed specifically in the differentiated P19EC sample (Figure 26C). Again a correlation was observed between the ethidium stained RNA bands and the expression of short RNA specific to the L1 $T_F$-type monomer.

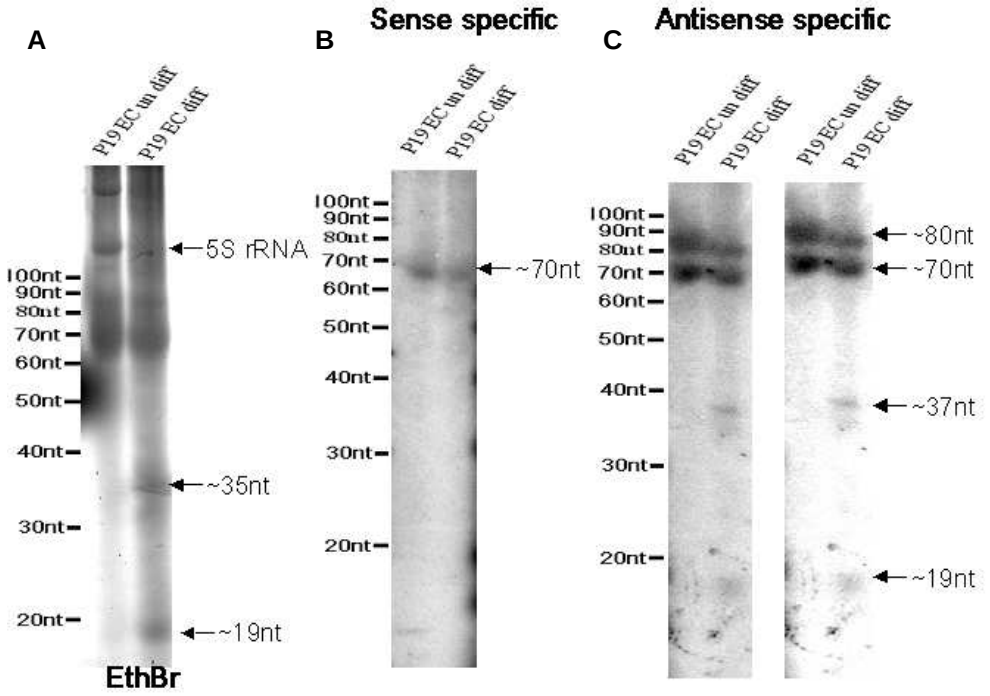


**Figure 26- Hybridisation of $T_F$-type L1 monomer to undifferentiated and two-day differentiated P19 EC RNA.** RNA was extracted from undifferentiated (P19 EC un diff) and differentiated (P19 EC diff) P19 EC cell line and quantified by $UV^{260}$ light absorbance. 10µg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 1% SDS buffer to the $T_F$-Type L1 monomer sense (B) and antisense-specific (C) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor- imaging with 96 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images. An over exposed image of the antisense-specific hybridisation is also shown to the right of the original (C).

### *4.3.4 A time course of embryonic cell differentiation reveals the induction of $T_F$-type monomer L1short RNA at two-days.*

The novel ~19nt $T_F$-type L1 antisense RNA was not observed in undifferentiated embryonic cells or the adult tissue screened. Therefore embryonic cell differentiation was focused on for further investigation of the induction of this transposon-associated short RNA.

Female mouse ES cells were grown as standard with differentiation induced by the removal of LIF from the culturing medium. Undifferentiated cells were grown in parallel without the addition of LIF for comparison. Total RNA was extracted from undifferentiated, two, four, six and ten-day differentiated ES cells and measured by $UV^{260}$ light absorbance. Equal amounts of the RNA samples

were then separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide. This showed the samples to be equally loaded with the two-day differentiated RNA having a slightly smeared appearance in comparison to the other samples (Figure 27A). The RNA was then electro-blotted onto nylon membrane and cross-linked using EDC.

The RNA was hybridised to the sense and antisense-specific probes of the L1 $T_F$-type monomer. The probes were removed between hybridisations (method 1) and the removal confirmed by phosphor-imaging. The sense-specific hybridisation identified a single sense RNA of ~70nt present in all samples as previously observed (Figure 27B). The antisense-specific hybridisation again identified several RNA bands, with the ~70nt and ~80nt RNA detected in all samples. In addition to these, RNA of ~38nt and ~32nt were also observed, specifically in the two-day differentiated ES sample (Figure 27C). However, the short RNA of ~19nt,previously detected were not easily identifiable, with only a faint signal observed.

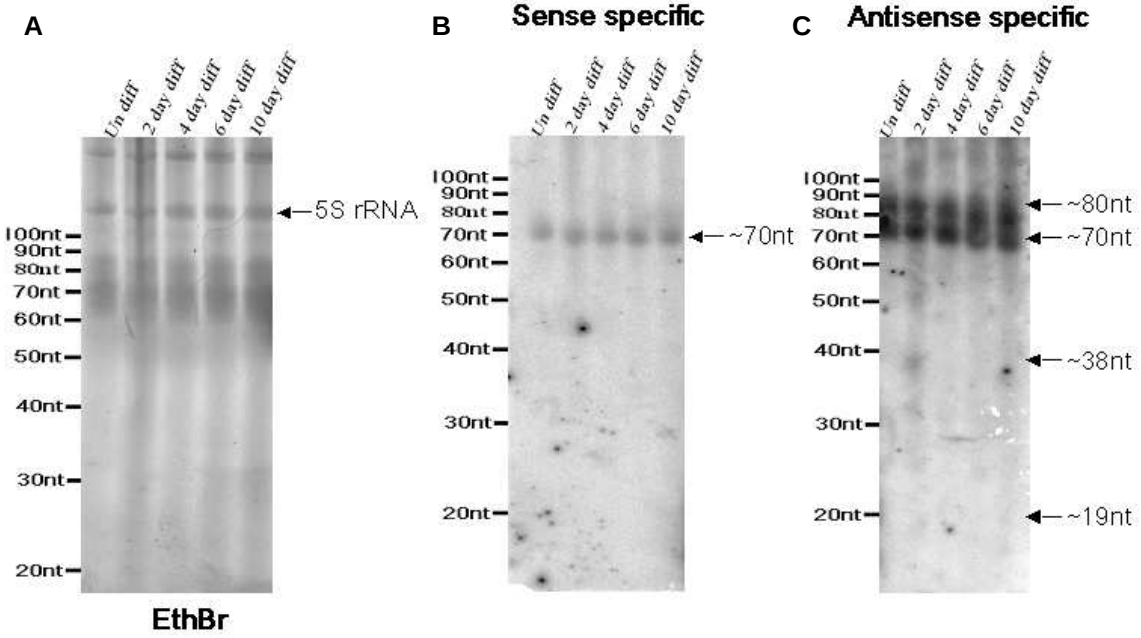


**Figure 27- Hybridisation of the $T_F$ L1 monomer probes to ES cell differentiation time course RNA.** RNA was extracted from undifferentiated (un diff) two-day (2day diff), four-day (4 day diff), six-day (6 day diff) and ten-day (10 day diff) differentiated female ES cells and measured by $UV^{260}$ light absorbance. 20µg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel by electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 1% SDS buffer to the L1 $T_F$-Type monomer sense-specific (B) and antisense-specific (C) probes at 40℃. 16 hour exposures are shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

### *4.3.5 A time course of embryonic cell differentiation screened for Long terminal repeat (LTR) transposon related short RNA.*

Subsequent hybridisation of the filter shown in Figure 27 to a full-length 5S rRNA sense-specific probe as a loading control unexpectedly showed RNA of ~19-26nt at the two-day time point alone (discussed in detail in chapter 6). These short 5S rRNA also correlated with a short RNA ethidium stained band ~19nt. This result, in addition to the short antisense-specific $T_F$-type RNA observations, suggested that the accumulation of short repeat-associated RNA may be pronounced at the two-day differentiation time-point. Therefore other repeat-associated short RNA may also be detected in these samples.

The intracisternal A-particle (IAP) and murine endogenous retrovirus-L (MuERV-L) were selected to be investigated for the presence of short RNA (~20-25nt). These transposons were selected because of their high levels of expression in the pre-implantation mouse embryo (Svoboda et al., 2004). Thus they may be highly expressed in the embryonic cell lines being used. The full-length transcript of each transposon was not ideal as a probe template because their length would results in a low specific activity probe. Instead prompted by the identification of short RNA specific to the L1 $T_F$-monomer, the conserved LTR promoter regions were investigated for the generation of short RNA. Probe templates representing the LTR sequences for the IAP (~340nt) and MuERV-L (~480nt) were generated by PCR to include T7 and T3 promoter recognition sites at 5´ and 3´ ends.

Hybridisation of the sense and antisense-specific IAP and MuERV-L LTR probes to undifferentiated and two-day differentiated ES cell RNA did not to reveal any RNA in the ~20-25nt range (data not shown). It was thought that the lack of short RNA detection may be due to the low specific activity of these probes. To assess this shorter probes, with increased specific activity were designed using short LTR sequences obtained from the cloning of short RNA from mouse oocytes (Watanabe et al., 2006). Again these sequences were selected because it was thought that maternal expressed short RNA may persist in the early mouse embryo and embryonic cell lines.

Two IAP2 LTR-specific short sequences of 22nt were selected. These were both annotated as sense RNA (Watanabe et al., 2006), however they were observed to

have 100% complementary. To verify their strand specificity the IAP2 short RNA were aligned to the IAP2-LTR sequence (Jurka et al., 2005) using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). This identified the two short IAP2 sequences to be derived from the same region of the LTR in sense and antisense orientation (data not shown). These co-expressed short IAP2 RNA were of particular interest, because of their potential to form a dsRNA. Therefore T7 DNA oligonucleotide templates were designed to generate probes specific to the sense and antisense-specific IAP2 short RNA.

Using the differentiated ES cell RNA samples from Figure 27 the short sense and antisense-specific probes were hybridised. This detected several RNA bands for the IAP2 sense-specific probe, among these was a discrete RNA of ~50nt (Figure 28A). The ~50nt RNA could be observed in all the differentiated samples, however it was most abundant in the two-day differentiation sample. There was an indication of possible short RNA in the two-day and six-day differentiated samples, however these were not observed as discrete bands. Hybridisation of the RNA with the IAP2 antisense-specific probe detected a single ~70nt RNA present in all samples (Figure 28B). The abundance of this RNA appeared to increase progressively with ES cell differentiation.

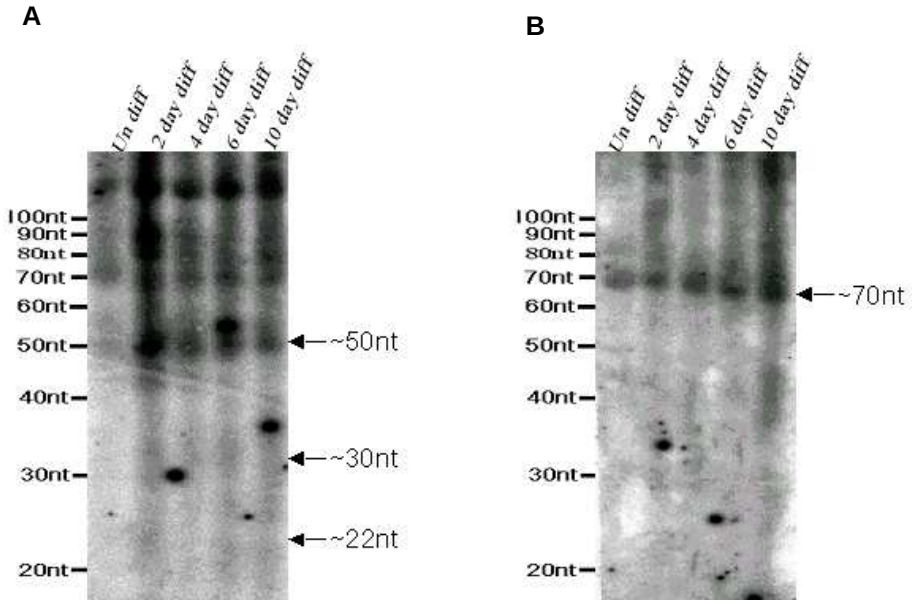


**Figure 28- Hybridisation of short IAP LTR-specific probes to ES cell differentiation time course RNA.** RNA samples from Figure 27 were stripped then hybridised in 2X SSC and 1% SDS buffer to the sense-specific (A) and antisense-specific (B) IAP2 probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. 16 hour exposures are shown for both hybridisations. The RNA of interest are indicated with an arrow (right) and marker sizes annotated on the left of the images.

For MuERV-L a single LTR specific short RNA sequence of 22nt was cloned from mouse oocyte RNA (Watanabe et al., 2006). To verify the strand and MuERV-L specificity of the short RNA, the sequence was aligned to the MERVL-LTR sequence (Jurka et al., 2005) using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). A T7 DNA oligonucleotide probe template was then designed. The differentiated ES cell RNA from Figure 27 were then hybridised to the short sense-specific MuERV-L probe. This showed there to be a single RNA of ~70nt present in all samples (Figure 29). This RNA also appeared to increase in abundance with the progressive differentiation of mouse ES cells.

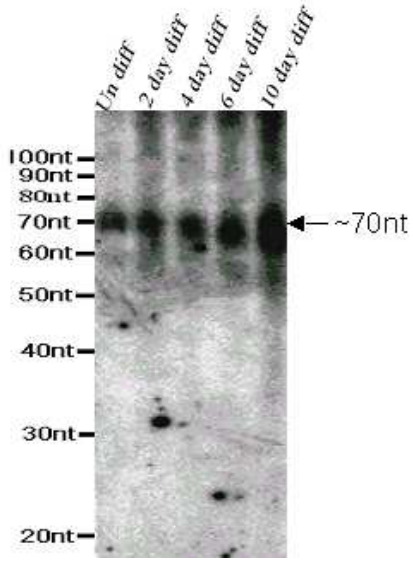


**Figure 29- ES time course and MuERV-L short probe hybridisation.** RNA samples from Figure 27 were stripped then hybridised in 2X SSC and 1% SDS buffer to the MuERV-L short at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with a 16 hour exposure shown. The RNA band of interest is indicated with an arrow (right) and marker sizes annotated on the left of the image.

### *4.3.6 B1 and L1 short RNA sequences determined by massively parallel signature sequencing (MPSS) of mouse embryos.*

(Mineno et al., 2006) describes the MPSS of 18-25nt RNA from the 9.5, 10.5 and 11.5 day mouse embryo. However, the full dataset including transposon and rRNA sequences was not published. This extended data was obtained by request (Personal communication Masayori Inouye (publication correspondent) and Yoshiyuki Notsu (Takara Bio Inc)). The complete database of short RNA present in mouse embryos was used to identify transposon-associated short RNA, that

may accumulate during the differentiation of embryonic cell lines in culture. The ability to design short transposon-specific probes was beneficial, because this would increase the sensitivity of short RNA detection by increasing the specific activity of the probe. In addition to the increased sensitivity, more sequence information on any transposon-associated short RNA detected would be gained with a short probe.

The database of short RNA present in mouse embryos was first screened for B1 sequences. For this, the full-length B1 F sequence (Jurka et al., 2005) was aligned with the complete unfiltered database, using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Table 7 shows the B1 like sequences identified. The number of transcripts per million (TPM) for each embryonic time point have been included to indicate the non-random accumulation of each sequence. All of the sequences identified were sense-specific, with the majority mapping to a single region of the B1 consensus (98-122nt). The 98-122nt region of the B1 element is part of the 29nt internal tandem repeat (discussed in chapter 3). These sequences also map to a second region of the B1 element (76-93nt), although with reduced identity due to divergence of the repeat. The 98-122nt region of the B1 element also has significant homology to a 4.5S sequence. Alignment of the sequences identified in the B1 screen to the 4.5S RNA sequence (M12658) also was carried using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi), these results are included in Table 7. From this analysis it was observed that nucleotides annotated as polymorphic to the B1 consensus sequence (highlighted in red) had 100% identity with the 4.5S sequence. Indicating the majority of the short RNA identified are likely to be derived from 4.5S and not B1 transcripts. However, the shared sequence of these RNA to the B1 consensus would allow targeting of both 4.5S and B1 if they are genuine functional short RNA.

| B1 sequence | Regions of identity & % | | | | | Embryonic day (TPM) | | |
|---|---|---|---|---|---|---|---|---|
| | Strand | B1 | 4.5 S | Sequence | % | 9.5 | 10.5 | 11.5 |
| AATCCCAGCACTTGGGAGG | Sense | 26-44 | | 1-21 | 100 | 4 | 15 | 6 |
| GTTCGAGGCCAGCCTGGGCTAC | Sense | 98-118<br>69-90 | 64-85 | 1-22<br>1-22<br>1-22 | 95<br>90<br>100 | 0 | 14 | 0 |
| TTCGAGGCCAGCCTGGGCTAC | Sense | 99-119<br>76-90 | 65-85 | 1-21<br>7-21<br>1-21 | 95<br>100<br>100 | 11 | 6 | 0 |
| TTCGAGGCCAGCCTGGGCTACA | Sense | 99-120<br>76-91 | 65-86 | 1-22<br>7-22<br>1-22 | 95<br>100<br>100 | 29 | 61 | 48 |
| GAGGCCAGCCTGGGCTACACAT | Sense | 103-122<br>76-91 | 68-89 | 2-21<br>4-19<br>1-22 | 100<br>100<br>100 | 123 | 82 | 178 |
| AGGCCAGCCTGGGCTACACATT | Sense | 103-122<br>76-91 | 69-90 | 1-20<br>3-18<br>1-22 | 95<br>100<br>100 | 69 | 66 | 94 |
| CGAGGCCAGCCTGGGCTACACA | Sense | 101-122<br>76-91 | 67-88 | 1-22<br>5-20<br>1-22 | 95<br>100<br>100 | 60 | 61 | 105 |
| TCGAGGCCAGCCTGGGCTACAC | Sense | 100-121<br>76-91 | 66-87 | 1-22<br>6-21<br>1-22 | 95<br>100<br>100 | 71 | 61 | 64 |
| GAGGCCAGCCTGGGCTACACA | Sense | 103-122<br>76-91 | 68-88 | 2-21<br>4-19<br>1-21 | 100<br>100<br>100 | 17 | 0 | 20 |
| GAGGCCAGCCTGGGCTACAC | Sense | 103-121<br>76-91 | 68-87 | 2-20<br>4-19<br>1-20 | 100<br>100<br>100 | 0 | 0 | 24 |
| GGCCAGCCTGGGCTACACATTT | Sense | 104-122<br>76-91 | 70-91 | 1-19<br>2-17<br>1-22 | 100<br>100<br>100 | 0 | 6 | 14 |
| CGAGGCCAGCCTGGGCTACATA | Sense | 103-120<br>76-93 | 67-86 | 1-22<br>5-22<br>1-22 | 91<br>100<br>95 | 0 | 11 | 6 |
| AGGCCAGCCTGGGCTACACA | Sense | 103-122<br>76-91 | 69-88 | 1-20<br>3-18<br>1-20 | 100<br>100<br>100 | 0 | 0 | 9 |

**Table 7- B1 related RNA identified in the database of short RNA present in mouse embryos.** The sequences are displayed in the column labelled “B1 sequence” with nucleotides common to 4.5S but polymorphic to the B1 consensus sequence indicated in red. Sequences are arranged by alignment to the B1 consensus with the region of alignment indicated (B1) from 5´ to 3´ with any alignment to 4.5S in (4.5S). The region of sequence identity to B1 transcript is also shown in the “sequence” column, with the percentage identity between the sequence and alignments indicated in the “%” column. The TPM value of each sequence at the 9.5 day, 10.5 day and 11.5 day (9.5,10.5 and 11.5) embryonic time points is also shown

Female mouse ES cells were cultured as standard and differentiation was then induced by the removal of LIF from the culturing medium. Undifferentiated cells were grown in parallel without the removal of LIF for comparison. Total RNA was extracted from undifferentiated two, four and six-day differentiated ES cells and measured by $UV^{260}$ light absorbance. Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide. This showed the equal loading of samples, however again the two-day differentiated RNA appeared slightly smeared in comparison to the other samples. The RNA was then electro-blotted onto nylon membrane and cross-linked using EDC.

The RNA was hybridised to probes specific for two of the sequences (26-44nt and 103-122nt) identified in the database. The probes were removed between hybridisations (method 2) and the removal confirmed by phosphor-imaging. Hybridisation to the B1 sense-specific 26-44nt probe revealed there to be an abundant RNA doublet of ~70nt in all but the two-day differentiated ES sample. In addition to the reduced abundance of the ~70nt doublet, the two-day differentiated sample had a number of additional hybridised lower RNA present in a smear (Figure 30B). However, none of these RNA were observed as discrete bands.

Hybridisation with the 103-122nt B1/68-89nt 4.5S specific probe revealed RNA of ~90nt in all samples, although it was reduced in the two-day differentiated ES sample. Several other RNA were observed in the two-day differentiated sample with possible bands at ~28nt and ~22nt observed (Figure 30C). However, these short RNA were not easily identifiable, or as discrete as the antisense-specific $T_F$ ~19nt RNA.

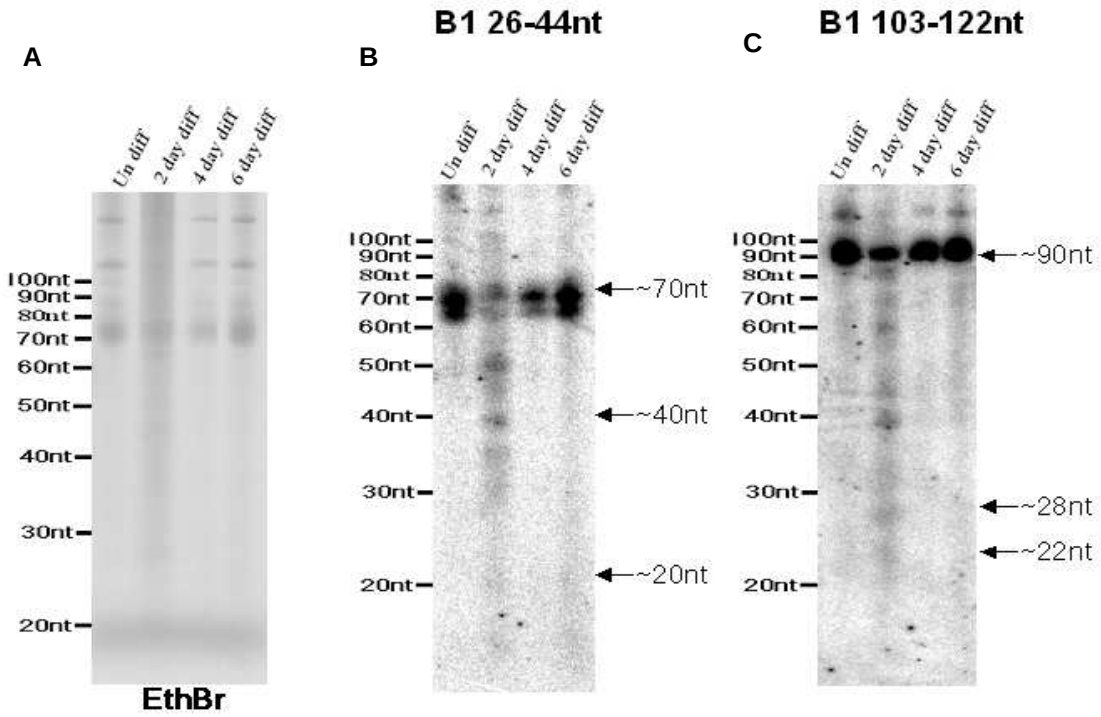


**Figure 30- B1 Takara clone hybridisation.** RNA was extracted from undifferentiated (un diff) two-day (2day diff), four-day (4 day diff) and six-day (6 day diff) differentiated female ES cells and measured by $UV^{260}$ light absorbance. 10µg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 1% SDS buffer to B1 sense-specific probes for the 26-44nt sequence (B) and 103-122nt (or specific to 68-89nt of the 4.5S sequence) (C) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with 16 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) and marker sizes annotated on the left of the images.

The L1 transposon A-type, $G_F$-type and $T_F$-type promoter monomers were then aligned with the complete unfiltered database of short RNA present in mouse embryos, using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). This did not identify any sequences with homology to the A-type and $G_F$-type monomer consensus sequences used. However, two sequences were identified with significant alignment to the $T_F$-type, one sense and the other antisense-specific (Table 8). Hybridisation of $T_F$-type monomer sequence probes to female ES RNA from a differentiation time-course failed to reveal any short RNA (data not shown).

| L1 $T_F$ sequence | Regions of identity & % | | | | Embryonic day (TPM) | | |
|---|---|---|---|---|---|---|---|
| | Strand | $T_F$ | Sequence | % | 9.5 | 10.5 | 11.5 |
| TGAGCATCTGCGGGAGACATCT | Sense | 41-62 | 1-22 | 95 | 0 | 0 | 8 |
| GAAGCAGTGTAGCTTCTGTAGT | Antisense | 75-96 | 1-22 | 95 | 0 | 9 | 0 |

**Table 8-$T_F$-type monomer related RNA identified in the database of short RNA present in mouse embryos.** The $T_F$ sequences are displayed in the column labelled "L1 $T_F$ sequence" with nucleotides polymorphic to the consensus sequence indicated in red. The region of sequence identity is also shown in the "Sequence" column, with the percentage identity of the alignment indicated in the "%" column. The TPM value for each sequence at the 9.5 day, 10.5 day and 11.5 day (9.5,10.5 and 11.5) embryonic time points is also shown.

The full-length L1 sequence (GenBank; M29324) was also aligned to the complete unfiltered sequence database using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Unlike B1, both sense and antisense-specific sequences were identified. These span the full-length of the element, with an increased abundance at the 3´of the element. These regions of high abundance short RNA map to the ORF2 and the 3´ UTR of the L1 element (Figure 31). A table showing the complete list of L1 short RNA identified has been included in appendix I. These sequence-tags await verification to assess if they represent short RNA.

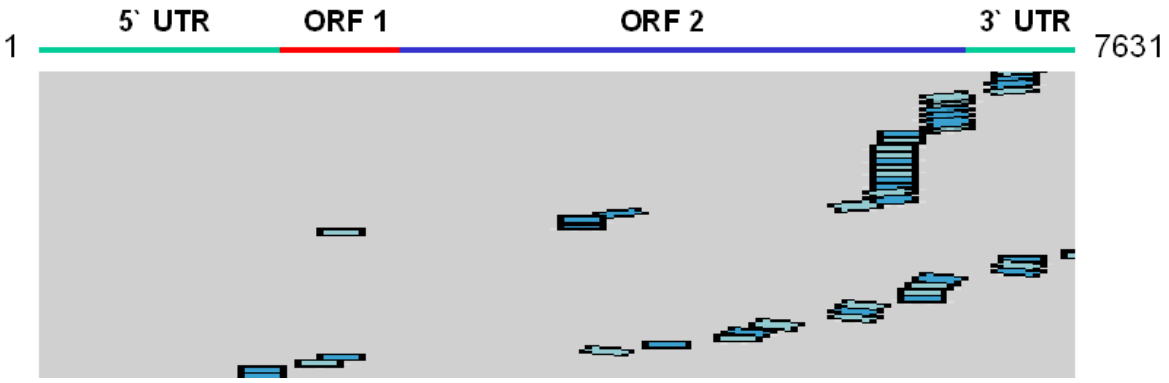


**Figure 31- Alignment of the full-length L1 to the database of short RNA present in mouse embryos.** This diagram depicts the alignment of the L1 specific sequences identified in the unfiltered embryonic dataset to the full-length L1 transcript. Each blue horizontal line indicates an individual sequence tag with the regions of L1 UTRs (green) ORF1 (red) and ORF2 (blue) indicated above.

## 4.4 Discussion

### *4.4.1 LINE-1 antisense ~19nt RNA identified in three differentiated embryonic cell lines.*

This chapter describes the extended use of mouse embryonic cell lines and adult tissue to identify transposon-associated mammalian short (20-25nt) RNA. First the active L1 transposons were investigated. This was because it was thought that highly transcribed transposons, if being processed, might have more abundant short RNA. For this, mouse L1 probes specific for different monomeric L1 promoter regions were used to investigate three active L1 transposon sub-families (Adey et al., 1994, DeBerardinis and Kazazian, 1999, Severynse et al., 1992). These probes identified several small sense and antisense-specific L1 RNA (~35-80nt) present in the RNA samples investigated (Figure 22, Figure 23 and Figure 24). In addition to these RNA, the antisense-specific $T_F$-monomer probe detected RNA of ~19nt RNA, specifically in the differentiated ES and P19EC cell RNA (Figure 24 and Figure 26). The ~19nt L1 RNA, being of a size characteristic of short interfering RNA, was of particular interest for investigating the hypothesis that RNAi directed suppression of transposable elements in mammals. However, the ~70nt RNA detected with the L1 monomer probes were also unexpected, being significantly shorter than the full-length L1 transcript and were previously unreported in the literature.

Perhaps not coincidently, the ~19nt antisense RNA was observed to associate with the accumulation of an ethidium bromide stainable band of ~19nt, not expected to be as a result of generalised RNA degradation (see chapter 7 for further discussion). It has been previously observed that a discrete SYBR gold stainable band of ~30nt associates with the accumulation of mammalian piRNA (Aravin et al., 2006, Girard et al., 2006). Therefore ~19nt ethidium stainable band may also indicate the accumulation of a subgroup of short RNA, induced by the differentiation of mouse embryonic cell lines in culture, which include the antisense $T_F$-monomer ~19nt RNA.

The identification of a short promoter specific-antisense transposon RNA is supportive of a post-transcriptional RNAi mediated suppression of L1-$T_F$ transposable elements. Previously, synthetic short-antisense RNA targeting the

5´UTR of murine L1 transcript, have been observed to effectively post-transcriptionally repress L1 transcript levels (Beraldi et al., 2006). Additionally there is growing evidence of promoter-specific antisense siRNA being capable of inducing heterochromatin formation in mammalian cells (Weinberg et al., 2006). Therefore the endogenous expression of the $T_F$-monomer ~19nt antisense RNA, could theoretically target the 5´UTR of L1-$T_F$ transposons both transcriptionally and post-transcriptionally *in vivo*. However, the ability of the endogenous ~19nt antisense RNA detected to be incorporated into the RNAi pathway has yet to be established.

The hypothesis of endogenous short RNA, mediating the repression of murine L1 through the RNAi pathway is supported in the literature, with Dicer knockdown being reported to increase transcription of L1 in mouse and human (Kanellopoulou et al., 2005, Yang and Kazazian, 2006). Additionally, sense and antisense ~21nt RNA derived from the human L1 5´UTR promoter region have recently been detected, using a gel blotting approach (Yang and Kazazian, 2006). The human L1 transposon unlike the mouse does not contain the tandem monomer-repeats. Instead the human L1 specific ~21nt RNA were derived from a 500nt region of bi-directional transcription in the human 5´UTR (Speek, 2001). The bi-directional transcription of the 5´UTR and the expression of Dicer were both established to be essential for the repression of human L1 transposons (Yang and Kazazian, 2006), thus implicating the RNAi pathway and short RNA in the suppression of the human L1 element.

The identification of endogenous short RNA, specific to the L1 promoter of two mammalian species may support a conserved targeting of the promoter region. However, because the promoter sequences of human and mouse L1 elements are unrelated, the L1 5´UTR short RNA are likely to have evolved separately if targeting this region. In fact, specifically targeting the promoter region of transcribed transposons may provide the most efficient method of repressing L1 transposable elements. This is because only active elements, which pose a threat to the host genome would be targeted by transcriptional or post-transcriptional repression. The specific targeting of active transposons may also explain why only short RNA ~19-25nt specific to the murine L1-$T_F$ 5´UTR (the most active mouse L1 sub-family) were at sufficient levels for detection by gel blotting.

Other short RNA were also observed of a size characteristic of miRNA precursors with the sense (~70nt) and antisense-specific (~70 and ~80nt) $T_F$-monomer probes (Figure 24). However, unlike the ~19nt RNA these were detected in all of the samples analysed. With the $T_F$-monomer probes both sense and antisense-specific ~70nt RNA were detected. This may indicate there to be a potential dsRNA of ~70nt. However, it is unknown if these RNA overlap to form potential dsRNA and precursors to the ~19nt antisense RNA. If these RNA were precursors to the ~19nt antisense RNA, their processing would have to be somehow restricted to the differentiated embryonic cell samples. Alternatively it could be argued that the ~70nt RNA observed were due to the cross-hybridisation of the probe with the tRNA. However, the variable hybridisation of the antisense $G_F$-monomer to the male and female ES cell RNA with ~80nt and~70nt bands hybridised respectively, would argue against the non-specific hybridisation of the tRNA (see page 107).

Further characterisation of the ~19nt antisense L1 RNA is required for understanding any potential function it may have in the repression of the L1 transposon. Firstly the sequence of the L1 $T_F$-monomer ~19nt antisense RNA would need to be defined to provide insight into its processing. Attempts were made to further define the ~19nt RNA by screening the database of short embryonic sequences present in mouse embryos. However, the single antisense $T_F$-monomer sequence identified was not found to associate with the ~19nt antisense RNA. Alternatively this could be undertaken using short DNA oligonucleotide probes, in a method similar to that used to map the Alu ~70nt RNA. Other future work on this project would primarily focus on assessing interactions of the ~19nt RNA with the RNAi pathway. These experiments could include assessing Dicers involvement in the processing of the ~19nt RNA, using Dicer knockout cells and the *in vitro* Dicer digestion of 70nt RNA. Also assessing the targeting ability of the short RNA through assessing the knockdown of reporter genes containing $T_F$-monomer sequence or the immunoprecipitation of the argonaute proteins, which are known components of the RISC complex.

### 4.4.2 The unexpected observation of short LTR and SINE transposon RNA in differentiated mouse embryonic cell line RNA.

LTR specific probes to IAP and MuERV-L transposons were designed using 22nt sequences cloned from mouse oocytes (Watanabe et al., 2006). The complementary IAP2 LTR-specific short RNA probes identified several defined RNA ~50-120nt which are significantly smaller in size than the full-length IAP element (Figure 28). The observation of a single ~70nt antisense IAP2 RNA was of particular interest due to the size similarity to miRNA precursors. The sense-specific RNA were less well defined, however a range of bands similar in size to miRNA precursors were observed. Given the 22nt probes used were complementary to each other this may suggest the potential for dsRNA formation of the ~70nt antisense RNA with several of the sense RNA observed. In addition to the ~50-120nt RNA there was also an indication of shorter RNA of ~30nt and ~22nt in the two and six-day differentiated samples with the IAP sense-specific probe. However, these lower RNA were not of the discrete nature expected for precisely processed short RNA. Hybridisation of the MuERV-L LTR specific short probe also identified a ~70nt RNA (Figure 29). However, no shorter RNA were observed.

The degree of similarity of the RNA populations between mouse oocytes (9,10 and 11 day) and cultured embryonic cell lines is unknown but is likely to vary greatly as angeogenesis has begun in the former but not in the latter. Therefore the inability to confirm the oocyte cloned LTR short RNA during the differentiation of embryonic cells was not unexpected. However, the observation of two 22nt RNA, with 100% complementary originating from the IAP2 LTR region is note worthy. These RNA have been miss annotated and this observation appears to have been overlooked in the original publication (Watanabe et al., 2006). Further investigation of the expression of these RNA could therefore be worthwhile.

The use of the full-length B1 sense and antisense-specific probes indicated potential antisense RNA of ~45nt, ~30nt and ~22nt in the differentiated ES samples (Figure 25). These RNA were in addition to the ~70nt antisense RNA described in chapter 3 (Figure 11). However, with the full-length antisense

probe these RNA were not discrete as would be expected for a functional short RNA. Analysis of short RNA sequences obtained from mouse embryos identified a number of short RNA with identity to the B1 transposon (Table 7). However, the majority of these appeared to be derived from the 4.5S RNA, which has some sequence similarity to the B1 transposon. Most of the sequences obtained mapped to a single region of the B1 element, with a single B1 specific sequence tag identified at the 5` region. Hybridisation of these two regions with short probes did identify what appeared to be discrete short RNA of various sizes down to ~20nt (Figure 30). However, the short RNA signals for the ~20nt and ~22nt RNA were very faint and would require further work to establish if they are genuine discrete short RNA. The appearance of these shorter RNA was accompanied by a reduction in the ~70n and ~90nt RNA of the 26-44nt and 103-122nt specific probes respectively. This may indicate a stepwise processing pathway of these RNA generating several intermediate bands to produce B1 specific RNA.

# 5

# B2 SINE short RNA (~20nt) accumulate during the differentiation of mouse embryonic cells in culture.

## Results 3

## 5.1 Summary

This chapter focuses on the identification of short RNA specific to the B2 retrotransposon in differentiated ES and P19 EC cell lines using a northern blot approach.

Probes were initially designed to the full-length B2 consensus sequence. The full-length sense and antisense-specific B2 probes were then hybridised to embryonic cell RNA extracted at different time periods of differentiation in cell culture. An RNA of ~80nt was unexpectedly observed with both probes. In addition, the B2 sense-specific probe appeared to detect the accumulation of shorter RNA, including a possible band of ~20nt in the two-day differentiated samples. However, these RNA were not well defined. In an aim to identify if the ~20nt band was genuine, short B2 related RNA identified in the database of short RNA present in mouse embryos (Mineno et al., 2006) were used to design probes. The improved specific activity associated with the use of shorter probes, lead to the detection of several discrete sense and antisense short B2 RNA (~90-18nt). Two of these, corresponding to the 1-22nt promoter region of the B2 sequence were 100% complementary to each other. Both of these probes were unexpectedly found to detect short RNA of ~20nt, predominantly after two days of ES or P19 EC cell differentiation. Several other RNA (~30-90nt) were also observed to be differentially expressed (usually up-regulated) in the two-day differentiated sample indicating there to be processing of the B2 transcripts. The potential processing of the ~20nt B2 RNA by Dicer was then investigated using RNA obtained from the differentiation of Dicer-null ES cells. This indicated there to be a reduction in B2 short RNA processing however a ~20nt RNA could still be observed.

Computer based analysis was also used to identify possible mRNA targets for the short B2 RNA. This identified many possible targets with 100% identity. However, using a dual luciferase sensor assay no targeting of the B2 short RNA was observed.

# 5.2 Introduction

## *5.2.1 The tRNA derived rodent SINE B2.*

The B2 retrotransposon is a non-autonomous SINE derived from tRNA sequence, which is predominantly restricted to the genomes of rodent species. However, it can also be found at low levels in the human genome (~50 copies)(Mayorov et al., 2000). B2 retrotransposition is mediated by autonomous L1 elements much like the 7SL derived B1 SINE discussed in chapter 3. The rate of B2 transposition is reportedly higher than that of B1, being more comparable to the human Alu SINE transposition rate (Dewannieux and Heidmann, 2005).

The modern murine B2 element is ~200nt in length and is transcribed by RNA polymerase III from the tRNA derived internal promoter. In addition to the tRNA derived sequence, B2 has acquired sequence of unknown origin containing an active RNA polymerase II promoter. This promoter is in the opposite orientation to the RNA polymerase III promoter (Ferrigno et al., 2001). A polyadenylated tail is also present at the 3` end of the B2 element. Figure 32 shows a diagram of the B2 element.

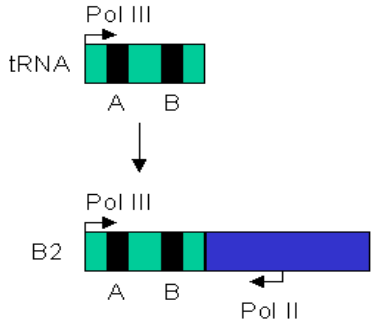


**Figure 32- Origins of the B2 transposable element from tRNA.** B2 is derived from the tRNA methionine (green) and includes the A and B box (black) essential for RNA polymerase III (Pol III) activity. In addition to the tRNA derived region there is an acquired 3´ region of unknown origin containing an active RNA polymerase II promoter (Pol II). This promoter is in the antisense orientation to the RNA polymerase III promoter. The B2 element also contains a polyadenylated tail at the extreme 3´.

B2 transcription is strongly developmentally regulated. It is transcribed at high levels in the oocyte, increasing upon fertilisation with subsequent down-regulation in the early embryo (White et al., 1989). B2 transcription has also been reported to be up-regulated during heat shock, where it is thought to function in the repression of RNA polymerase II (Espinoza et al., 2004).

## 5.3 Results

### *5.3.1 B2 short RNA (~80-20nt) unexpectedly identified in two-day differentiated ES and P19 EC cells.*

Because the B2 transposon represented another highly abundant transposon transcribed in the early embryo, RNA harvested from ES cells undergoing differentiation at timed intervals was also examined for the presence of B2 short RNA ~20-25nt. For this a full-length B2 specific consensus probe template was generated by PCR to include T7 and T3 RNA polymerase promoter recognition sites at the 3´ and 5´ends (see materials and methods).

Female mouse ES cells were grown as standard with differentiation induced by the removal of LIF from the culturing medium. Undifferentiated cells were grown in parallel for comparison. Total RNA was extracted from undifferentiated, two, four, six and ten-day differentiated ES cells and the concentration measured by $UV^{260}$ light absorbance. Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide. This showed the equal loading of samples and the two-day differentiated RNA to have a slightly smeared appearance (Figure 33A). The RNA was then electro-blotted onto nylon membrane and cross-linked using EDC.

The RNA was hybridised to the full-length B2 sense and antisense-specific probes. Probe removal (method 2) was confirmed by phosphor-imaging between hybridisations. The full-length B2 (~200nt) were not observed because the gel conditions used were optimised for the observation of RNA <100nt. However, an unanticipated sense RNA of ~80nt was identified in all samples. Shorter RNA bands of ~60nt and ~32nt were also unexpectedly observed in the two-day differentiated ES sample (Figure 33B). A band of ~20nt in the two-day differentiated sample may have also been present. However, high levels of background in this area of the filter made this difficult to establish. The antisense-specific probe also unexpectedly identified an RNA of ~80nt in all samples, with a RNA of ~60nt in the two-day differentiated ES sample alone (Figure 33C).

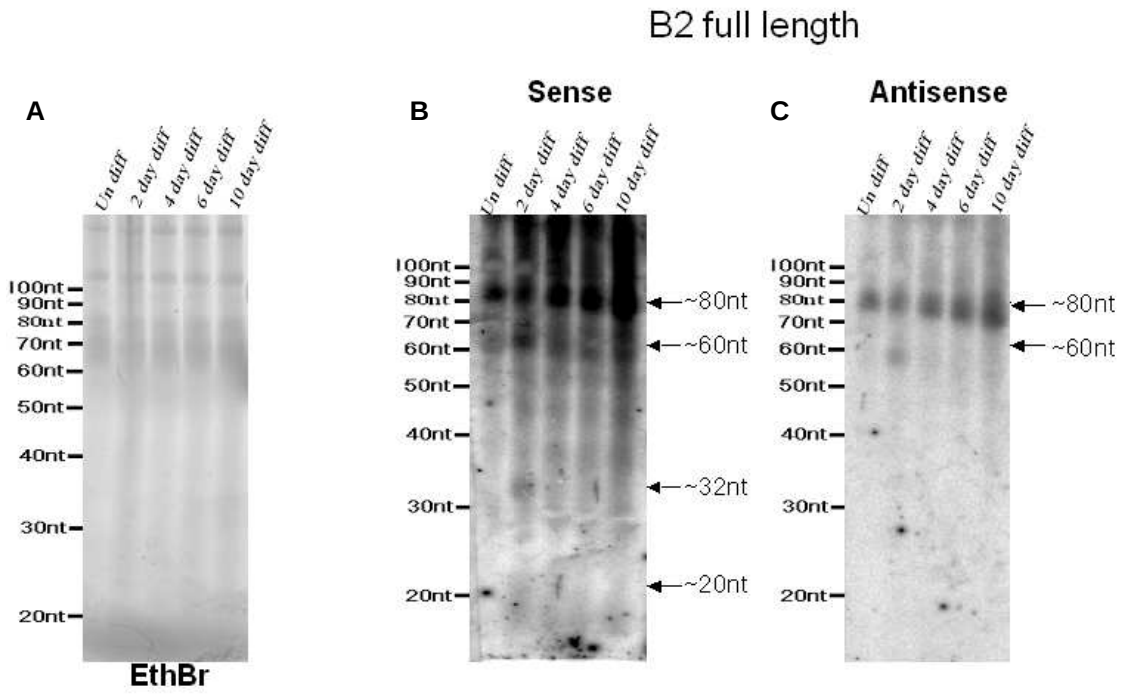


**Figure 33- ES time course and full-length B2.** RNA was extracted from undifferentiated (un diff), two-day (2day diff), four-day (4 day diff), six-day (6 day diff) and ten-day (10 day diff) differentiated female ES cells and measured by $UV^{260}$ light absorbance. 20μg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 1% SDS buffer to the full-length B2 sense-specific (B) and antisense-specific (C) probes at 40℃, wash 0. 2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with 16 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

The differentiation time course was then repeated using P19EC cells to test if these RNA were a specific feature of ES cells, or a more general phenomenon. Mouse P19 EC cells were grown as standard and were treated with retinoic acid to induce differentiation. Undifferentiated cells were grown in parallel for comparison. Total RNA was extracted from undifferentiated, one, two, four and six-day differentiated P19 EC cells and the concentration measured by $UV^{260}$ light absorbance. Equal amounts of RNA were separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide (Figure 34A). This staining, although weak, did show the equal loading of samples. The two-day differentiated RNA were again observed to be slightly smeared in comparison to the other RNA samples.

The RNA was then hybridised to the full-length B2 sense-specific probe. This showed there to be a sense RNA of ~80nt in all samples as observed for the female ES cell differentiation time-course. This ~80nt RNA appeared to be

reduced in the two-day differentiated sample where shorter RNA bands of ~42nt, ~35nt and 20nt were observed (Figure 34B). Although the lower bands were not well defined, the ~20nt B2 sense RNA was more visible in the two-day differentiated P19 EC cell RNA compared with to the ES cell RNA hybridisation (Figure 33B). Hybridisation of the full-length antisense-specific B2 was not carried-out for the P19 EC differentiation time course RNA. Therefore it was not known if antisense-specific short RNA also accumulated in the P19 EC RNA.

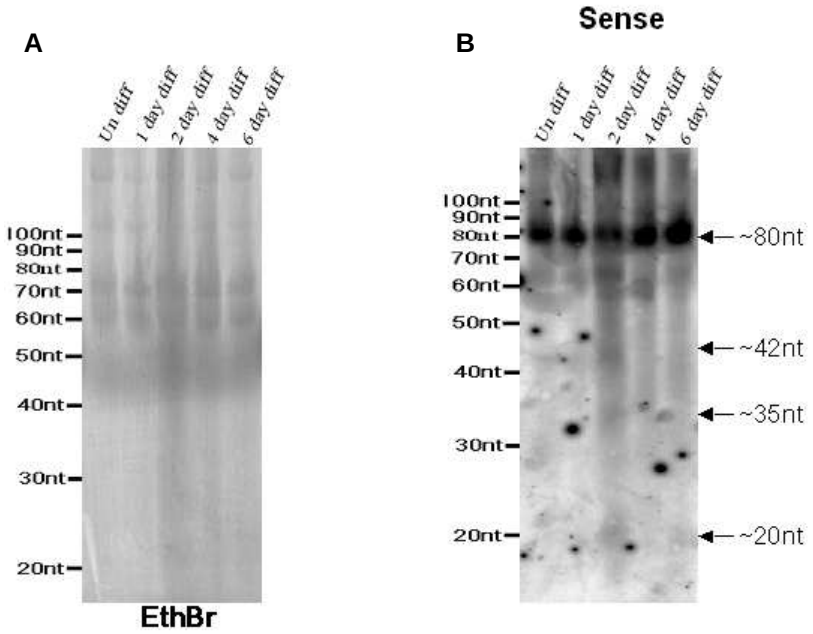


**Figure 34-P19 EC cell differentiation time course hybridised to full-length B2.** RNA was extracted from undifferentiated (un diff) one-day (1 day diff), two-day (2 day diff), four-day (4 day diff) and six-day (6 day diff) differentiated P19 EC cells and measured by $UV^{260}$ light absorbance. 40µg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 1% SDS buffer to the full-length B2 sense-specific (B) probe at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with a 16 hour exposure shown. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

### *5.3.2 B2 related RNA identified in the database of short RNA present in mouse embryos are expressed in two-day differentiated ES and P19 EC cells.*

The detection of B2 RNA, of a size characteristic of siRNA and miRNA was of most interest in terms of investigating a hypothesis of transposon regulation through RNAi. However, with the full-length B2 probes, these RNA were poorly resolved. To improve the detection of any genuine ~20nt RNA, short T7 probes

were designed to B2 sequences identified in the unfiltered database of short RNA present in mouse embryos (Mineno et al., 2006).

The database was screened for B2 sequences by the alignment of the full-length B2 transcript (Jurka et al., 2005) to the complete unfiltered sequence data using the BLAST (bl2seq) alignment tool on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Table 9 lists the B2 sequences identified. The number of transcripts per million (TPM) for each embryonic time point have been included to indicate the non-random accumulation of each sequence. Nucleotides highlighted in red indicate polymorphism to the particular B2 sequence used. However, all of these B2 variations were subsequently identified in the mouse genome using the BLAST alignment tool on the ensemble website (http://www.ensembl.org/index.html).

| B2 sequence | Regions of identity & % | | | | Embryonic day (TPM) | | |
|---|---|---|---|---|---|---|---|
| | Strand | B2 | Sequence | % | 9.5 | 10.5 | 11.5 |
| GGGCTGGAGAGATGGCTCAG | Sense | 1-20 | 1-20 | 100 | 17 | 23 | 68 |
| GGGCTGGAGAGATGGCTCA | Sense | 1-19 | 1-19 | 100 | 0 | 0 | 18 |
| CGGGCTGGAGAGATGGCTCA | Sense | 1-19 | 2-20 | 100 | 0 | 0 | 10 |
| GGGCTGGAGAGATGGCTC | Sense | 1-18 | 1-18 | 100 | 0 | 0 | 9 |
| TTGAGCCATCTCTCCAGCCCT | Antisense | 1-19 | 2-20 | 100 | 11 | 0 | 0 |
| CTGAGCCATCTCTCCAGCCCTA | Antisense | 1-20 | 1-20 | 100 | 0 | 8 | 0 |
| ACTGCTCTTCCAGAGGTCCTGA | Sense | 37-58 | 1-22 | 100 | 19 | 4 | 2 |
| TGACTGCTCTTCCAGAGGTCCT | Sense | 35-56 | 1-22 | 100 | 9 | 2 | 0 |
| GAGTTCAATTCCCAGCAACCA | Sense | 57-77 | 1-21 | 100 | 9 | 0 | 0 |
| GAGCTCAATTCCCAGCAACCA | Sense | 57-77 | 1-21 | 95 | 11 | 0 | 0 |
| CTCTCCTGGTGTGTCTGAAGA | Sense | 117-137 | 1-21 | 95 | 0 | 13 | 0 |

**Table 9- B2 related RNA identified in the database of short RNA present in mouse embryos.** The B2 sequences are displayed in the column labelled “B2 sequence” with nucleotides polymorphic to the B2 consensus sequence used, indicated in red (all polymorphisms were found to be present in the mouse genome). The sequences are arranged by alignment to the B2 consensus with the region of alignment indicated (B2) from 5´ to 3´. The region of sequence identity to the B2 consensus is also shown in the “Sequence” column, with the percentage identity between the sequence and alignments indicated in the “%” column. The TPM value of each sequence at the 9.5 day, 10.5 day and 11.5 day (9.5,10.5 and 11.5) embryonic time points is also shown.

Of particular interest were the sense and antisense-specific sequences at the extreme 5´ of B2 sequence. This was because in addition to these sequences being the most abundant B2 identified, they also formed a hypothetical double-stranded RNA of ~20nt. T7 probe templates were designed to enable the specific detection of three of the B2 sequences identified through the computer analysis. These mapped to the 5´ sense and antisense-specific regions of the B2 transcript (1-22nt) and the internal sense-specific region (35-58nt).

Female mouse ES cells were grown as standard with differentiation induced by the removal of LIF from the culturing medium. Undifferentiated cells were grown in parallel for comparison. Total RNA was extracted from undifferentiated, two, four and six-day differentiated ES cells and the concentration measured by $UV^{260}$ light absorbance. Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide (Figure 35A). This showed the equal loading of samples and the two-day differentiated RNA to be quite smeared. The RNA was electro-blotted onto nylon membrane and cross-linked using EDC.

The RNA was then hybridised to the B2 sense and antisense-specific 1-22nt probes. Hybridisation to the sense-specific 1-22nt region clearly identified discrete RNA of ~20nt and ~18nt (Figure 35B). These are both within the size range of short functional RNA such as siRNA and were predominantly observed to accumulate at the two-day differentiation time point. In addition, several other RNA bands also accumulated in the two-day differentiated sample. These included RNA of ~70nt, ~50nt and ~35nt as indicated in Figure 35B. Conversely an RNA of ~90nt seen in all other samples was observed to disappear in the two-day differentiated RNA sample.

The probe was removed (method 2 materials and methods) with removal confirmed by phosphor-imaging (Figure 35C). The RNA was then hybridised to the B2 1-22nt antisense-specific probe (Figure 35D). This showed there to be an antisense RNA of ~20nt in the two-day differentiated sample. Similar to the sense ~20nt and ~18nt RNA, the appearance of the antisense ~20nt RNA was accompanied by the disappearance of a ~90nt RNA, which is observed in all other samples.

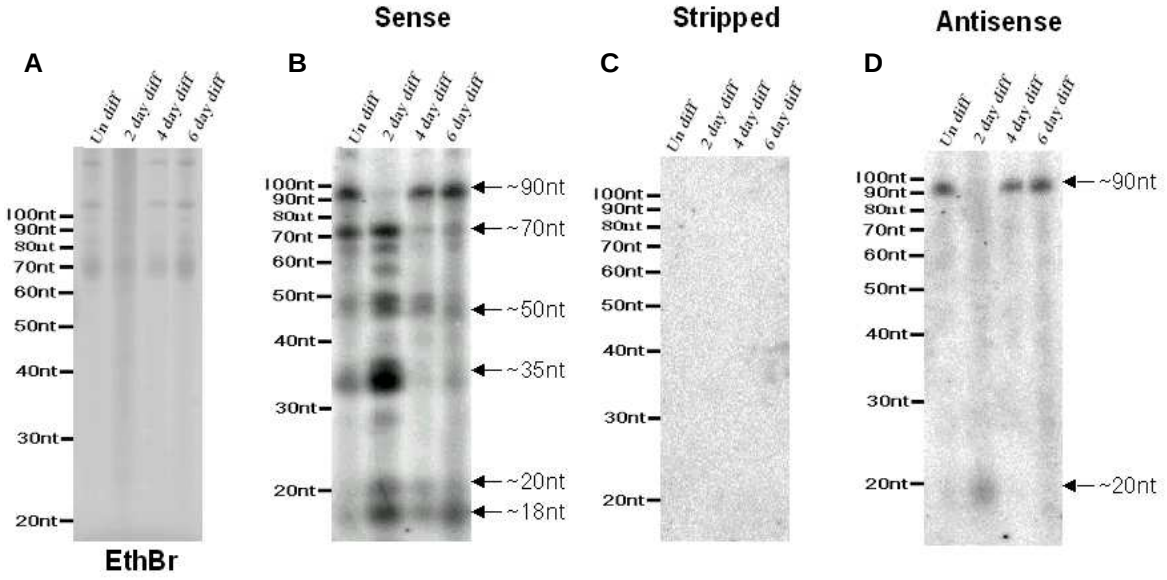


**Figure 35- B2 1-22nt sense and antisense probe hybridisation to RNA extracted from an ES differentiation time course.** RNA was extracted from undifferentiated (un diff) two-day (2day diff), four-day (4 day diff) and six-day (6 day diff) differentiated female ES cells and measured by $UV^{260}$ light absorbance. 10µg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then sequentially hybridised in 2X SSC and 1% SDS buffer to the B2 1-22nt sense-specific (B) and antisense-specific (D) probes at 40℃, wash 0.2X SSC 0.2% SDS 4 0℃. Probe removal was carried out between hybridisations (C). The radioactive signal was recorded by phosphor-imaging with 16 hour exposures shown for both hybridisations and a one hour exposure of the stripped filter. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

The probe 1-22nt antisense probe was removed from the ES cell RNA, this was verified by phosphor-imaging. The RNA was then hybridised to the B2 35-58nt sense-specific probe (Figure 36). This showed there to be a sense RNA of ~30nt and ~20nt predominantly in the two-day differentiated sample, with a ~30nt RNA also accumulating in the six-day differentiated sample. No discrete longer RNA were observed with this probe, however a smear ~70-80nt was observed in the undifferentiated and two-day differentiated samples.

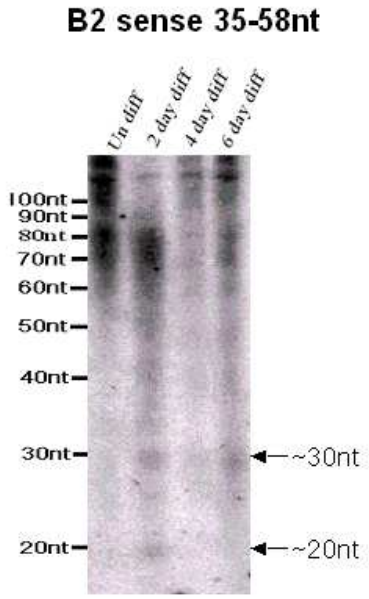


**Figure 36- ES time course and B2 35-58nt probe hybridisation.** ES differentiation time course RNA samples from Figure 35 were stripped then hybridised in 2X SSC and 1% SDS buffer to the B2 35-58nt sense-specific probe at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with a 16 hour exposure shown above. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

The 1-22nt B2 sense and antisense-specific probes also hybridised to short RNA of ~20nt in P19 EC cell RNA (data not shown). These were again predominantly expressed in the two-day differentiated P19 EC cells and correlated with the disappearance of ~90nt band.

### *5.3.3 Reduced accumulation of B2 short RNA during the differentiation of Dicer-null ES cells in culture.*

Mouse Dicer is encoded by a single genomic locus, with homozygous null mutations unable to process known miRNA. The Dicer-null ES cells used were produced by the laboratory of G. Hannon (Cold Spring Harbor Laboratories, USA). This was done by replacing the wild type Dicer allele of ES cells already containing a Dicer-null allele, with a "floxed" allele in which loxP sites flanked exons 22 and 23 (location of the RNase III catalytic domain). Cre treatment of ES cells was then used to delete exons 22 and 23 of the floxed Dicer allele by recombination between the loxP sites. A complete description of the Dicer-null ES cell generation and properties is published (Murchison et al., 2005). These Dicer-null embryonic cells along with the AB 2.2 control cell-line were subjected to differentiation. This was carried out to assess if Dicer was involved in the processing of the B2 specific short RNA.

The growth and differentiation of the Dicer-null and control (AB 2.2) ES cells was carried out as described in the materials and methods. Total RNA was extracted from undifferentiated, two, four and six-day differentiated ES cells and measured by $UV^{260}$ light absorbance. Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide (Figure 37A). This showed the equal loading of the majority of samples. However, the loading of the undifferentiated and two-day differentiated AB 2.2 cell RNA was difficult to determine because of smearing in comparison to other samples. The RNA was then electro-blotted and cross-linking to nylon membrane using EDC.

The RNA was first hybridised to the B2 (1-22nt) sense-specific probe because this represented the most abundant B2 short RNA observed. This identified short B2 sense RNA of ~20nt for the control undifferentiated and two-day differentiated samples. To a lesser extent short sense RNA of ~20nt were also observed for the two-day differentiated Dicer-null RNA sample (Figure 37B). Coinciding with the appearance of the ~20nt sense RNA, was the presence of several other RNA as had previously been observed, although in the two-day Dicer-null sample this was less pronounced. The sense RNA of ~90nt previously observed to be absent in the two-day differentiated ES (Figure 35B) and P19 EC cell RNA (data not shown) appeared to be reduced with the induction of the ~20nt B2 sense RNA in the control samples. However, no change of the ~90nt sense RNA was observed for the two-day Dicer-null sample. An RNA of ~18nt could also be observed in all samples, although at reduced levels in the Dicer-null samples. The B2 1-22nt sense-specific probe was then removed (method 2) with verification by phosphor-imaging. The RNA was subsequently hybridised to a probe specific for mmu-mir-292as (Figure 37C), which is known to be highly expressed in mouse ES cells (Houbaviy et al., 2003). This was carried out to assess the null phenotype, as the Dicer-null cells should not be capable of producing mature miRNAs. This hybridisation showed an RNA corresponding to the mature miRNA in all the control ES cell RNA samples. No signal was observed in the Dicer-null RNA samples, which is consistent with no miRNA processing in these cells.

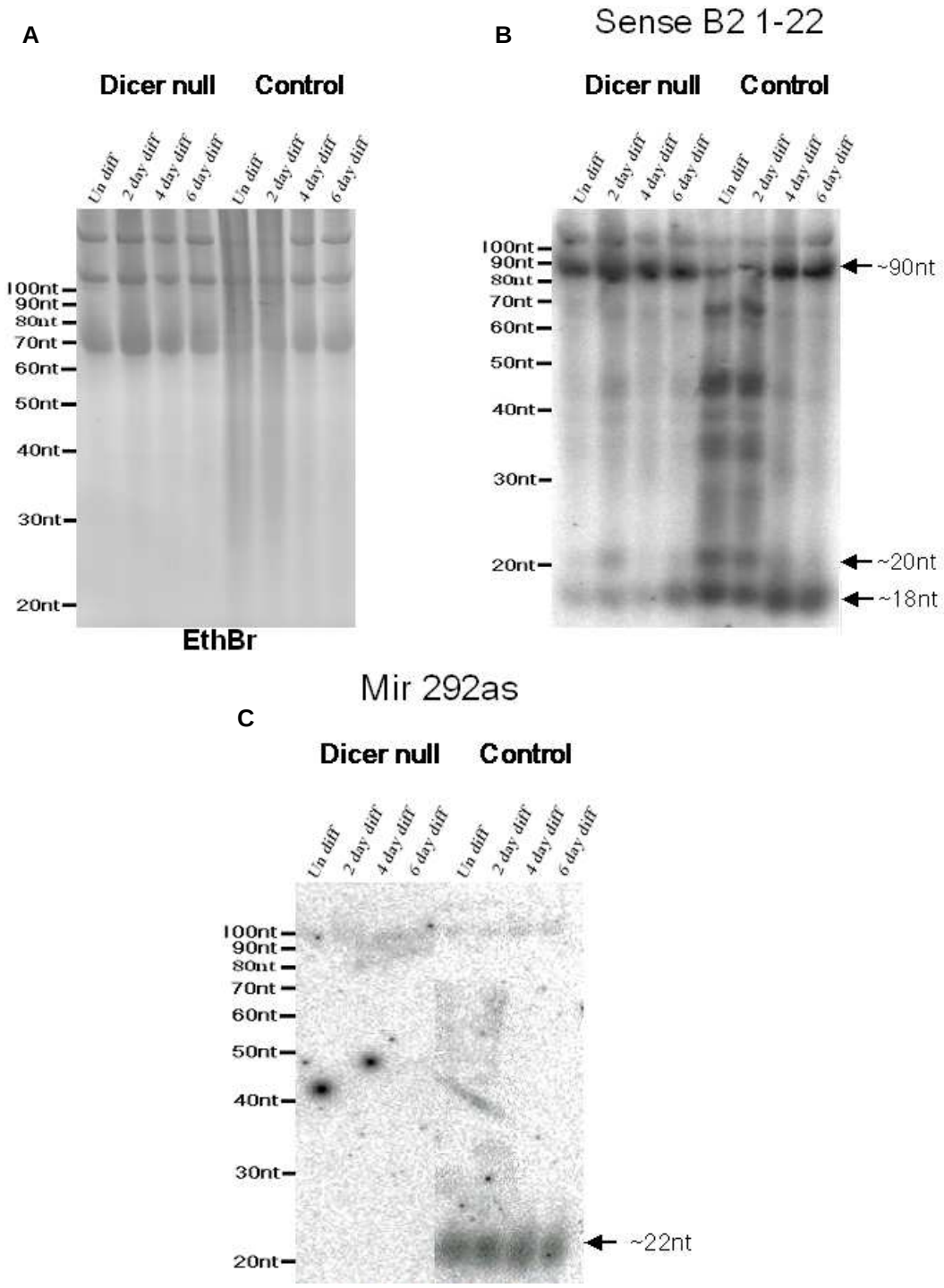


**Figure 37- B2 5 prime sense specific RNA in Dicer-null.** RNA was extracted from 0 (Un diff) and 2-day (2 day diff), 4-day (4 day diff) and 6-day (6 day diff) differentiated Dicer-null and ES AB 2.2 cells were measured by $UV^{260}$ light absorbance. 40µg/well of each RNA sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (A) prior to electro-blotting and EDC cross-linking to nylon membrane. The RNA was hybridised in 2X SSC and 1% SDS buffer to the B2 (1-22nt) sense-specific (B) and mmu-mir-292as (C) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with 16 hour exposures shown for both hybridisations. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

### 5.3.4 The B2 short ~20nt RNA identified are complementary to many mRNA sequences in the mouse genome.

There are many transposable elements incorporated into mRNA transcripts in the mouse genome. If the short B2 RNA identified were capable of interaction and regulation through the RNAi pathway, these mRNA may also be targeted (collateral damage). To identify targetable mRNA, the mouse expressed sequence tags (EST) database was screened for sequences complementary to the short B2 sequences identified.

First the B2 1-21 sense sequence was screened against the mouse EST database. This uncovered many targetable sites located within mRNA sequences, with some containing multiple target sites. Table 10 displays some of the highest identity mRNA target sites identified. The length and percentage of identity for each target site and its position within the mRNA transcript have been included.

| Protein name | mRNA accession no | Identity | B2 1-21 Sense | mRNA | Location |
|---|---|---|---|---|---|
| Commd7 | NM_133850 | 100% | 1-21 | 2366-2346 | 3`UTR |
| Tiprl | NM_145513 | 100% | 1-21 | 3153-3133 | 3`UTR |
| Rsad1 | NM_001013381 | 100% | 2-21 | 3529-3510 | 3`UTR |
| | | 100% | 1-19 | 3007-2989 | 3`UTR |
| | | 100% | 2-20 | 3352-3334 | 3`UTR |
| Tmco3 | NM_172282 | 100% | 1-20 | 3800-3781 | 3`UTR |
| | | | 2-19 | 3292-3275 | 3`UTR |
| Zfp689 | NM_175163 | 100% | 1-20 | 2758-2739 | 3`UTR |
| Cd300a | NM_170758 | 100% | 1-20 | 2110-2091 | 3`UTR |
| | | 100% | 2-20 | 2562-2544 | 3`UTR |
| Padi2 | NM_008812 | 100% | 1-20 | 3740-3721 | 3`UTR |

**Table 10- High identity B2 1-21 sense-specific mRNA target sites.** The B2 1-21nt sense-specific probe sequence was aligned to the mouse EST database, 11 target sites with 100% identity are included in the table above. The protein names identified with each mRNA transcript are displayed in the column labelled “Protein name” with the related accession number “mRNA accession no”. The percentage of complementarity between the mRNA and the predicted B2 1-21 short sense RNA is displayed in “identity” with the length of the alignment in “B2 1-21 sense” column. Finally the nucleotide position and the location of the targetable site within the mRNA are shown in the “mRNA” and “location” columns respectively.

Subsequently the B2 1-21 antisense-specific and 35-58nt sense-specific sequences were screened against the mouse EST database. This also showed there to be many targetable sites located within mRNA sequences. Again the

highest complementary mRNA target sites identified for the B2 1-21 antisense and 35-58nt sense short RNA are displayed in Table 11 and Table 12 respectively.

| Protein name | mRNA accession no | Identity | B2 1-21nt antisense | mRNA | Location |
|---|---|---|---|---|---|
| Hemk2 | NM_026366 | 100% | 1-21 | 1477-1457 | 3`UTR |
| Ddo | NM_027442 | 100% | 1-21 | 2529-2509 | 3`UTR |
| Dscr6 | NM_133229 | 100% | 2-21 | 1061-1041 | 3`UTR |
| Atpaf2 | NM_145427 | 100% | 2-21 | 1595-1576 | 3`UTR |
| Sap18 | NM_009119 | 100% | 2-21 | 931-912 | 3`UTR |
| Pts | NM_011220 | 100% | 2-20 | 794-776 | 3`UTR |
| Ociad2 | NM_026950 | 100% | 2-20 | 1372-1354 | 3`UTR |

**Table 11- High identity B2 1-21 antisense-specific mRNA target sites.** The B2 1-21nt antisense-specific probe sequence was aligned to the mouse EST database, 7 target sites with 100% identity are included in the table above. The protein names identified with each mRNA transcript are displayed in the column labelled “Protein name” with the related accession number “mRNA accession no”. The percentage of complementarity between the mRNA and the predicted B2 1-21 short antisense RNA is displayed in “identity” with the length of the alignment in “B2 1-21 antisense” column. Finally the nucleotide position and the location of the targetable site within the mRNA are shown in the “mRNA” and “location” columns respectively.

| Protein name | mRNA accession no | Identity | B2 35-58nt sense | mRNA | Location |
|---|---|---|---|---|---|
| Mllt11 | NM_019914 | 100% | 1-22 | 1689-1668 | 3`UTR |
| Homez | NM_183174 | 100% | 1-22 | 2978-2957 | 3`UTR |
| Pigv | NM_178698 | 100% | 2-22 | 2563-2543 | 3`UTR |
| Ube2j2 | NM_021402 | 100% | 1-22 | 3036-3015 | 3`UTR |
| Rsad1 | NM_001013381 | 100% | 2-22 | 3317-3297 | 3`UTR |

**Table 12- High identity B2 35-58nt sense-specific mRNA target sites.** The B2 35-58nt sense-specific probe sequence was aligned to the mouse EST database, 5 target sites with 100% identity are included in the table above. The protein names identified with each mRNA transcript are displayed in the column labelled “Protein name” with the related accession number “mRNA accession no”. The percentage of complementarity between the mRNA and the predicted B2 35-58nt short sense RNA is displayed in “identity” with the length of the alignment in “B2 35-58nt S” column. Finally the nucleotide position and the location of the targetable site within the mRNA are shown in the “mRNA” and “location” columns respectively.

Using a gene ontology web tool (http://proto.informatics.jax.org/prototypes/GOTools/web-docs/), the genes identified using the three B2 short RNA were investigated for trends in protein function. From this analysis, the potential mRNA targets were observed to be enriched for encoding genes involved in primary metabolic processes (data not shown). It was also striking that all of the target-sites were within the 3´UTR of the mRNA transcripts with multiple target sites identified within some 3´UTR sequence. However, this 3´UTR preference could be explained if these

sequences were derived from the retrotransposition of B2 elements. This is because UTR regions would be more tolerant of transposition events as they would not cause disruption to the mRNA reading frame. To assess this possibility the mRNA were subsequently aligned with the full-length B2 consensus sequence to observe if the target-sites were derived from the insertion of a B2 transposable element. This showed all but one of the mRNA target-sites to contain sequence complementary to the B2 transposon extending beyond the region of short B2 identity to varying degrees. Indicating that the majority if not all of the target-sites identified in the tables above are present due to the transposition of a B2 element into the mRNA sequence. A single 1-21nt target-site identified for the Ddo mRNA had no alignment to the full-length B2 sequence beyond the region of complementarity to the short 1-21nt antisense B2 RNA. This target site would therefore appear not to be derived from B2 insertion into the mRNA, or is derived from a B2 insertion, which has undergone complete divergence but has selectively maintained this region of the B2 sequence.

RT-PCR was used to assess if the mRNA transcripts identified were targeted by the B2 short RNA. However, the PCR amplification (that was to serve as an internal control for RNA quality and efficiency of cDNA synthesis) of the housekeeping genes was not constant. Therefore comparison of the mRNA containing target sites for the short B2 RNA across the different time points of embryonic cell differentiation could not be reliably achieved.

### *5.3.5 Analysis of B2 short RNA using sensor constructs does not confirm targeting ability.*

To further assess if the B2 short RNAs identified were capable of targeting complementary mRNA sequences for destruction, a widely used method for assessing the targeting ability of short RNA was chosen. This involves the insertion of the target sequence into a reporter gene (usually within the 3´ UTR). Such "sensor-constructs" are then transfected into cultured cells that are known to express the short RNA of interest. The targeting ability is then assessed by monitoring any reduction of the reporter, in the presence of the short RNA under study.

Using an easily measurable reporter gene was critical to allowing multiple short RNA to be tested for their ability to target complementary sequence within mRNA transcripts. For this reason the firefly luciferase gene was selected. The firefly luciferase sensor-constructs were designed using DNA oligonucleotides of the B2 short RNA sequences and their reverse complement with an additional four nucleotides (CTAG) at the 5´end of each. The complementary oligonucleotides were then annealed to produce dsDNA with XbaI compatible 5´ overhangs and inserted into the XbaI site in the 3´ UTR of the firefly luciferase gene contained within the pGL3 plasmid (Figure 38). Clones containing the complementary B2 short RNA sequence within the 3´ UTR of the firefly luciferase were then selected by selective PCR across the region of insertion. Sensor-constructs to the micro RNA, Mmu-mir-15 and Mmu-mir-16 were also generated by the same method. These were created to provide a positive control for testing the accessibility of the target site to short RNA. In addition to these, negative controls to all of the sensor-constructs were also designed. For the B2 sensors these contained miss-matches to the short RNA predominantly within the seed-region (2-8nt), which is reported to be critical in target recognition by complementary short RNA (Doench and Sharp, 2004). The miRNA controls used the same sequence inserted in the opposite orientation. See materials and methods for the exact sequences used to create the sensors and controls.

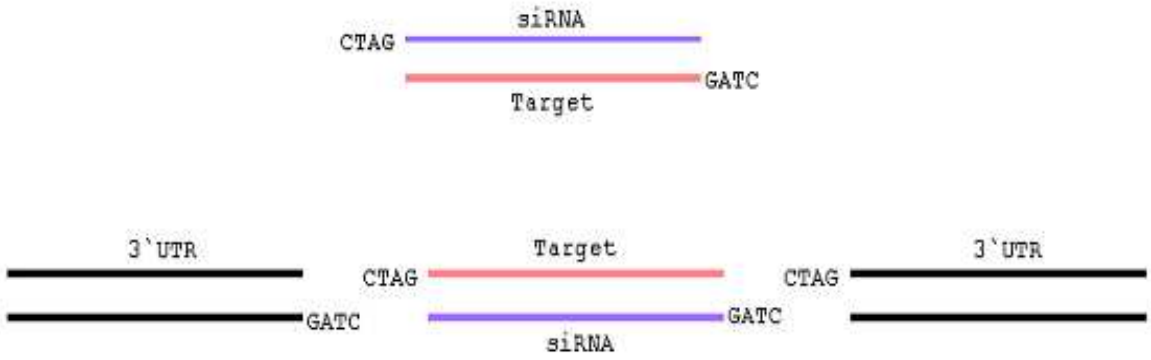


**Figure 38-The design of the target sequence insertion into the sensor construct.** The diagram above displays the siRNA (purple) and complementary target sequence (pink) annealed with CTAG 5´ overhangs. These overhangs are complementary to the GATC 3´ overhand produces by the XbaI endonuclease cleavage in the firefly luciferase 3´ UTR. Ligation of the annealed DNA oligonucleotides in the orientation shown was required for the generation of the sensor construct.

Male and female mouse ES cells were grown in differentiating conditions for 24 hours. The sensor-constructs (B2s 1-23nt, B2as 1-23nt, Mmu-mir-15, Mmu-mir-16 and miss-match controls) were then individually co-transfected with unmodified

renilla luciferase vectors. The co-transfection of renilla luciferase was used to provide an internal control for transfection efficiency. Figure 39 illustrates the specific targeting of the firefly sensor mRNA relative to the renilla luciferase mRNA. Additional transfections of unmodified pGL3 co-transfected with renilla luciferase, individual transfections of each plasmid and a no plasmid control were also included. These transfections provided controls for the specificity of the dual luciferase assay. The transfected cells were lysed after 24 hours (two-days differentiation) and the lysate divided in two aliquots. Half was used to analyse firefly and renilla luciferase expression levels, with the rest used to extract RNA to test for the presence of the B2 short RNA in the cells.

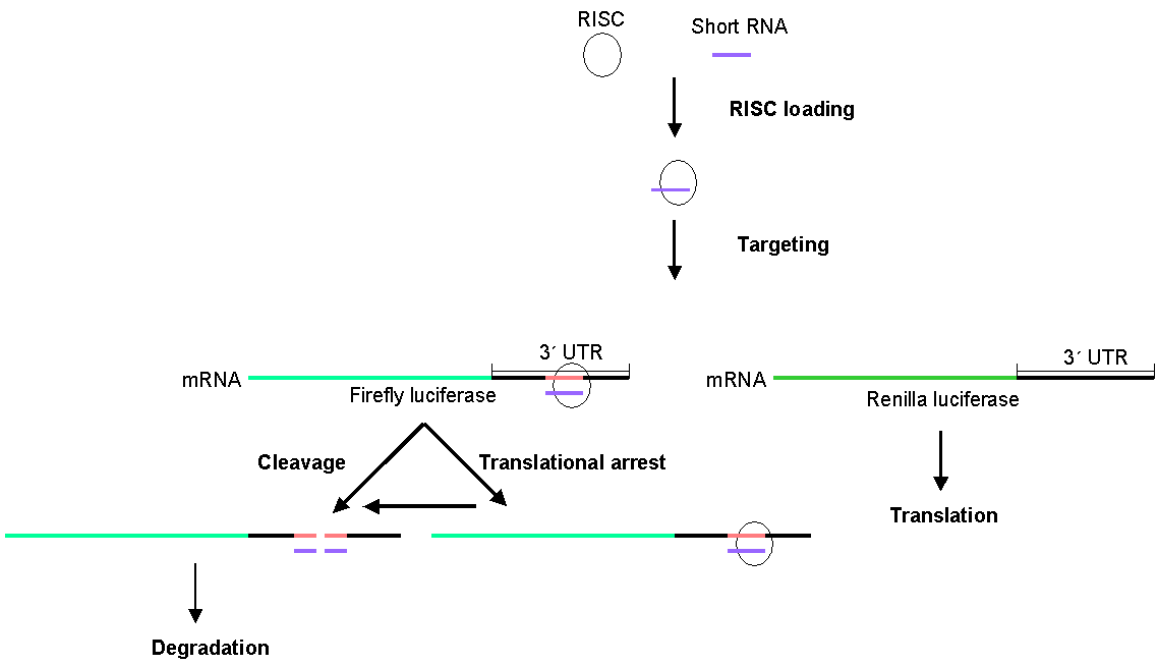


**Figure 39-The use of dual luciferase sensor constructs in short RNA targeting analysis.** The luciferase genes are shown with the firefly luciferase containing the target site (pink) in the 3` UTR. If the short RNA of interest (purple) associates with RISC it may guide the complex to complementary sequences (target). Association of the short RNA guided RISC to the firefly luciferase mRNA containing target sequences would result in the cleavage or translational arrest of the mRNA transcript. Therefore reducing protein levels. However, if the short RNA does not associate with a RISC like complex targeting will no be observed and the firefly luciferase protein levels will remain unaffected. The renilla luciferase mRNA would be unaffected by the short RNA targeting because they do not contain the short RNA target site. These mRNA would therefore be translated and renilla luciferase protein levels would be unaffected. Comparison of the ratios of firefly to renilla luciferase therefore allows any targeting to be observed without the ambiguity of differential transfection between samples.

The cell lysate for each transfection was examined by a single read of firefly and renilla luciferase levels using luminometery. The values obtained from this

analysis were then used to produce a renilla to firefly ratio of the luciferase proteins. Targeting would therefore be observed by an increase of the renilla to firefly ratio relative to the control samples. For the female ES cells there may have been minimal knockdown of the Mmu-mir-15 sensor construct (Figure 40). However, the change compared with its polymorphic control was small and not replicated in subsequent analysis (data not shown). For the other transfections no significant increase in luciferase ratios were observed between the sensor and polymorphic controls. The B2 1-23nt sense and antisense sensors even showed varying degrees of ratio reduction.

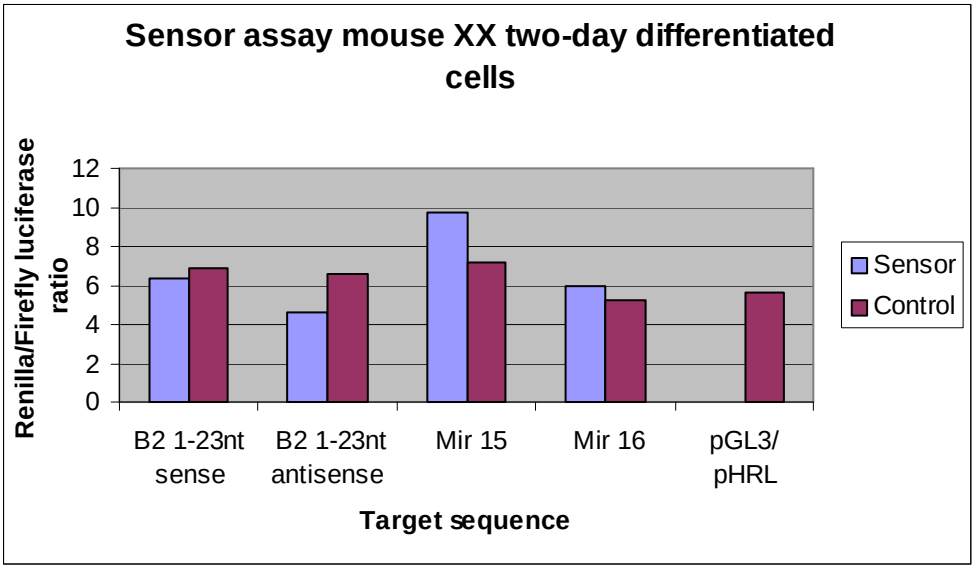


**Figure 40-The renilla/firefly ratios for the B2 sensor constructs in two-day differentiated ES XX cells.** ES XX cells were grown under differentiating conditions for 24 hours before transfection with 10ng pHRL plasmid (renilla) and 90ng of pGL3 (firefly) construct containing the target sequences (sensor) or polymorphic target sequences (control) for the 5´ B2 sense-specific (B2 1-23nt sense), 5´ B2 antisense-specific (B2 1-23nt antisense), Mmu-mir-15 (Mir 15) and Mmu-mir-16 (Mir 16) short RNA. The unmodified pGL3 plasmid was also cotransfected with pHRL plasmid at these concentrations to provide an additional control (pGL3/pHRL). Cells were lysed after 24 hours of transfection with the lysate used for luminometery with the dual luciferase kit to assess levels of renilla and firefly protein. The renilla to firefly ratio was calculated to generate the above graph where sensor constructs target sequence is indicated in the x-axis with the sensor (blue) and their specific polymorphic control sensor (purple) ratios displayed.

The male ES cell sensor transfections also did not display any targeting, with three of the four transfections also showing a decrease in renilla to firefly Ratio (Figure 41). To test if the lack of targeting observed was as a result of the polymorphic controls also being targeted, they were compared to the renilla to firefly ratios of the unmodified pGL3 plasmid. This comparison did show the up-

regulation of renilla to firefly luciferase ratio for the majority of the polymorphic controls (Figure 40 and Figure 41). However, these increases were small and do not appear to indicate targeting of the polymorphic controls as an explanation for the lack of sensor construct targeting observed.

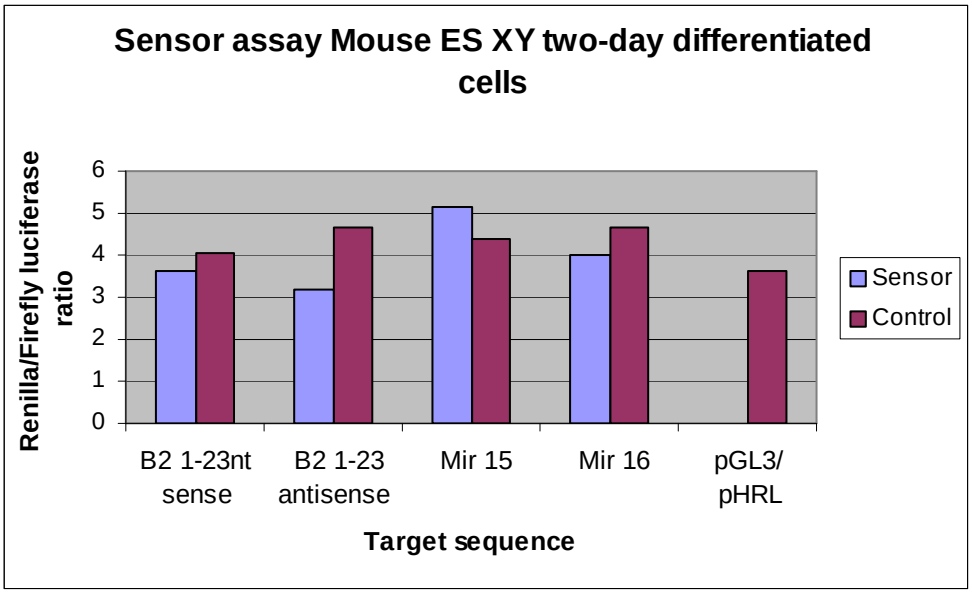


**Figure 41- The renilla/firefly ratios for the transfection of B2 sensor constructs into two-day differentiated ES XY cells.** ES XY cells were grown under differentiating conditions for 24 hours before transfection with 10ng pHRL plasmid (renilla) and 90ng of pGL3 (firefly) construct containing the target sequences (sensor) or polymorphic target sequences (control) for the 5´ B2 sense specific (B2 1-23nt sense), 5´ B2 antisense specific (B2 1-23nt antisense), Mmu-mir-15 (Mir 15) and Mmu-mir-16 (Mir 16) short RNA. The unmodified pGL3 plasmid was also cotransfected with pHRL plasmid at these concentrations to provide an additional control (pGL3/pHRL). Cells were lysed after 24 hours of transfection with the lysate used for luminometery with the dual luciferase kit to assess levels of renilla and firefly protein. The renilla to firefly ratio was calculated to generate the above graph where sensor constructs target sequence is indicated in the x-axis with the sensor (blue) and their specific polymorphic control sensor (purple) ratios displayed.

The RNA extracted after the 24-hour transfection of both male and female ES cells was found to be of poor quality when separated by denaturing-polyacrylamide gel electrophoresis and stained using ethidium bromide. Therefore levels of the short B2 sense 1-22nt, B2 antisense 1-22nt, Mmu-mir-15 and Mmu-mir-16 RNA could not be compared for these sensor transfections. The low quality of the RNA may have arisen from the lysis solution not being sufficiently inhibitory to RNase activity. This could be easily resolved in future experiments by splitting all samples into two aliquots prior to lysis and

extraction each aliquot using dedicated methods for luciferase and RNA respectively.

## 5.4 Discussion

### *5.4.1 B2 sense and antisense short RNA in differentiated embryonic cell lines.*

The initial analysis of embryonic mouse RNA (ES and P19 EC) with full-length B2 sense and antisense-specific probes unexpectedly identified RNA of ~80nt in all samples. In addition to this RNA, there was an indication of shorter RNA in the two-day differentiated samples including a sense RNA of ~20nt (see Figure 33 and Figure 34). This unexpected observation of B2 RNA, significantly smaller than the full-length element (~200nt) prompted further analysis of the B2 transposon.

Using the embryonic mouse short RNA sequence dataset (Mineno et al., 2006) several sequences (sense and antisense) mapping to regions of the B2 transposon were identified (Table 9). Because short RNA expressed in mouse embryos may be similar to those expressed in embryonic stem cells, three short RNA probes were designed to be hybridised to the most abundant sequences identified (1-22nt sense, 1-22nt antisense and 35-58nt sense). The subsequent hybridisation of these short B2 probes identified three defined RNA of ~20nt in both ES (Figure 35 and Figure 36) and P19 EC RNA (data not shown). These short RNA were predominantly observed in two-day differentiated ES and P19 EC samples. Of most interest were the 1-22nt sense and antisense-specific short RNA. This was because they were abundant, formed a hypothetical dsRNA and were derived from promoter region of the B2 transposon. Therefore if capable of targeting through the RNAi pathway they could theoretically be involved in both TGS and PTGS. Like the L1 $T_F$-monomer short antisense RNA, the short B2 RNA identified were previously unreported in the literature.

Subsequent analysis of the B2 RNA focused on the complementary 1-22nt ~20nt RNA. DsRNA of this size have been commonly associated with siRNA in other species, generated from the processing of longer dsRNA (Bernstein et al., 2001, Zhang et al., 2004). Long dsRNA are reported to be generated from the bi-directional transcription of a region and/or the action of RdRP on sense transcripts in other species (Lipardi et al., 2001, Sijen et al., 2001). However, to date no mammalian homologue of RdRP has been identified and few endogenous

mammalian siRNA have been proposed. This suggests that the potential production of dsRNA in mammals maybe limited to fortuitous bi-directional transcription. Because B2 elements are interspersed throughout the genome, there is significant potential for the production of antisense transcripts from the non-specific read-through transcription of antisense-orientated B2 elements within transcribed regions of the genome (Dahary et al., 2005). In addition to the fortuitous production of antisense B2 transcripts, B2 transposons are also capable of the bi-directional transcription of individual elements. This is because B2 transposons contain an active internal RNA polymerase II promoter in antisense orientation to the tRNA derived RNA polymerase III promoter (Ferrigno et al., 2001). Therefore transcription from both the RNA polymerase II and III promoters would result in the partial bi-directional transcription of the 1-133nt region of the B2 transposon. The possibility of the B2 short being derived from this specific bi-directional transcription is supported by recent observations of bi-directional transcription of the human L1 promoter associating with the production of siRNA (Yang and Kazazian, 2006).

In addition to the complementary ~20nt RNA observed with the sense and antisense-specific 1-22nt B2 probes several other RNA were observed (~30-90nt). The majority of these RNA were identified with the B2 1-22nt sense-specific probe and were found to be up-regulated with the induction of the ~20nt RNA in the two-day differentiated samples (Figure 35). However, RNA of ~90nt (resembling the ~80nt RNA observed with the full-length B2 probes (Figure 33)) were observed to be specifically down-regulated in the two-day differentiated samples with both sense and antisense-specific probes. The two ~90nt B2 RNA are known to have up to 22nt of complementary sequence capable of forming dsRNA. If the region of complementarity between these sequences extends beyond this region these sequences may form a long dsRNA, creating a potential dsRNA precursor which could be recognised by Dicer and other dsRNA specific RNase to generate the ~20nt RNA detected.

This hypothesis is supported by the reciprocal expression pattern of the ~90nt and ~20nt RNA and the appearance of potential intermediate RNA (B2 1-22nt sense-specific hybridisation) in the two-day differentiated samples. Antisense-specific intermediate RNA may also be expected from this type of processing. However, their absence could be explained by the preferential destruction of

the antisense strand. A potential pitfall of this hypothesis is that dsRNA are widely reported to induce the interferon response in mammalian cells. A dsRNA of ~90nt in length would theoretically be capable of inducing an interferon response. However, undifferentiated embryonic stem cells have been reported not to elicit an interferon response to long dsRNA (Yang et al., 2001). Therefore the potential B2 ~90nt dsRNA may have avoided triggering an interferon response through their endogenous expression in ES cells, not sufficiently differentiated to elicit an interferon response to dsRNA. In addition other factors, such as the cellular localisation of dsRNA also play a critical role in a cells ability to induce an interferon response.

In order to establish if the RNAi pathway was involved in the processing of the 1-22nt B2-specific ~20nt RNA, Dicer-null ES cells were differentiated. From this it was established that the accumulation of the ~20nt sense-specific 1-22nt B2 RNA was substantially reduced when compared with differentiated control ES cells (Figure 37B). In addition to this the ~90nt RNA in the two-day differentiated Dicer-null sample was unchanged and there was a reduced accumulation of the intermediate RNA. This partially supports the hypothesis of Dicer-dependant processing of the ~90nt RNA to generate the intermediate and ~20nt B2 sense-specific RNA. However, the complete absence of Dicer activity was confirmed by the elimination of miRNA production (Figure 37C) and so minimal Dicer activity is not thought to explain the low levels of short B2 RNA observed. Therefore the ability of the Dicer-null ES cells to generate some B2 sense-specific intermediate and ~20nt RNA suggests that Dicer is not exclusively involved in this process. Instead an alternative Dicer-independent low level processing of the short RNA is suggested to exist specifically in the two-day differentiated sample (potentially the action of other RNase). However, these results could also be accounted for by a non-specific secondary effect of the Dicer knockout, which affects the processing of the B2 short RNA. Further analysis of B2 RNA short RNA processing is required to eluted what role Dicer plays in this.

The potential targeting by the three ~20nt B2 RNA identified could have major implications in cellular gene regulation, though the repression of B2 transposon expression and mRNA transcripts containing complementary sequences. Therefore their potential recruitment into the RNAi pathway had to be established. There were many potentially targetable mRNA sequences identified

containing complementary sequences due to the insertion of B2 transposable elements (Table 10,Table 11 and Table 12). The semi-quantitative RT-PCR analysis of these mRNA sequences at the different time-points of embryonic cell differentiation was chosen as a simple method of indicating any targeting by the B2 short RNA. However, across the embryonic time course the PCR amplification of several housekeeping genes (that were to serve as an internal control for RNA quality and cDNA synthesis) were not observed to be constant and so reliable measurement could not be made by this method. Therefore an alternative approach involving the insertion of targetable sequences (complementary to the short B2 sequences identified) into the 3` UTR of firefly luciferase reporter genes was used. This approach did not indicate any targeting by the B2 short RNA sequences. However, targeting was also not observed for the positive controls containing Mmu-mir-15 and Mmu-mir-16 target sequences, thus the sensors may not be sufficiently sensitive. Therefore further analysis is also required to establish if the short B2 RNA are capable of targeting complementary sequences for repression like siRNA and miRNA.

# 6

# Consistent processing of 5S rRNA and accumulation of small (~20nt) RNA in human and mouse cell lines and human tissue

## Results 4

## 6.1 Summary

Chapters 4 and 5 have described transposon related short RNA identified during the differentiation of embryonic stem (ES) and embryonic carcinoma (EC) cell lines. These investigations also serendipitously uncovered previously unreported and unanticipated short 5S ribosomal RNAs (rRNA).

This chapter focuses on the discrete 5S rRNA short RNA ~19-26nt in length. These were originally identified in differentiated embryonic mouse and human cell lines. Subsequent analysis of various cancer cell lines and human breast tissue samples also identified short 5S rRNA.

Mapping of short 5S rRNA to regions of the full-length transcript was carried out using mouse ES RNA samples. This identified two short sense and one antisense 5S rRNA to be derived from the extreme 5` (~22nt), 3` (~26nt) and central region (~26nt) of the full-length 5S rRNA sequence respectively. Subsequent analysis of the unfiltered embryonic mouse RNA sequence dataset previously described confirmed highly abundant sense-specific short 5S rRNA to be derived from these regions. However, no antisense-specific sequences were identified in the embryonic mouse RNA data. To assess if this was a general feature of rRNA processing the embryonic mouse RNA dataset was used to identify highly abundant short 18S rRNA and 28S rRNA sequences. From this a single short ~19nt 28S rRNA was identified to be expressed at all points of the mouse ES differentiation time course.

The involvement of Dicer in the processing of short 5S rRNA transcripts was also assessed. This demonstrated that *in vitro* digestion of the intermediate 5S rRNA with human recombinant Dicer produced short 5S rRNA. However, the differentiation of Dicer-null ES cells did not completely eliminate the accumulation of short 5S rRNA implying incomplete Dicer dependence. Complementary sequences to short 5S rRNA were then embedded into the 3`UTR of a firefly luciferase reporter construct to assess the targeting ability of the short 5S rRNA, however no targeting was observed.

# 6.2 Introduction

## *6.2.1 5S rRNA and derived transposons*

5S rRNA is a component of the large subunit of all cytoplasmic and most organellar ribosomes. In eukaryotic organisms the ribosome is composed of two subunits: the large 60S consisting of 5S, 28S and 5.8S rRNAs and the small 40S subunit containing 18S rRNA (Szymanski et al., 2003). A number of proteins are also associated with each of the ribosomal subunits, making it functional. The precise role of 5S rRNA within the ribosome is unclear. However, it has been speculated to influence large subunit stability, because 5S rRNA (released by nuclease cleavage) coincides with a proportional decrease in transcriptional activity (Holmberg and Nygard, 2000).

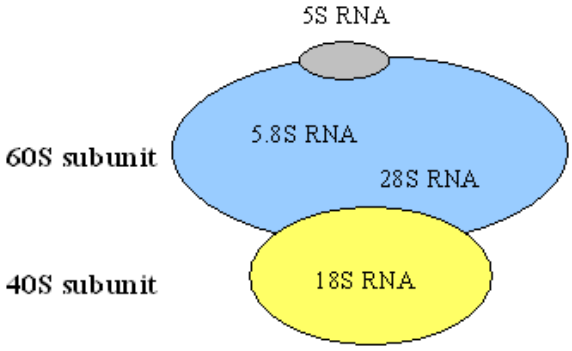


**Figure 42- Eukaryotic rRNA arrangement within the ribosome.** The above image illustrates the rRNA positioning within eukaryotic ribosomal subunits. This shows the 5.8S and 28S rRNAs associating with the large 60S subunit along with 5S rRNA and the 18S rRNA associating with the small 40S subunit.

The secondary structure of ribosome bound 5S rRNA is highly conserved among eukaryotes, despite significant sequence variation between species as summarised in Figure 43 (Szymanski et al., 2002). However, this structure may differ when 5S rRNA is unbound or is associating with other proteins.

The 5S rRNA gene is encoded by a repeating-unit containing a ~120nt RNA polymerase III transcribed region. The size of the repeat-unit varies between species, however the size of the transcribed region remains constant. The consensus sequence of the 5S rRNA transcript is also highly conserved and is identical between human and mouse (Suzuki et al., 1994). However, the spacer regions diverge greatly between the organisms with the exception of the D-box. This is a promoter element located ~30nt upstream of the transcriptional start-site (Nielsen et al., 1993).

**Figure 43- Eukaryotic 5S rRNA secondary structure.** Image taken from (Szymanski et al., 2002), generated from the alignment of eukatyotic 5S rRNA sequences showing common secondary structure and any sequence variation present. The nomenclature used was N: A, C, G or U; R: A or G; W: A or U; S: G or C; M: A or C; K: G or U. Black squares represent rare insertions.

In mouse, 5S rRNA gene copies (~1.6kb) are arranged in a head-to-tail tandem-array at a single genomic location on chromosome 8 (Suzuki et al., 1994). The majority of human 5S rRNA genes (~2.3kb) are also found in tandem on chromosome 1. However, an additional smaller second cluster of 5-10 copies (1.6kb) has been identified at a separate genomic location (Sorensen and Frederiksen, 1991). In both mouse and human there are ~100 copies of the 5S rRNA gene present in the main tandem repeat cluster (Drouin, 2000, Sorensen and Frederiksen, 1991).

In addition to 5S rRNA gene copies, there are many other 5S rRNA copies dispersed throughout the genome. These dispersed 5S rRNA copies are not known to associate with the ribosome. The term pseudogene originated to describe protein-coding sequence related to known genes, which had lost their ability to encode a functional protein. Although the functional contribution of a non-coding RNA can be less easily determined, the term could be equally well used to describe non-coding RNA that have lost their functional ability. Therefore for the purposes of this thesis the dispersed 5S rRNA genes, which are reported to have limited or no transcriptional activity will be referred to as pseudogenes. The reduced transcriptional activity of the 5S rRNA pseudogenes is often attributed to sequence divergence or the absence of the D-box promoter located within the 5S rRNA spacer sequence. The absence of the D-box promoter alone is reported to result in a ten-fold reduction in 5S rRNA transcription (Nielsen et al.,

1993). Some of the pseudogenes are also thought to have originated through transposition because they are flanked by inverted repeats (Drouin, 2000), which is a common feature of retrotransposition.

## 6.3 *Results*

### *6.3.1 The identification of novel short 5S rRNA in cultured mouse embryonic cells induced to undergo differentiation.*

Non-coding RNA such as tRNA and rRNAs are widely used as indicators of equal loading and RNA quality in gel electrophoresis. This is because of their high abundance and ubiquitous transcription. The ethidium bromide staining of RNA within polyacrylamide gels does provide some control for equal loading of samples. However, verification of equal transfer to a northern blot requires the hybridisation of cross-linked RNA to a probe specific for an internal control RNA. The internal control of 5S rRNA was selected for use because its transcript size of 121nt would not obscure the region under study (<100nt). Hybridisation of the 5S rRNA "sense-specific" probe (i.e. will hybridise to a sense transcript) was expected to produce a single band of ~120nt with an intensity proportional to the concentration of RNA loaded in each lane of the northern blot.

RNA extracted from differentiating female mouse ES cells was hybridised to the full-length 5S rRNA sense-specific probe. Unexpectedly, discrete RNA of ~24nt and ~26nt were observed, which closely resemble miRNA and piRNA in size respectively (Aravin et al., 2006, Bartel, 2004, Kanellopoulou et al., 2005). Like the short transposon-associated RNA, these short 5S rRNA were only observed in the two-day differentiated sample (Figure 44B). In addition to these, several other intermediate RNA (~40-100nt in size) hybridising with the 5S rRNA sense-specific probe were unexpectedly identified predominantly in the two-day differentiated sample. The probe was then removed from the RNA (method 2) and the removal was verified by phosphor-imaging. The RNA was then hybridised to the Mmu-mir-292as probe to assess the loading and accessibility of the short RNA. Mmu-mir-292as was detected at comparable levels in all of the ES samples (Figure 44C).

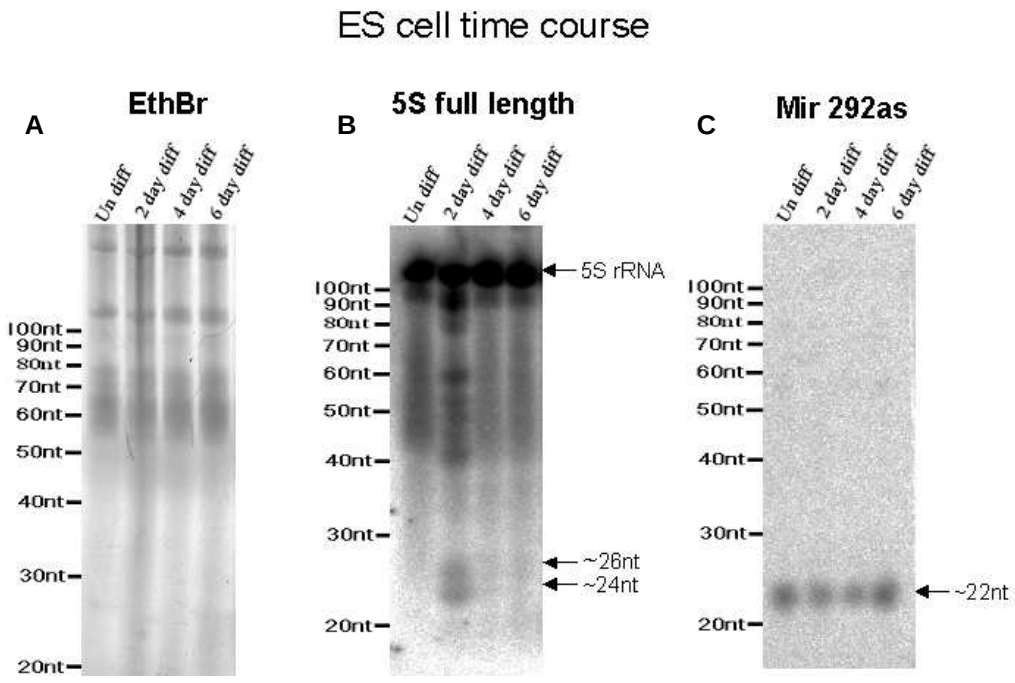


**Figure 44- 5S rRNA short RNA in mouse embryonic stem cells.** RNA was extracted from undifferentiated (un diff), two-day (2day diff), four-day (4 day diff) and six-day (6 day diff) differentiated female mouse ES cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 10µg/well of each sample was then loaded and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gels were stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the full-length 5S sense-specific probe (B) and Mmu-mir-292as (C) probes at 40℃, wash 0.2X SSC  0.2% SDS 40℃. The radioactive signals were recorded by phosphor-imaging with one-hour exposures shown above. Bands for the full-length 5S rRNA and those of similar size to miRNA are indicated with arrows (right) with marker size annotated on the left of images.

To test whether this accumulation of short 5S rRNA was unique to ES cells or was a more general process, RNA extracted from differentiating mouse P19 EC cells was assessed. Hybridisation of the full-length 5S rRNA sense-specific probe also identified the accumulation of short ~19nt and ~26nt RNA with several intermediate RNA (~40-100nt in size) in the two-day differentiated samples of P19 EC cells (Figure 45B). The 5S rRNA probe was then removed (method 2) and the removal was verified by phosphor-imaging. The RNA was subsequently hybridised to the Mmu-mir-21 probe to assess the loading and accessibility of the short RNA. Mmu-mir-21 was detected to be slightly under-represented in the two and four-day differentiated P19 EC samples (Figure 45C).

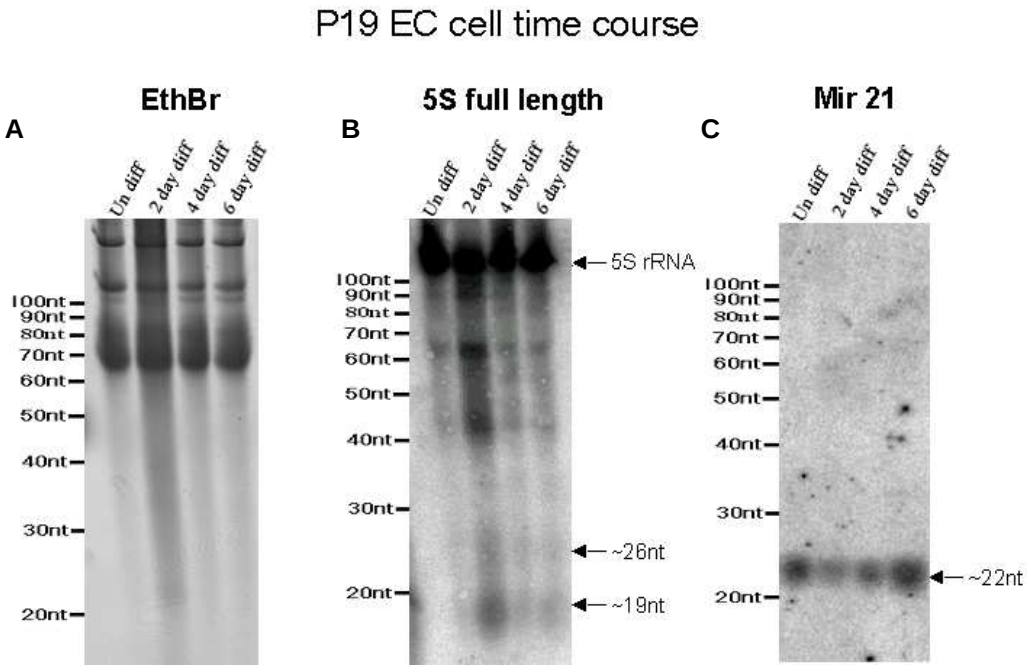


**Figure 45- 5S rRNA short RNA in mouse P19 embryonic carcinoma cells.** RNA was extracted from undifferentiated (un diff), two-day (2day diff), four-day (4 day diff) and six-day (6 day diff) differentiated mouse P19 EC cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 20μg/well of each sample was then loaded and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gels were stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the full-length 5S sense-specific probe (B) and Mmu-mir-21 (C) probes at 40℃, wash 0.2X SSC 0. 2% SDS 40℃. The radioactive signals were recorded by phosphor-imaging with one-hour exposures shown above. Bands for the full-length 5S rRNA and those of similar size to miRNA are indicated with arrows (right) with marker size annotated on the left of images.

Several RNA extractions from differentiating P19 and EC cells were consistently observed to have a smeared appearance in the two-day sample. Examples of this are shown in Figure 44A and Figure 45A. There was no evidence of increased cell death in culture at this time point (which might result in increased generalised RNA degradation). This was assessed by the absence of "floating" cells detached from the culturing surface. However, no specific measurements of apoptotic or necrotic markers have been made to assess cell viability. The smeared appearance of these samples may therefore be due to a general processing of the cellular RNA. However, it is not known if this would affect all of the RNA in some cells or some of the RNA in all of the cells.

### 6.3.2 Short 5S rRNA were identified during the induction of differentiation in the human embryonic carcinoma cell line NT2.

The conservation of siRNA and miRNA sequence and expression between species can indicate function (Sempere et al., 2004). To investigate if short 5S rRNA could also be identified in human, the embryonic carcinoma cell line NT2 (NT2 from this point forward) was selected. These cells were selected because they are similar in terms of their origin to the mouse embryonic cells previously found to express 5S rRNA (Figure 45B).

NT2 cells were grown as standard and differentiation was induced by the addition of retinoic acid. Undifferentiated cells were grown in parallel for comparison. Total RNA was extracted from undifferentiated, two, four and six-day differentiated cells and the concentration measured by $UV^{260}$ light absorbance. The NT2 RNA samples were then separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide. The ethidium bromide stain of NT2 RNA (Figure 46A), in contrast to RNA from the mouse embryonic cell time courses, (Figure 44A) displayed no differential smearing at any time point. The RNA was then electro-blotted onto nylon membrane and cross-linked using EDC.

Because the human and mouse consensus sequence for 5S rRNA are identical, the same full-length 5S rRNA sense-specific probe was used for the hybridisation of NT2 RNA. This identified short 5S rRNA of ~22nt predominantly in the two-day differentiated sample (Figure 46B). Lower levels of the ~22nt 5S rRNA could also be observed in the four and six day differentiated samples. Intermediate 5S rRNA (~30-70nt in size) were also observed in all samples. However, the size distribution of these intermediate RNA was significantly altered in the two-day differentiated sample when compared to later time points and undifferentiated samples. This included the disappearance of the ~30nt RNA at the same time point as the induction of the ~22nt 5S rRNA.

The 5S rRNA probe was then removed (method 2) and the removal was verified by phosphor-imaging. The RNA was then hybridised to the Hsa-mir-21 probe to

assess the loading and accessibility of the short RNA. Hsa-mir-21 was detected at comparable level in all samples (Figure 46C).

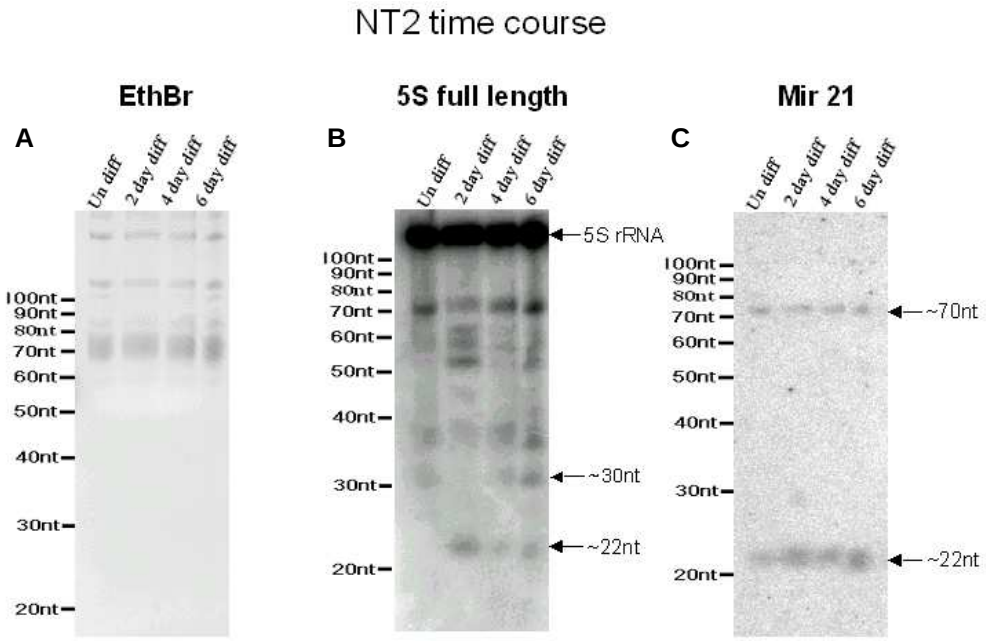


**Figure 46- 5S rRNA short RNA in human NT2 cell differentiation.** RNA was extracted from undifferentiated (un diff), two-day (2day diff), four-day (4 day diff) and six-day (6 day diff) differentiated human NT2 cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 5µg/well of each sample was then loaded and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the full-length 5S sense-specific (B) and Mmu-mir-21 (C) probes at 40℃, wash 0.2X SSC 0.2% S DS 40℃. The radioactive signals were recorded by phosphor-imaging with 16 hour exposures shown above. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

### 6.3.3 Determining the region of 5S rRNA from which short RNA are derived by oligonucleotide mapping.

Short 5S rRNA were observed during the differentiation of mouse (Figure 44 and Figure 45) and human (Figure 46) embryonic cell lines. However, it was unknown which regions of the 5S rRNA transcript were generating the short 5S rRNA. Knowledge of the positioning of the short RNA relative to the full-length transcript could indicate if specific (as is observed in siRNA and miRNA processing) or generalised processing is responsible for the short 5S rRNA production.

Mapping of the short 5S rRNA was carried out using DNA oligonucleotides as probes to specifically hybridise with either sense (sense-specific) or antisense (antisense-specific) 5S rRNA transcripts (Figure 47). For this, eight oligonucleotide probes (25nt in length with the exception of the final oligonucleotide) were designed to hybridise across the entire 121nt 5S rRNA consensus transcript. A 10nt overlap was introduced between adjacent oligonucleotides to reduce the effect of partial probe hybridisation on the hybridisation efficiency of the short 5S rRNA. The exact sequences of probes used can be found in the materials and methods section (Table 3).

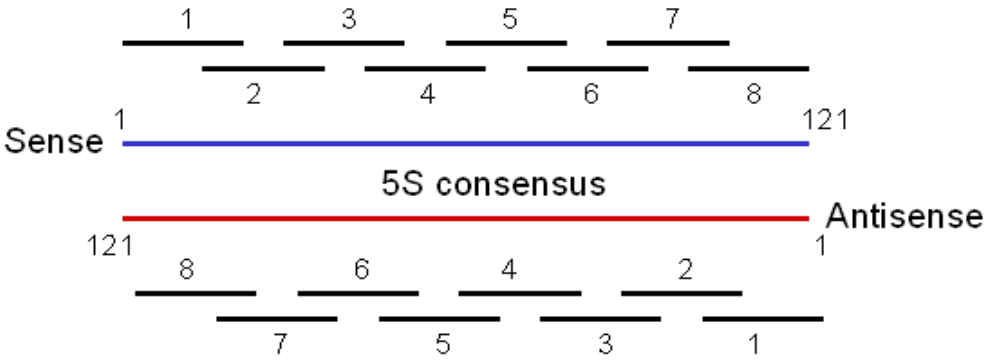


**Figure 47- Overlapping oligonucleotide design.** The 5S rRNA transcript is represented above (not to scale) with blue (sense) and red (antisense) strands. The oligonucleotide probes numbered 1-8, the black lines indicate their region of hybridisation relative to the consensus sequence.

Total RNA was extracted from undifferentiated and two-day differentiated mouse ES cells. Equal amounts of each sample were loaded on to a denaturing-polyacrylamide gel in triplicate before size separation by gel electrophoresis. The RNA was subsequently stained with ethidium bromide (Figure 48A) prior to electro-blotting and cross-linking to nylon membrane using EDC. The northern

blot was then cut into three pieces. Each piece containing a single lane of undifferentiated and two-day differentiated ES cell RNA.

The sense-specific oligonucleotide probes were split into three pools 1-2, 3-5 and 6-8 (see Figure 47 for positioning). These pools were then radioactively end-labelled using $^{32}$P-γATP as described in the materials and methods. Figure 48B shows the hybridisation of the three pools of oligonucleotide probe to the ES RNA. This identified two short RNA from different regions of the 5S transcript one of ~22nt (pool 1) and another of ~26nt (pool 3).

The individual oligonucleotide probes for pools one (Figure 48C) and three (Figure 48D) were then radioactively end-labelled using $^{32}$P-γATP. These were consecutively hybridised to the ES RNA with probe removal between hybridisations (method 1). The pre-hybridisation phosphor-image confirming the effective removal of probe hybridised to the region of interest (~19-30nt) are shown alongside each hybridisation. These results show the 5´of the 5S rRNA transcript corresponds to a ~22nt RNA (5S sense-specific probes (1) 1-16nt & (2) 6-31nt) with the RNA of ~26nt mapping to the extreme 3´of the transcript (5S sense-specific probe (8) 95-121nt). The position of the mapped short 5S rRNA relative to the folded full-length 5S rRNA is also shown in Figure 48E. Other intermediate RNA (~38-100nt) hybridised with the two discrete short 5S rRNA were also identified. These may indicate possible precursors to the processing of the short RNA.

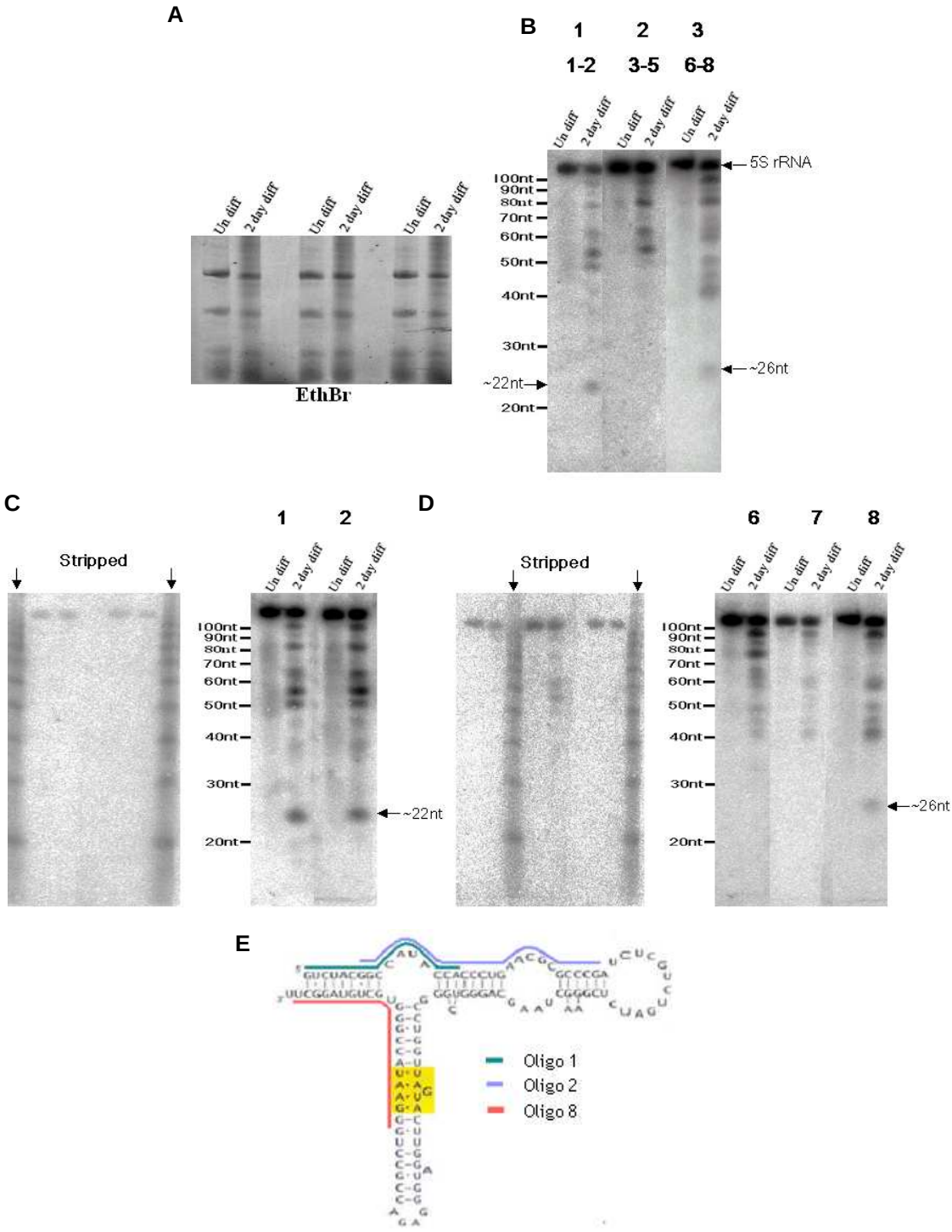


**Figure 48- Sense specific 5S rRNA mapping.** RNA was extracted from undifferentiated (un diff) and two-day (2day diff) differentiated mouse ES cells. The concentration of the RNA was measured by $UV^{260}$ light absorbance. 20µg/well of each sample was loaded and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide (A) before electro-blotting and cross-linking to nylon membrane using EDC. The RNA samples were then hybridised in 2x SSC and 2% SDS buffer to three (1-2, 3-5, 6-8) 5S sense-specific oligonucleotide probe pools (B) and the individual oligonucleotide probes of pool 1 (C) and 3 (D) at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive si gnals were recorded by phosphor-imaging. One-hour exposures are displayed for all hybridisations with bands of interest indicated on the right and the marker on the left. The position of oligonucleotide probe hybridisation relative to the folded 5S rRNA consensus is shown in (E).

The sense-specific oligonucleotide probes were removed from the ES cell RNA (method 1). DNA oligonucleotide probes specific for antisense 5S rRNA, were then used to identify short antisense 5S rRNA complementary to those observed in Figure 48. The pre-hybridisation phosphor-images confirming the effective removal of probe hybridised to the region of interest (~19-30nt) are shown alongside each hybridisation.

The antisense-specific oligonucleotide probes were separated into three pools 1-3, 4-6 and 7-8 (see Figure 47 for positioning). These pools were radioactively end-labelled using $^{32}$P-γATP as described in the materials and methods. Figure 49A shows the hybridisation of the three oligonucleotide probe pools to the ES RNA. This identified a single antisense ~26nt RNA (pool 2) in the two-day differentiated sample. Probes were then removed (method 1) before hybridisation of the ES cell RNA to the individual oligonucleotide probes of pool 2. Figure 49B shows that the antisense RNA of ~26nt hybridises predominantly with probe 5 (antisense 5S rRNA 36-61nt). This maps the RNA to the antisense of the 5S rRNA right arm of the hairpin loop as indicated in Figure 49C. Therefore no overlap between the two short, sense 5S rRNA and the short antisense 5S rRNA was observed.

Figure 49 also shows significant hybridisation to a band identical in size to the full-length transcript with oligonucleotide probe 5 (36-61nt) and to a lesser extent with the other antisense-specific probes. This was unexpected as it was assumed that any antisense-specific transcripts would be as a result of read through transcription and result in hetero disperse transcripts. Precisely sized molecules, particularly exactly the same size as the sense RNA were not expected. The possibility of cross-hybridisation to the 5S rRNA sense transcript by the antisense-specific oligonucleotide probe 5 was investigated using the BLAST (bl2seq) alignment tool. This showed there to be no regions of extensive complementarity to the sense transcript. Manual alignment of the alignment probe to the sequence base paired in the reported 5S rRNA secondary structure was then carried out discounting non-canonical base pairs. This showed there to be minimal hybridisation potential (possible 11 of 17 bases paired with mismatch and gaps) between the sequences (Figure 49D).

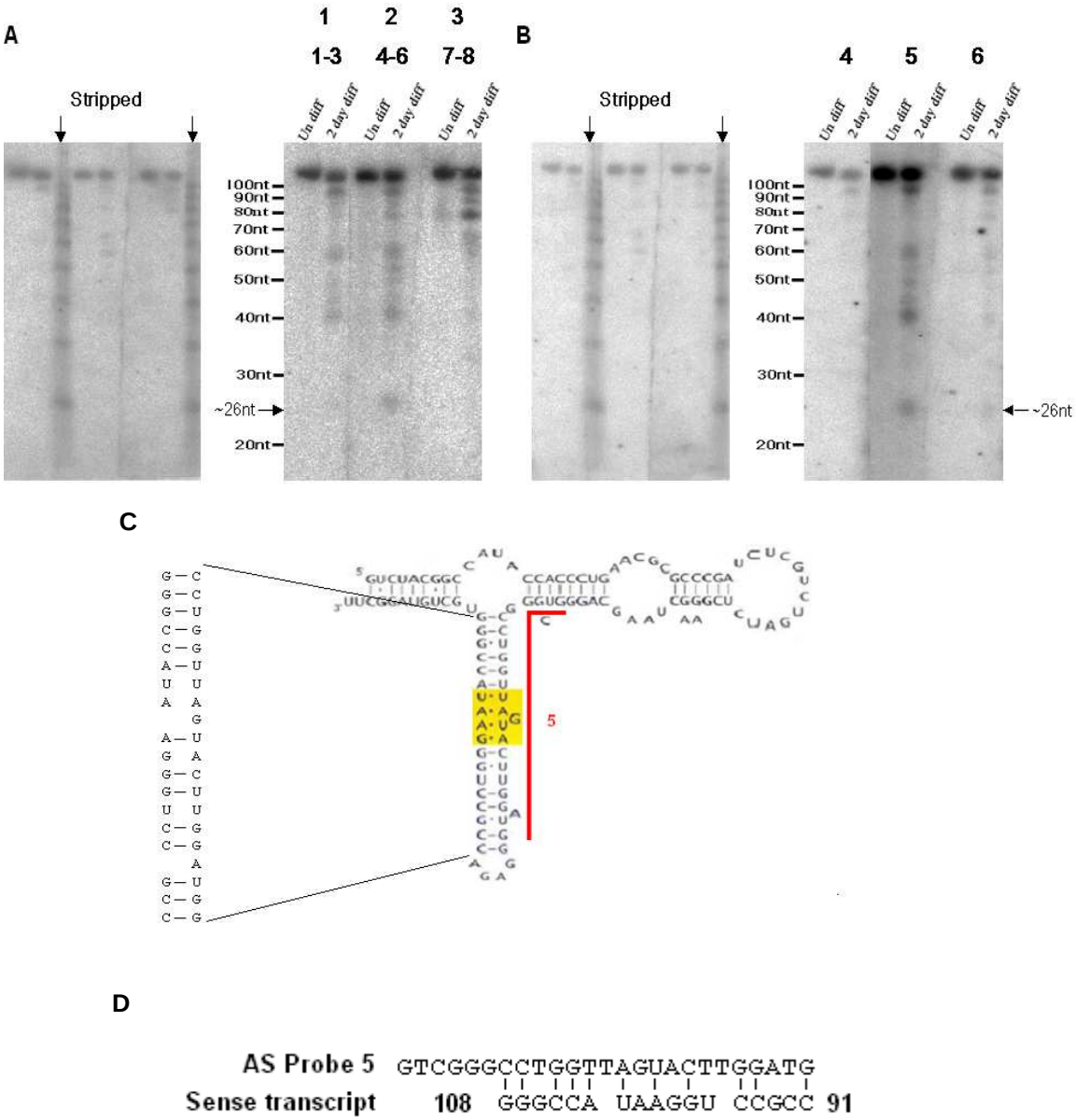


**Figure 49-Antisense-specific short RNA mapping.** Northern blots for undifferentiated (un diff) and two-day differentiation (2 day diff) mouse ES cell RNA produced in Figure 48 were stripped. They were then hybridised in 2x SSC and 2% SDS buffer to the three (1-3, 4-6, 7-8) antisense-specific 5S rRNA oligonucleotide probe pools (A) and the individual oligonucleotide probes in pool 2 (B) at 40℃, wash 0.2X SSC 0.2% SDS 40℃.The radio active signals were recorded by phosphor-imaging. One-hour exposures are displayed for all hybridisations with bands of interest indicated on the right and the marker on the left. The sequence of the antisense-specific oligonucleotide probe 5 relative to the folded 5S rRNA sequence is shown along with nucleotide alignment of the hairpin minus the non-canonical base pairing(C). The best alignment of the antisense oligonucleotide 5 sequence to the region of the 5S rRNA sense transcript (91-108nt) which forms the left arm of the hairpin in the secondary structure (D).

### *6.3.4 Full-length 5S rRNA antisense transcripts*

The identification of a ~120nt RNA transcript hybridising to an antisense-specific 5S rRNA probe in both samples was of interest, because this may indicate the presence of full-length antisense 5S rRNA transcripts. The most likely explanation of a genuine antisense-specific 5S rRNA, identical in size to the sense 5S rRNA would be from RNA dependant RNA polymerase (RdRp) transcription using the sense 5S rRNA as a template. However, there is no reported RdRp in mammals. Alternatively, because this signal was observed after successive hybridisations of the RNA to sense-specific 5S rRNA probes, it may have been as a result of hybridisation to sense-specific probe i.e. antisense sequence that had not been removed by stripping (see Figure 49).

To investigate this possibility, a similar experiment to that of Figure 49 was carried out with a freshly prepared blot. Equal amounts of female ES RNA extracted from undifferentiated, two-day and four-day differentiated cells were loaded onto a denaturing-polyacrylamide gel. The RNA was subsequently stained with ethidium bromide prior to electro-blotting and cross-linking onto nylon membrane. The RNA was then hybridised to the 5S rRNA antisense-specific oligonucleotide 5 probe that had been end labelled using $^{32}$P-γATP. This hybridisation again identified an RNA of ~120nt in all three samples (data not shown). Therefore confirming the original ~120nt signal observed to be due to direct hybridisation to the ES cell RNA and not any residual complementary probe.

To further assess the extent of hybridisation to the ~120nt antisense 5S rRNA identified, two fresh blots, identical to that described above were hybridised separately to the 5S rRNA antisense-specific oligonucleotide probe (Figure 50A) and the reverse complement of the 5S rRNA antisense-specific oligonucleotide probe (Figure 50B). These filters were then subjected to a series of washes of increasing stringency. The radioactive signals were recorded for five minutes after each wash by phosphor-imaging. The radioactive signals were then quantified using imaging software to assess the extent of signal reduction after each wash. For this, areas containing the ~120nt band at the three time points were selected. After subtraction of the generalised background signal calculated from an area of equal size without specific signal, the value of 100% was

attributed to each of the ~120nt bands observed at 40ºC. The percentage of starting signal observed after each subsequent wash was then calculated for the undifferentiated, two-day and four-day differentiated samples separately. The average of these values was then used to generate a graph (Figure 50C) indicating the percentage of the original 40ºC signal remaining after each wash step.

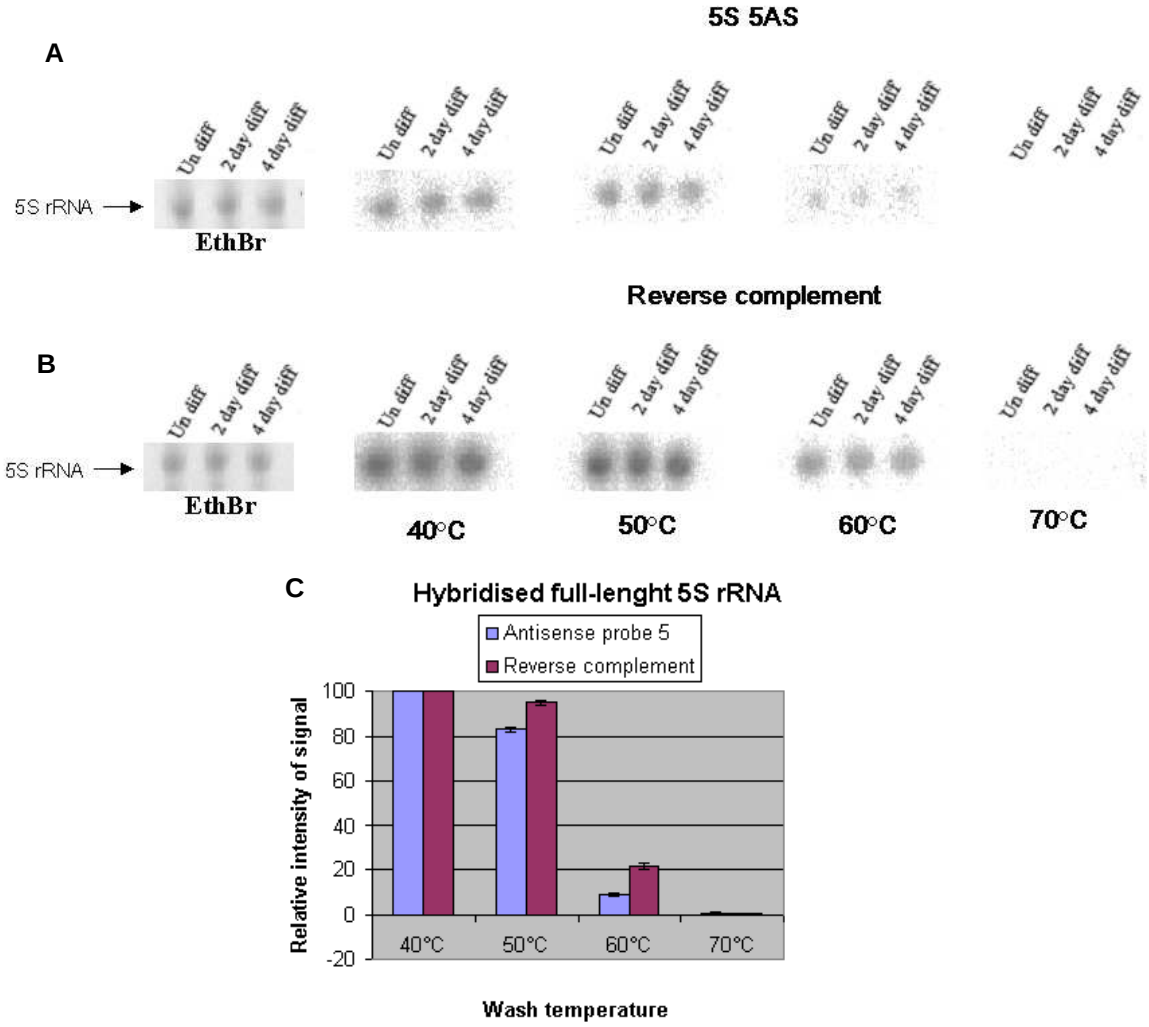


**Figure 50- Hybridisation of the ~120nt antisense 5S rRNA.** RNA was extracted from undifferentiated (un diff), two-day (2day diff) and four-day (4 day diff) differentiated female mouse ES cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 10µg/well of each sample was then loaded in duplicate and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide before being electro-blotted and cross-linked to nylon membrane using EDC. The membrane was cut to contain a single lane of each RNA sample, then hybridised in 2X SSC and 2% SDS buffer to the 5SrRNA antisense-specific oligonucleotide probe 5 (A) and its reverse complement (B) at 40℃. The filter was then subjected to a series of washes at 40℃, 50° C, 60℃ and 70℃. The radioactive signals were recorded after each wash by phosphor-imaging for five minutes. These images have been cropped to show the ~120nt antisense-specific and sense-specific 5S rRNA and are shown above. Using the imaging software the RNA bands were quantified and used to generate a graph showing the comparative decrease in signal with increasing stringencies of wash (C), sd are also included.

Assuming that the sense-specific reverse complement of the 5S rRNA antisense-specific oligonucleotide probe is 100% complementary to the majority of the full-length 5S rRNA transcripts, these results indicate that there is significant complementarity between the antisense-specific probe and a ~120nt RNA. The slightly lower levels of probe retention with the antisense-specific probe may be explained by a lower proportion of the ~120nt RNA to having 100% complementary to the probe or slight miss-match between the sequences. These results therefore support the presence of a ~120nt RNA containing perfect or near perfect antisense sequence and not a highly miss-matched cross-hybridisation of sense 5S rRNA.

Having indicated the ~120nt RNA not to be as a result of cross-hybridisation to the sense-specific DNA oligonucleotides, we wished to assess if the RNA represented a full-length antisense 5S rRNA transcript. To investigate this possibility equal amounts of female ES RNA extracted from undifferentiated, two-day and four-day differentiated cells were loaded in triplicate onto a denaturing-polyacrylamide gel. The RNA was subsequently stained with ethidium bromide (Figure 51A) prior to electro-blotting and cross-linking onto nylon membrane. The northern blot was then cut into three pieces, each containing a single lane of undifferentiated, two-day and four-day differentiated ES RNA. DNA oligonucleotide probes specific to the extreme 5` (1) and 3`(8) of the theoretical 5S rRNA antisense transcript were radioactively end-labelled using $^{32}$P-$\gamma$ATP along with antisense-specific oligonucleotide 5. Each probe was then individually hybridised to one of the northern blot filters. This clearly identified the ~120nt RNA using the 5S rRNA antisense-specific oligonucleotide 5 probe. However, no signal was observed with the oligonucleotide 8 probe and only a very faint ~120nt signal was observed with the oligonucleotide 1 probe with longer exposure times (data not shown). Therefore it would appear that although the region of similarity between the ~120nt RNA and the antisense 5S rRNA sequence may extend beyond the region of 5S rRNA antisense-specific oligonucleotide 5 hybridisation, these sequences are more divergent from the 5S rRNA consensus used. There will only be definitive confirmation of such an antisense RNA by cloning and sequencing of the hybridised RNA.

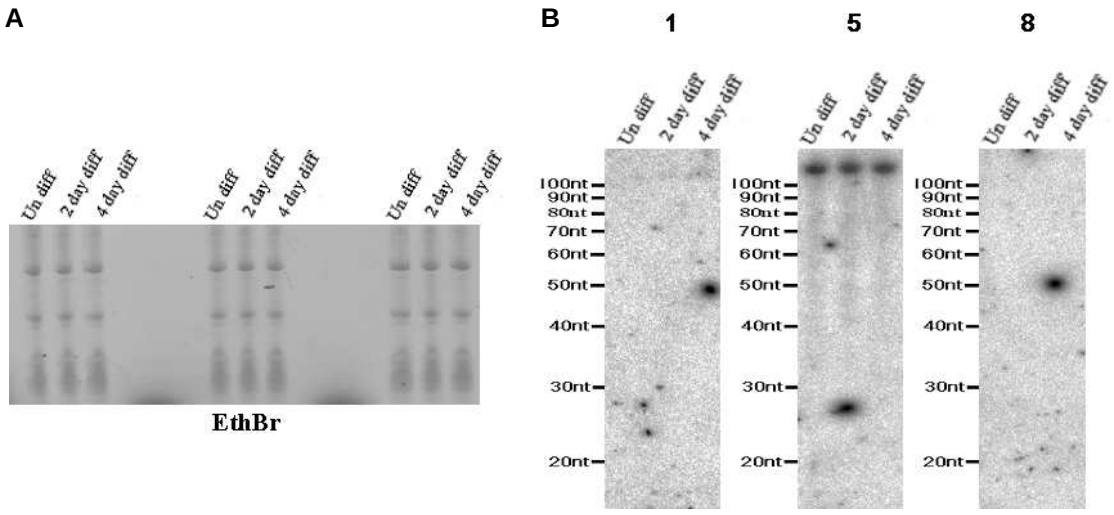


**Figure 51-Hybridisation of the ~120nt RNA with antisense-specific 5S rRNA probes across the sequence.** RNA was extracted from undifferentiated (un diff), two-day (2day diff) and four-day (4 day diff) differentiated female mouse ES cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 10µg/well of each sample was then loaded in triplicate and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the 5S rRNA antisense-specific oligonucleotide probes 1, 5 and 8 at 40℃, w ash 0.2X SSC 0.2% SDS 40℃. The radioactive signals were recorded by phosphor-imaging with one-hour exposures shown.

### *6.3.5 5S rRNA sequences identified in the massively parallel signature sequencing (MPSS) of mouse embryos.*

Hybridisation of DNA oligonucleotide probes to two-day differentiated ES cell RNA identified two sense-specific and one antisense-specific short 5S rRNA. However, because this hybridisation method is limited in resolution, it is possible that other small 5S rRNA may have been missed with this assay. Ideally high throughput sequencing of the short RNA (~19-26nt) induced in the two-day differentiated embryonic cell lines would have been used to identify the sequences of the short 5S rRNA derived from the full-length transcript. However, generating such datasets is expensive. As an alternative, an already described database containing thousands of short RNA derived from mouse embryos was made available (Mineno et al., 2006) and analysed for 5S rRNA sequences.

The full-length 5S rRNA consensus transcript was aligned with the complete unfiltered database of short RNA present in mouse embryos (discussed in chapter 4) using the BLAST (bl2seq) alignment tool. Table 13 shows there to be 61 separate sequences with identity to the 5S rRNA consensus transcript. All of the sequences identified were sense-specific and non-uniformly distribution across the 5S rRNA transcript. The number of transcripts per million for each embryonic time point has also been included. This highlights the variation in abundance between the different sequences. Many of the highest abundance sequences had significant complementarity to the short sense 5S rRNA identified in the ES RNA (Figure 48). However, there also were sequences mapping to regions where no short 5S rRNA were detected in the ES RNA hybridisations.

| 5S rRNA sequence | Regions of identity & % | | | | Embryonic day (TPM) | | |
|---|---|---|---|---|---|---|---|
| | Strand | 5S rRNA | Sequence | % | 9.5 | 10.5 | 11.5 |
| GTCTACGGCCATACCACCCT | Sense | 1-20 | 1-20 | 100 | 32 | 4 | 13 |
| CGTCTACGGCCATACCACCCTG | Sense | 1-21 | 2-22 | 100 | 9 | 7 | 11 |
| GTCTACGGCCATACCACCCTG | Sense | 1-21 | 1-21 | 100 | 0 | 48 | 26 |
| GTCTACGGCCATACCACCCTGA | Sense | 1-22 | 1-22 | 100 | 364 | 170 | 125 |
| TCTACGGCCATACCACCCTGAA | Sense | 2-23 | 1-22 | 100 | 9 | 4 | 0 |
| ATACCACCCTGAACGCGCCCGA | Sense | 11-32 | 1-22 | 100 | 9 | 23 | 0 |
| TACCACCCTGAACGCGCCCG | Sense | 12-31 | 1-20 | 100 | 4 | 2 | 8 |
| TACCACCCTGAACGCGCCCGA | Sense | 12-32 | 1-21 | 100 | 96 | 107 | 102 |
| TACCACCCTGAACGCGCCCGA | Sense | 12-33 | 1-22 | 100 | 11 | 8 | 0 |
| TCTCGGAAGCTAAGCAGGGTCG | Sense | 43-64 | 1-22 | 100 | 41 | 63 | 193 |
| CCTCGGAAGCTAAGCAGGGTCG | Sense | 44-64 | 2-22 | 100 | 15 | 0 | 0 |
| CTCGGAAGCTAAGCAGGGTCGG | Sense | 44-65 | 1-22 | 100 | 15 | 0 | 0 |
| TCGGAAGCTAAGCAGGGTCGG | Sense | 45-65 | 1-21 | 100 | 7 | 2 | 14 |
| CGGAAGCTAAGCAGGGTCGGG | Sense | 46-66 | 1-21 | 100 | 13 | 0 | 20 |
| GGAAGCTAAGCAGGGTCGGGC | Sense | 47-68 | 1-21 | 100 | 15 | 7 | 22 |
| GCTAAGCAGGGTCGGGCCTGGT | Sense | 51-72 | 1-22 | 100 | 2 | 0 | 15 |
| TCGGGCCTGGTTAGTACTTGGA | Sense | 62-84 | 1-22 | 100 | 11 | 0 | 0 |
| AGTACTTGGATGGGAGACCGC | Sense | 74-94 | 1-21 | 100 | 35 | 5 | 0 |
| CTACTTGGATGGGAGACCGCCT | Sense | 76-96 | 2-22 | 100 | 0 | 15 | 6 |
| TACTTGGATGGGAGACCGCCTG | Sense | 76-97 | 1-22 | 100 | 94 | 32 | 40 |
| ACTTGGATGGGAGACCGCCT | Sense | 77-96 | 1-20 | 100 | 11 | 0 | 0 |
| ACTTGGATGGGAGACCGCCTG | Sense | 77-97 | 1-21 | 100 | 11 | 0 | 0 |
| ACTTGGATGGGAGACCGCCTGG | Sense | 77-98 | 1-22 | 100 | 1031 | 306 | 391 |
| CTTGGATGGGAGACCGCCTGGG | Sense | 78-99 | 1-22 | 100 | 89 | 30 | 0 |
| TTGGATGGGAGACCGCCTGGGA | Sense | 79-100 | 1-22 | 100 | 60 | 27 | 8 |
| TGGATGGGAGACCGCCTGGGAA | Sense | 80-101 | 1-22 | 100 | 9 | 0 | 8 |
| GGATGGGAGACCGCCTGGGAA | Sense | 81-101 | 1-21 | 100 | 11 | 0 | 7 |
| GGATGGGAGACCGCCTGGGAAT | Sense | 81-102 | 1-22 | 100 | 31 | 0 | 0 |
| TGGGAGACCGCCTGGGAATAC | Sense | 84-104 | 1-21 | 100 | 13 | 23 | 8 |
| GGGAGACCGCCTGGGAATACCG | Sense | 85-106 | 1-22 | 100 | 11 | 0 | 7 |
| GAGACCGCCTGGGAATACCGGG | Sense | 87-108 | 1-22 | 100 | 22 | 2 | 7 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AGACCGCCTGGGAATACCGGGT | Sense | 88-109 | 1-22 | 100 | 7 | 11 | 16 |
| GACCGCCTGGGAATACCGGG | Sense | 89-108 | 1-20 | 100 | 9 | 0 | 0 |
| CGACCGCCTGGGAATACCGGGT | Sense | 89-109 | 2-22 | 100 | 11 | 0 | 6 |
| GACCGCCTGGGAATACCGGGTG | Sense | 89-110 | 1-22 | 100 | 99 | 20 | 121 |
| ACCGCCTGGGAATACCGGGTG | Sense | 90-110 | 1-21 | 100 | 91 | 50 | 88 |
| ACGCCTGGGAATACCGGGTGCT | Sense | 90-112 | 2-22 | 100 | 9 | 2 | 4 |
| CCGCCTGGGAATACCGGGTGC | Sense | 91-111 | 1-21 | 100 | 0 | 13 | 12 |
| CCGCCTGGGAATACCGGGTGCT | Sense | 91-112 | 1-22 | 100 | 119 | 66 | 77 |
| CGCCTGGGAATACCGGGTGC | Sense | 92-111 | 1-20 | 100 | 136 | 29 | 46 |
| CGCCTGGGAATACCGGGTGCT | Sense | 92-112 | 1-21 | 100 | 92 | 6 | 42 |
| CGCCTGGGAATACCGGGTGCTG | Sense | 92-113 | 1-22 | 100 | 3452 | 698 | 633 |
| GCCTGGGAATACCGGGTGCTGT | Sense | 93-114 | 1-22 | 100 | 311 | 102 | 114 |
| CCTGGGAATACCGGGTGCTGTA | Sense | 94-115 | 1-22 | 100 | 144 | 84 | 81 |
| CTGGGAATACCGGGTGCTGTAG | Sense | 95-116 | 1-22 | 100 | 232 | 74 | 132 |
| TGGGAATACCGGGTGCTGTAGG | Sense | 96-117 | 1-22 | 100 | 114 | 69 | 118 |
| GGGAATACCGGGTGCTGTAGG | Sense | 97-117 | 1-21 | 100 | 63 | 34 | 86 |
| GGAATACCGGGTGCTGTAGGCT | Sense | 98-119 | 1-22 | 100 | 69 | 64 | 51 |
| GAATACCGGGTGCTGTAGGC | Sense | 99-118 | 1-20 | 100 | 11 | 7 | 5 |
| GAATACCGGGTGCTGTAGGCT | Sense | 99-119 | 1-21 | 100 | 11 | 5 | 5 |
| GAATACCGGGTGCTGTAGGCTT | Sense | 99-120 | 1-22 | 100 | 126 | 163 | 143 |
| AATACCGGGTGCTGTAGGCTT | Sense | 100-120 | 1-21 | 100 | 31 | 40 | 16 |
| AATACCGGGTGCTGTAGGCTTT | Sense | 100-121 | 1-22 | 100 | 167 | 141 | 86 |
| ATACCGGGTGCTGTAGGCTT | Sense | 101-120 | 1-20 | 100 | 37 | 7 | 11 |
| ATACCGGGTGCTGTAGGCTTT | Sense | 101-121 | 1-21 | 100 | 108 | 104 | 35 |
| TACCGGGTGCTGTAGGCTTT | Sense | 102-121 | 1-20 | 100 | 170 | 145 | 141 |

**Table 13- 5S rRNA related RNA identified in the database of short RNA present in mouse embryos.** The sequences identified in the database with alignment to the 5S rRNA consensus are displayed in the column labelled “5S rRNA sequence” with nucleotide polymorphisms to the 5S rRNA consensus sequence indicated in red. The sequence are arranged by region of alignment to the 5S rRNA consensus sequence from 5´ to 3´, with the region of alignment indicated (5S rRNA). The region of sequence identity to 5S rRNA is also shown in the “sequence” column, with the percentage identity between the sequence and 5S rRNA consensus indicated in 5S rRNA and sequence columns given in the “%” column. The abundance of each sequences at the 9.5 day, 10.5 day and 11.5 day (9.5,10.5 and 11.5) embryonic time points is also shown.

Table 13 has been simplified into the schematic diagram shown in Figure 52. For this the highest abundance 5S rRNA sequences were selected and superimposed onto the folded 5S rRNA consensus sequence. The abundance for each sequence was calculated as the sum of the three embryonic time points analysed. The abundance of each sequence is colour coded to match the related RNA sequence. Also indicated in Figure 52 are the regions of probe hybridisation for sense-specific oligonucleotides 1 (1-16), 2 (6-31), and 8 (96-121nt) which identified the ~22nt and ~26nt short 5S rRNA (Figure 48).

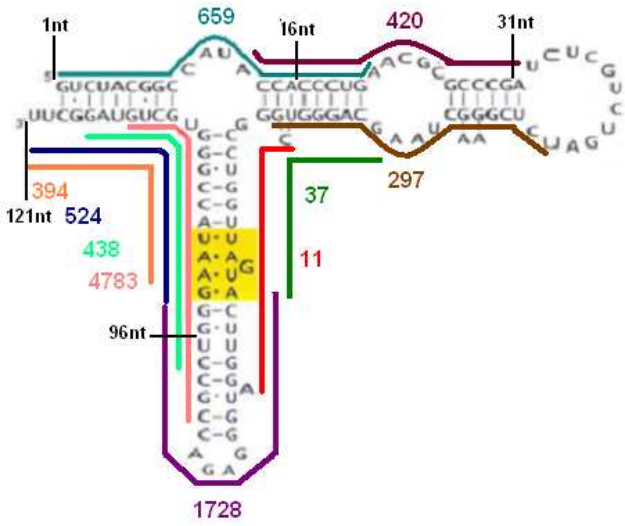


**Figure 52- The alignment of the most abundant short 5S rRNA present in mouse embryos on to the 5S rRNA secondary structure.** Each coloured line is a sequence identified in the embryonic database with identity to 5S rRNA. Each sequence is colour-coded with the number relating to the abundance of sequence in the dataset (calculated by the sum of the three embryonic time points). All of the sequences within Table 13 are not represented. Instead the most abundant sequences for each region have been selected to give an impression of the abundance of the short sequences generated across the 5S rRNA transcript.

Figure 52 shows sense-specific probe 8 (sense-specific 96-121nt) to have significant complementarity to four highly abundance sequences, originating from the 3´of the 5S rRNA transcript. Sense-specific probes 1 and 2 also overlap with two abundant 5´ sequences from the embryonic database. Therefore several sequences have been identified which overlap the short 5S rRNA identified in ES cells. The embryonic sequences also identify common cleavage points of the 5S rRNA. However, several short RNA not identified during the hybridisation of ES cell RNA were also observed. Of these the 77-98nt sequence was by far the most abundant. The sense-specific probe 6 used in the hybridisation analysis was complementary to 17nt of the 21nt 77-98nt sequence.

Therefore theoretically should have been sufficient for hybridisation. However, the ability of this sequence to form a hairpin may have prevented its identification. Alternatively the small 5S rRNA sequences present in the mouse embryo may not be exactly the same as those in the two-day differentiated ES cell RNA, due to variation in either the processing and/or stability.

### *6.3.6 Other rRNA sequences identified in the massively parallel signature sequencing (MPSS) of mouse embryos*

The apparent processing of 5S rRNA into short species was unexpected and it was speculated that processing of other rRNA into short species might also occur. It was reasoned that to investigate this possibility, sequences within the database of short RNA present in mouse embryos, with identity to 18S rRNA and 28S rRNA could be used to design probes for the hybridisation analysis of ES cell RNA. The identification of all the 18S rRNA and 28S rRNA embryonic sequences would have been impractical and unnecessary for this purpose. Instead the most abundant embryonic sequences were identified for 18S and 28S rRNA (personal communication Carles Codony-Servat). It was assumed that the abundance of the sequence would reflect the level of short RNA within the mouse embryo. Therefore by inference, these short 18S rRNA and 28S rRNA sequences may be the most easily detectable in ES cell RNA.

Table 14 shows the rRNA sequences selected. T7 DNA oligonucleotide probe templates were designed to hybridise each of these sequences (see materials and methods for probe design).

| rRNA | Most abundant rRNA sequences | rRNA alignment | Embryonic day (TPM) | | |
|---|---|---|---|---|---|
| | | | 9.5 | 10.5 | 11.5 |
| 28S (1) | TCAACAAGTACCGTAAGGGAAA | 357-378 | 21641 | 38037 | 16908 |
| 28S (2) | TCCCCGCGGGGCCCCGTCGTC | 2777-2797 | 10806 | 8437 | 20179 |
| 18S | CGAAAGCATTTGCCAAGAATGT | 995-1016 | 9945 | 6503 | 6863 |

**Table 14- The most abundant 18S rRNA and 28S rRNA related sequences identified in the database of short RNA present in mouse embryos**. The table shows highest abundance matches for rRNA in the unfiltered embryonic sequence database. The sequences are shown (most abundant rRNA sequences) along with the rRNA it aligns to (rRNA) and the position of the alignment to the rRNA transcript (rRNA alignment). The abundance at each of the mouse embryonic time points (9.5,10.5 and 11.5 day embryo's) is also indicated.

Female mouse ES cells were grown as standard with differentiation induced by the removal of LIF from the culturing medium. Undifferentiated cells were grown in parallel for comparison. Total RNA was extracted from undifferentiated, two, four and six-day differentiated ES cells and measured by $UV^{260}$ light absorbance. Short 5S rRNA when induced, are very abundant within the short RNA population. Using as little as 1μg of total RNA, the short 5S rRNA can be detected by northern blot hybridisation (see Figure 60 for example). Therefore for the detection of other short rRNA, 1μg of total RNA for each sample was loaded in triplicate and separated by denaturing-polyacrylamide gel electrophoresis. The RNA was then electro-blotted and cross-linked to nylon membrane using EDC. The nylon membrane was cut into three pieces. Each piece contained RNA from undifferentiated, two, four and six-day differentiated ES samples.

The RNA was hybridised to the three rRNA specific probes. No hybridisation was observed for the 28S (1) and the 18S rRNA-specific probes (data not shown). However, the 28S rRNA (2) probe identified two short RNA of ~19 and ~28nt (Figure 53). Unlike the short 5S rRNA, these were expressed at all time points of the ES cell differentiation. These also appeared to accumulate with progressive differentiation, being most abundant in the six-day differentiated sample. The sequence of the 28S rRNA (2) probe is not conserved between human and mouse.

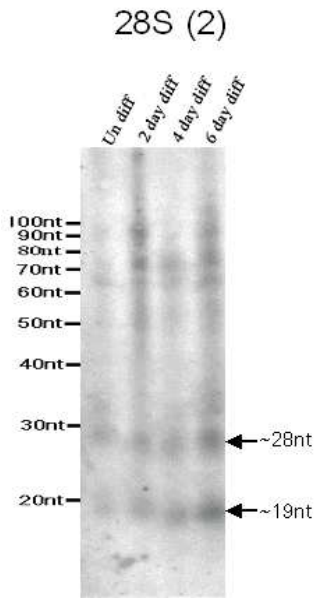


**Figure 53- 28S rRNA (2) probe hybridisation to ES cell RNA.** RNA was extracted from undifferentiated (Un diff), two-day (2-day), four-day (4-day) and six-day (6-day) differentiated female mouse ES cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 1μg/well of each sample was then loaded and separated by 10% denaturing-polyacrylamide gel by electrophoresis prior to electro-blotting and cross-linking to nylon membrane using EDC. The ES RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the 28S rRNA (2) specific probe at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor-imaging with a 16 hour exposure shown above. RNA bands of interest indicated with an arrow (right) and marker size annotated on the left of the image.

This suggests that unexpected processing of the rRNA might not be restricted to 5S rRNA with a short 28S rRNA observed to accumulate to similar levels in mouse ES cells. However, the accumulation of short 5S rRNA specifically at the two-day differentiation time point is not also a general feature of rRNA processing.

### *6.3.7 The 5´ rapid amplification of cDNA ends (RACE) of intermediate 5S rRNA to identify the genomic source of the short RNA.*

Mammalian genomes in addition to containing highly conserved, tandemly arrayed copies of 5S rRNA have many other copies dispersed throughout the genome. These dispersed copies can vary with regard to the 5S rRNA consensus sequence and may represent 5S rRNA pseudogenes (Drouin, 2000, Nishihara et al., 2006). The current model is that dispersed 5S rRNA copies are not involved in ribosome formation, with all ribosome associated 5S rRNA derived from the tandemly arrayed 5S rRNA. However, there is no direct evidence that these "pseudogene" RNA are not also involved.

It was unknown if the short 5S rRNA identified were as a result of 5S rRNA contributing to the ribosome or a feature of dispersed 5S rRNA transcript processing. It was reasoned that variation from the 5S rRNA consensus could be used to differentiate between RNA derived from the tandemly arrayed and dispersed 5S rRNA copies. Therefore, sequencing of 5S rRNA RACE products was carried out to provide direct sequence information and so allow polymorphisms to be identified.

The short 5S rRNA (~19-26nt) would only be able to provide ~10nt of discriminatory sequence using this method. It was doubtful that the length of discriminatory sequence obtained from the short 5S rRNA, would be sufficient to differentiate between the tandemly arrayed 5S rRNA and pseudogenes. It was hypothesised that the intermediate RNA (~40-100nt) were generated though the same processing pathway as the short 5S rRNA due to their accumulation along with the short RNA and thus were derived from the same genomic transcripts. Therefore, the intermediate 5S rRNA were chosen for sequencing, increasing the length of discriminatory sequence obtained and the likelihood of sufficient polymorphic sites being identified to differentiate between source sequences.

The intermediate 5S rRNA of ~80nt (Figure 48C) which were observed to be co-induced with the ~22nt sense short 5S rRNA were selected for analysis. The ~22nt 5S rRNA mapped to the 5´ of the 5S rRNA transcript. From this it was deduced that the 3´end of the ~80nt 5S rRNA would be ~60-80nt downstream of the 5S rRNA transcription start site.

Three 5S rRNA sequences that have been annotated as transposed (Drouin, 2000) or pseudogenes (Emerson and Roeder, 1984, Leah et al., 1990) copies were aligned with the 5S rRNA consensus sequence. Two 5S rRNA specific primers were then designed to be complementary to the regions of highest conservation at the hypothetical 3´ of the ~80nt RNA. Degenerate bases were included where there was <100% identity between sequences. The aim was to avoid bias toward the 5S rRNA consensus. The alignment of sequences and positioning of primers is shown in Figure 54.

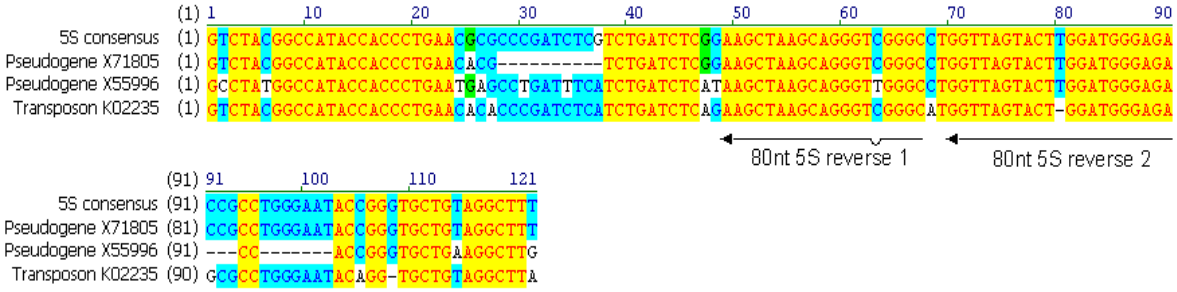


**Figure 54- 5S rRNA consensus alignment to mouse pseudogenes.** The 5S rRNA consensus sequence has been aligned to two known mouse pseudogenes (accession numbers X71805, X55996) and one transposed (accession number K02235) 5S rRNA copy. The primers designed for the RACE analysis are indicated by arrows below the sequence (80nt 5S reverse 1 and 2). Yellow= 100%, blue=75% and green= 50% conservation between the aligned sequences.

RNA extracted from ES cells after two-days of differentiation was separated by denaturing-polyacrylamide gel electrophoresis. The 60-80nt fraction of the gel was excised and the contents eluted overnight (see materials and methods). After concentration of the RNA (without de-phosphorylation and re-phosphorylation) a 5´adaptor was ligated. Reverse transcription was carried out using reverse transcriptase with random hexamers to produce the PCR template. PCR of the template was carried out with the adaptor primer and either 5S rRNA specific primer 1 or 2 (as shown in Figure 54). PCR with both 5S rRNA specific primers amplified DNA of the expected size. The amplified DNA was then ligated into a plasmid vector and transformed into bacteria. Colonies were screened by PCR across the region of insertion. Several colonies containing inserts were selected and sequenced.

An alignment of the sequences obtained to the 5S rRNA consensus sequence is shown in Figure 55. Colonies sequenced from reverse primer 1 RACE show complete identity to the 5S rRNA consensus with the exception of plasmid 1. This sequence was absent the first nucleotide of the 5S rRNA sequence. Using the BLAST algorithm on the ensemble website (http://www.ensembl.org/index.html) to search the mouse genome no matches for this single nucleotide deletion were identified. This indicated the variation observed was likely to be as a result of a PCR or cloning artefact. Sequences obtained from reverse 2 primer RACE are also identical to the consensus, with the exception of plasmid 7. The 19nt deletion of the 5´end in the sequence was also not identified in the mouse genome. This sequence may also be derived from a PCR or cloning artefacts or alternatively may represent the 60nt intermediate 5S rRNA, which was not sufficiently excluded by gel purification.

The sequencing of the RACE products indicates that the ~80nt 5S rRNA are derived from tandemly arrayed or un-diverged dispersed 5S rRNA copies. We infer from this that the short ~22nt identified with the same probe (Figure 48C) are also derived from consensus 5S rRNA. However, we cannot exclude the possibility that the short RNA are derived from different progenerator 5S rRNA transcripts.

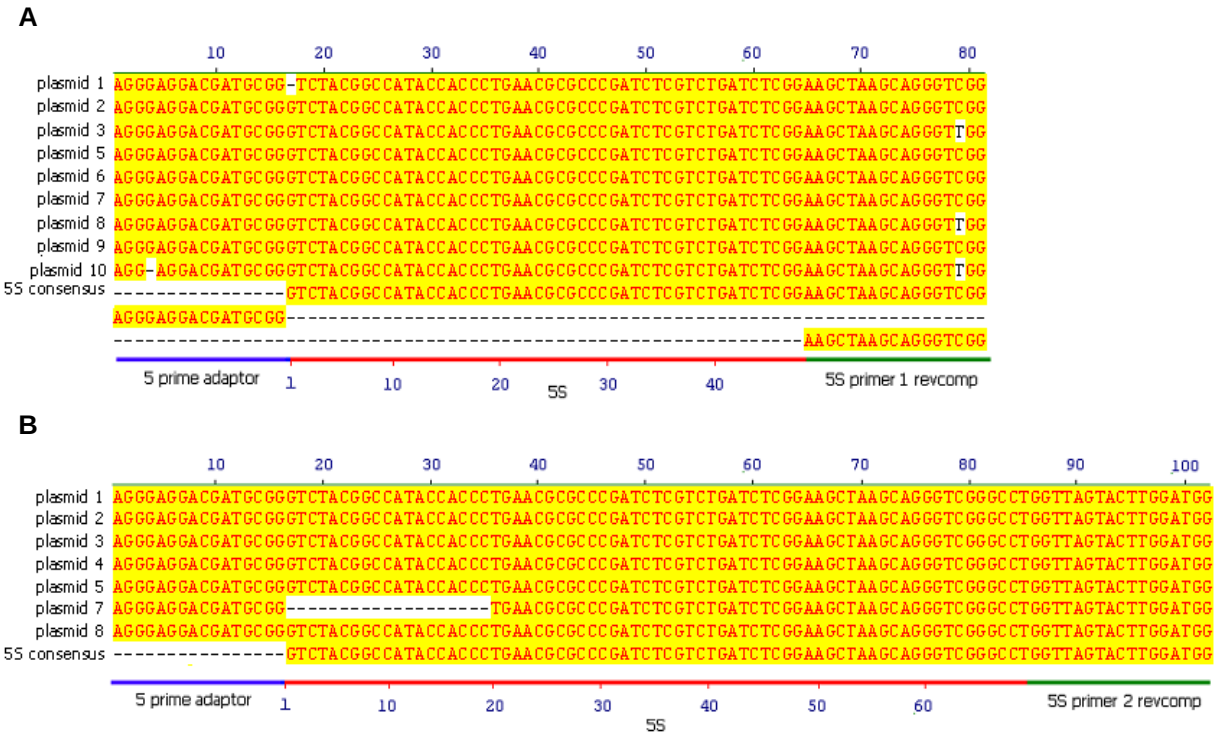


**Figure 55- Race sequences.** Clones containing RACE PCR products from 5S 80nt primer 1 (A) and 5S 80nt primer 2 (B) PCR were sequenced and aligned to the 5S rRNA consensus sequence. The line below the alignments indicate primer specific regions of the sequence (Blue=5´ adaptor primer and green= 5S rRNA specific primer), with the red section of the line representing RACE derived 5S rRNA sequence.

### *6.3.8 Recombinant human Dicer processes intermediate 5S rRNA into short (siRNA-like) RNA.*

Short discrete RNA are known to be produced by the action of the RNase III enzyme, Dicer on longer dsRNA substrates. Therefore it was investigated whether Dicer may be involved in the processing pathway of the short 5S rRNA identified (~19-26nt).

RNA was extracted from undifferentiated and two-day differentiated P19EC cells and measured by $UV^{260}$ light absorbance. Equal amounts of RNA were then separated by denaturing-polyacrylamide gel electrophoresis. The 50-80nt

fraction was excised and the contents eluted overnight. After concentration, the purified RNA was measured by $UV^{260}$ light absorbance. An equal volume and concentration of undifferentiated and differentiated RNA was incubated in the appropriate buffer with and without recombinant human Dicer (see materials and methods). The RNA samples were subsequently separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (Figure 56A). The stained image of the RNA shows no general degradation to have taken place with the tRNA appearing as discrete bands ~70nt. The RNA was then electro-blotted and EDC cross-linked to nylon membrane.

The RNA was hybridised to the full-length 5S rRNA sense-specific probe. This identified a ~22nt sense RNA in undifferentiated and 2 day differentiated RNA treated with Dicer. To determine whether the short 5S rRNA generated were similar to those identified in the DNA oligonucleotide mapping (Figure 48), the RNA was subsequently hybridised to 5S sense-specific DNA oligonucleotide probes 2 (hybridising 6-31nt) and 8 (hybridising 95-121nt). Probe removal between hybridisations was carried out (method 2) and verified by phosphor-imaging.

The 6-31nt sense-specific probe identified a short ~22nt 5S rRNA in the Dicer-treated sample alone (Figure 56). This RNA was comparable in size to the short endogenous RNA detected in two-day differentiated female ES cell RNA with this probe. However, in additional to this, a short RNA of ~28nt was observed in both Dicer-treated and untreated samples. This RNA was significantly shorter than the gel purified fraction and maybe derived from generalised processing of the intermediate 5S rRNA by any nuclease present under the conditions used. Alternatively the ~28nt RNA may have originally fractionated alongside the 50-80nt RNA despite their smaller individual length, perhaps because of incomplete denaturation affecting their mobility. However, with the subsequent gel purification and denaturation, effective denaturation of these samples may have resulted in the correct size in the second denaturing-polyacrylamide gel. The 95-121nt sense-specific probe also identified short RNA of ~22nt and ~26nt specifically in the Dicer-digested samples. Of these the ~26nt RNA was of comparable size to the endogenous RNA observed in the two-day differentiated female ES cells with this probe.

Therefore short RNA of the expected size were detected specifically in the Dicer treated 50-80nt gel purified RNA samples (extracted from two-day differentiated P19 EC cells). This would appear to indicate a Dicer specific processing of the intermediate 5S rRNA into the shorter RNA. However, the subsequent hybridisation of the 5S rRNA sense-specific DNA oligonucleotide probe pool 2 (probes 3-5), also detected Dicer specific short RNA (Figure 56E). This was unexpected because these probes did not identify any short RNA (<30nt) when hybridised to the two-day differentiated ES cell RNA. This may highlight differences in the *in vivo* and *in vitro* processing of the intermediate 5S rRNA. The ability of Dicer to process 5S rRNA specific sequences within this gel-purified fraction does implicate these to have substantial dsRNA character.

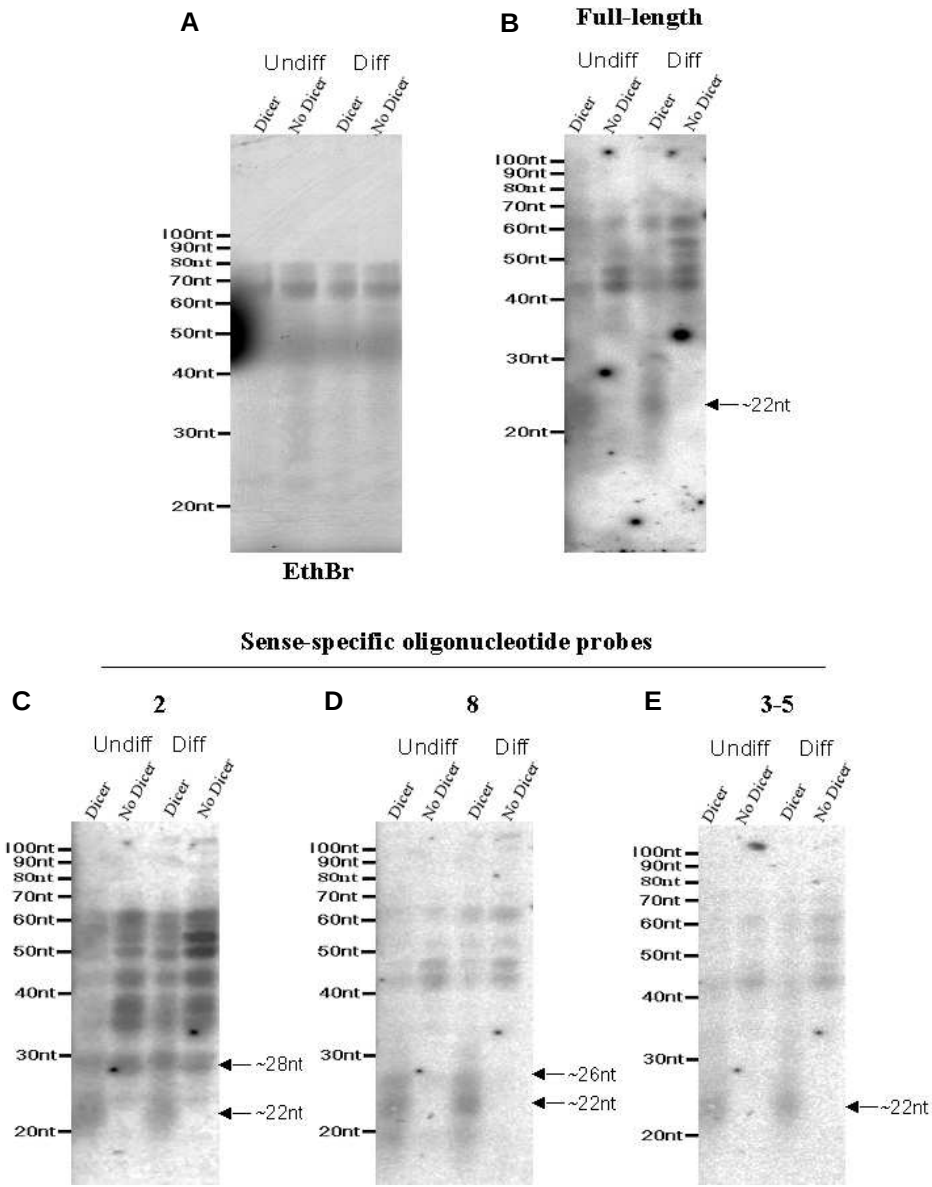


**Figure 56-*In vitro* Dicer digest of gel purified P19 EC RNA.** RNA was extracted from undifferentiated (Un diff) and two-day (2-day) differentiated P19 EC cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. An equal amount of each RNA sample was then loaded and separated by 10% denaturing-polyacrylamide gel electrophoresis. The 50-80nt fraction of RNA was excised and the contents eluted overnight. The concentration of RNA in these samples was then measured by $UV^{260}$ light absorbance before 5µg of each sample was treated with Dicer in parallel Dicer null controls were carried out. The samples were then loaded and separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide (A) before being electro-blotted and cross-linked to nylon membrane using EDC. The RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the full-length 5S rRNA sense-specific (B), 2 (C), 8 (D) and 3-5 (E) 5S rRNA sense-specific DNA oligonucleotide probes, at 40℃, wash 0.2x SSC 0.2% SDS 40℃. The r adioactive signals were recorded by phosphor-imaging with 16 hour exposures shown above. The RNA bands of interest are indicated with an arrow (right) with marker size annotated on the left of the images.

### 6.3.9 The reduction of short 5S rRNA accumulation in Dicer-null embryonic stem cells.

To further assess if Dicer was involved in the *in vivo* production of short 5S rRNA, Dicer-null cells were subjected to differentiation and RNA was extracted for similar analysis to that carried out in Figure 44. The origin of these cells has already been described in chapter 5.

The mouse Dicer-null and control (AB 2.2) ES cells were grown as standard with differentiation induced by the removal of LIF from the culturing medium. Undifferentiated cells were grown in parallel for comparison. Total RNA was extracted from undifferentiated, two, four and six-day differentiated cells and measured by $UV^{260}$ light absorbance. Equal amounts of each sample were separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide. This showed both the undifferentiated and two-day differentiated AB 2.2 cell RNA to be smeared in appearance (Figure 57A). Previously this had only been observed for two-day differentiated RNA samples (Figure 44). Undifferentiated cell populations can contain a variable number of differentiated cells, which may explain this observation. The RNA was then electro-blotted and cross-linking to nylon membrane using EDC.

To verify the Dicer-null cells were functionally null, the RNA was hybridised to a mmu-mir-292as specific probe, because the accumulation of mature micro RNA is known to absolutely depend on Dicer (Kanellopoulou et al., 2005) so even low levels of Dicer activity might produce detectable miRNA. This miRNA was selected because of its high expression in mouse ES cells. This showed an RNA corresponding in size to the mature miRNA across all time points of differentiation for the control ES cell RNA. However, there was no signal observed in the Dicer-null RNA samples. This result was consistent with the expected phenotype of the Dicer-null cells. The mmu_mir_292as specific probe was removed (method 2) and verified by phosphor-imaging.

The RNA was subsequently hybridised to the 5S rRNA sense-specific DNA oligonucleotide probe 8, which hybridises nucleotides 95-121 of the sense transcript (Figure 57B). This identified short RNA of ~19nt and ~26nt, similar to those previously observed in ES cells. These were observed in the

undifferentiated and two-day differentiated control and faintly in the two and six-day differentiated Dicer-null ES samples (Figure 57B). The intensity of the RNA signal observed in the Dicer-null samples was significantly reduced in comparison to the control samples.

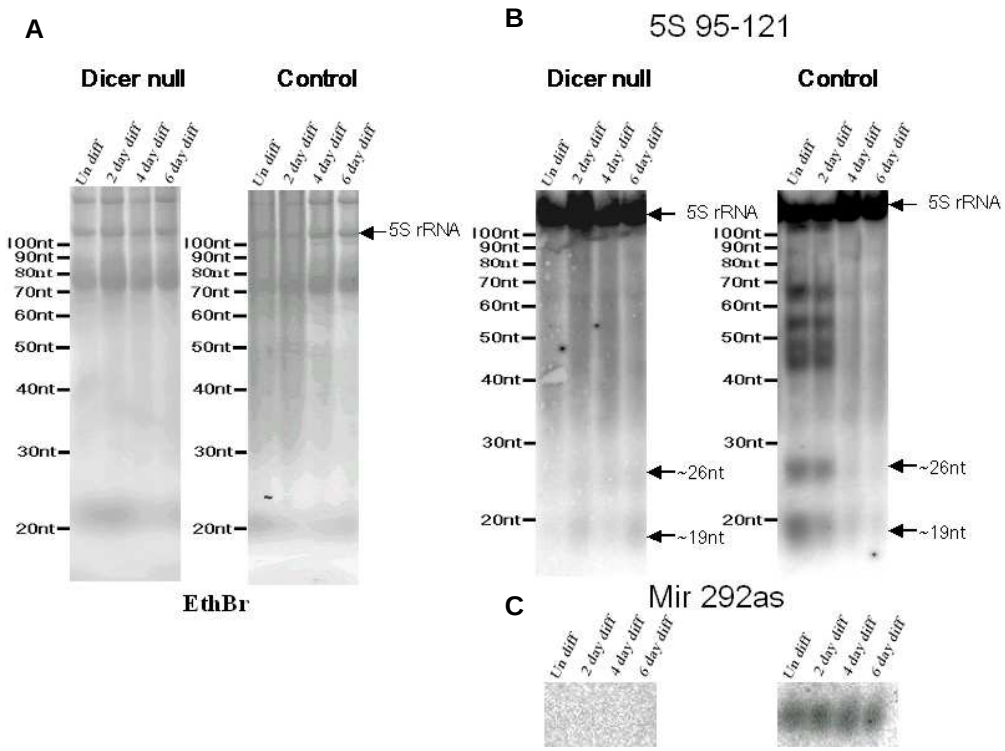


**Figure 57- Short 5S rRNA in Dicer-null ES cells.** RNA was extracted from Undifferentiated (Un diff), two-day (2 day diff), four-day (4 day diff) and six-day (6 day diff) differentiated mouse ES Dicer-null and ES AB 2.2 cells. The concentration of RNA in these samples was measured by $UV^{260}$ light absorbance. 50μg/well of each RNA sample was then loaded separated by 10% denaturing-polyacrylamide gel electrophoresis. The gel was stained with ethidium bromide (A) prior to electro-blotting and cross-linked to nylon membrane using EDC. The ethidium bromide staining at ~20nt indicates the progression of the loading dye. The RNA samples were then hybridised in 2X SSC and 2% SDS buffer to the 5S rRNA (95-121nt) sense-specific (B) and mmu-mir-292as (C) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The ra dioactive signals were recorded by phosphor-imaging with 16 hour exposures shown above. The RNA bands of interest indicated with an arrow (right) and marker size annotated on the left of the image.

### 6.3.10 The identification of short 5S rRNA in human cancer cell lines.

Because the short 5S rRNA were observed during the differentiation of all embryonic cells analysed, it was aimed to observe if the differentiation of cells of non-embryonic origin would similarly cause the accumulation of short 5S rRNA. The treatment of the CaCo-2 carcinoma cell line with sodium butyrate has been reported to induce differentiation (Harrison et al., 1999).

The CaCo-2 cells were differentiated by the addition of butyrate for three-days to the culturing medium at two different concentrations previously reported to induce differentiation (Harrison et al., 1999). Control cells were incubated without butyrate in parallel for comparison. Total RNA was extracted from the undifferentiated and differentiated CaCo-2 cells using the same method as for the ES cells (personal communication Carles Codony-Servat) and measured by $UV^{260}$ light absorbance. Equal amounts of each sample were then separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (Figure 58A). This showed the samples to be equally loaded. The RNA was then electro-blotted and cross-linked to nylon membrane using EDC.

The RNA was hybridised to the full-length 5S rRNA sense-specific probe which identified an RNA of ~22nt in the untreated CaCo-2 RNA. This short RNA appeared to be down regulated in the butyrate treated samples. An additional faint band of ~19nt was also observed for both untreated and 5μM butyrate treated CaCo-2 RNA. The probe was removed (method 2) and verified by phosphor-imaging, before the consecutive hybridisation of the 1-22nt and 92-113nt sense-specific 5S rRNA probes. This was done to identify if the CaCo-2 short 5S rRNA mapped to the same region as either of the short RNA observed in the two-day differentiated ES sample.

Figure 58C shows hybridisation of the 1-22nt sense-specific 5S rRNA probe. This identified a RNA of ~22nt predominantly in the untreated CaCo-2 RNA as observed with the full-length probe. The reduction of the ~22nt RNA in the butyrate treated CaCo-2 RNA also correlated with the appearance of a ~38nt RNA. The other intermediate 5S rRNA were also hybridised. These highlighted the full-length intermediate bands that contain 5´sequence. Hybridisation of the

RNA to the 95-113nt sense-specific probe identified a short RNA of ~19nt and possibly ~24nt (Figure 58D). The short 5S rRNA identified with the 92-113nt probe was detected predominantly in untreated and 5μM butyrate samples. The ~19nt RNA was observed to be slightly down-regulated in the 2μM butyrate treated RNA sample. The 5´ and 3´ intermediate 5S rRNA bands were also observed to predominantly correlate with different intermediate bands detected with the full-length 5S probe (~60nt and ~40nt bands respectively). After probe removal the CaCo-2 RNA were hybridised to the hsa-mir-21 specific probe. This was done to confirm equal RNA transfer and hybridisation efficiency of the RNA samples (Figure 58E). Thus, cells of a non-embryonic origin also accumulate short 5S rRNA. In contrast to ES cells these decreased in abundance following the differentiation. However, all cells accumulate intermediate size RNA that could be processed by Dicer as shown in Figure 56.

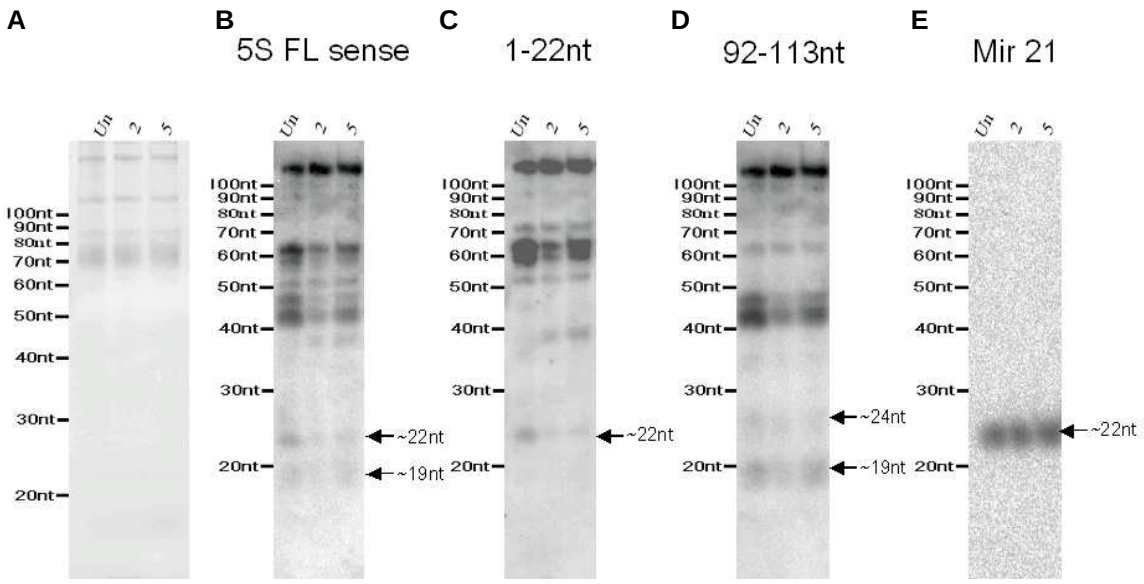


**Figure 58- 5S rRNA short RNA in CaCo2.** RNA was extracted from CaCo-2 cells, untreated (Un) and treated with 2μg (2) or 5μg (5) of sodium butyrate (personal communication Carles Codony-Servat) and measured by $UV^{260}$ light absorbance. 5μg/well of each sample was separated by 10% denaturing-polyacrylamide gel electrophoresis and the gel stained with ethidium bromide (A). The RNA was then electro-blotted and EDC cross-linked to nylon membrane. The RNA was hybridised in 2X SSC and 2% SDS buffer to the full-length 5S rRNA sense-specific (B), 1-22nt sense-specific (C), 95-113nt sense-specific (D) and hsa-mir-21 (E) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosphor- imaging with one-hour exposures shown for the 5S rRNA hybridisations and 16 hour exposure shown for the hsa-mir-21 hybridisation. RNA bands of interest indicated with an arrow (right) and marker size are annotated on the left of the image.

To test whether the short 5S rRNA could be observed in other cancer cell lines, RNA was extracted from the human cancer cell lines Hela (cervical origin), MCF-7 (breast origin), MDA-MB-435 and MDA-MB-436 (possible breast origin). These cell lines were selected based on their availability in the laboratory.

An equal amount of each RNA sample was separated by denaturing-polyacrylamide gel electrophoresis and stained with ethidium bromide (Figure 59A). The RNA was then electro-blotted and cross-linked to nylon membrane using EDC. Because the 5S rRNA consensus sequence is identical between human and mouse the same 5S rRNA probes were used for hybridisation. Hybridisation of the RNA to the full-length 5S rRNA sense-specific probe identified short RNA of ~19nt and ~22nt (Figure 59B). These were observed at variable levels in Hela and MDA-MB-435 cell lines but absent in MCF-7 and MDA-MB-436 cell lines.

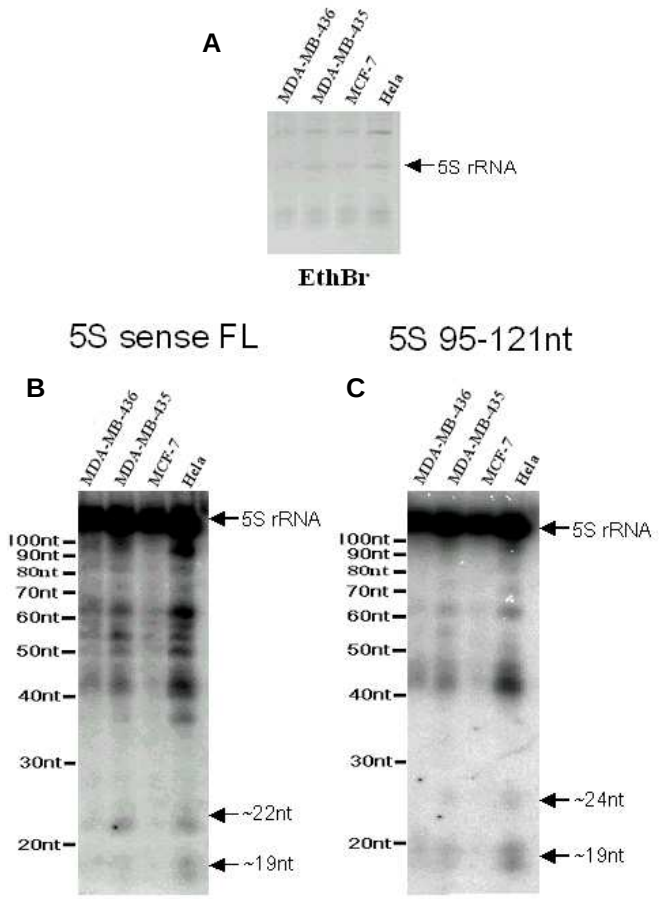


**Figure 59-Short 5S rRNA in human cancer cell lines.** RNA was extracted from MDA-MB-436, MDA-MB-435, MCF-7 and Hela cancer cell lines. Each RNA sample was measured by $UV^{260}$ light absorbance and 1μg/well was separated by 10% denaturing-polyacrylamide gel electrophoresis. The RNA was stained with ethidium bromide (A) before electro-blotting and cross-linking to nylon membrane using EDC. The RNA was hybridised in 2X SSC and 2% SDS buffer, to the full-length 5S rRNA sense-specific (B) and 5S rRNA (95-121nt) sense-specific (C) probes at 40℃, wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorde d by phosphor-imaging with 16 hour exposures shown. The RNA bands of interest indicated with arrows (right) and the marker size annotated on the left of the images.

### 6.3.11 Short 5S rRNA identified in human breast tissue.

Figure 58 and Figure 59 indicate the presence of short 5S rRNA in cell lines of non-embryonic origin. However, it remained possible that these small RNA were an artefact of culturing cells *in vivo*. To assess if the short 5S rRNA could also be identified in human tissue, RNA was extracted from a number of frozen human breast tumour samples representing a range of types and stages of breast cancer. RNA from breast reduction patients was also extracted to provide "normal" control tissue (Table 15). This material was obtained from the pathology department at the Western Infirmary, Glasgow and was anonymous and randomly selected with approval of the Local Research Ethics Committee. After extraction, RNA concentration was measured by $UV^{260}$ light absorbance.

| Name | Tissue | Age |
|---|---|---|
| Normal breast A | Normal | 26 |
| Normal breast B | Normal | 19 |
| Human breast tumour 1 | Grade 1 epithelial carcinoma | 66 |
| Human breast tumour 2 | Grade 1 epithelial carcinoma | 87 |
| Human breast tumour 3 | Grade 2 epithelial carcinoma | 74 |
| Human breast tumour 4 | Grade 3 epithelial carcinoma | 59 |
| Human breast tumour 5 | Grade 3 epithelial carcinoma | 39 |
| Human breast tumour 6 | Lobular carcinoma | 46 |
| Human breast tumour 7 | Lobular carcinoma | 74 |
| Human breast tumour 8 | Ductal carcinoma *in situ* | 74 |

**Table 15- Breast tissue extracted for RNA.** The table above displays the names that will be used to refer to RNA from the related patients (name column). The type of tissue and grade of cancer (where known) are shown in the tissue column, with the age of the patients detailed in the age column.

1μg/well of each RNA sample was separated by denaturing-polyacrylamide gel electrophoresis. The ethidium bromide stain of the RNA (Figure 60A) is shown to provide an indication of RNA quality. Despite the faint appearance of bands due to the low concentration of RNA loaded, no overt degradation of the RNA was observed with all samples appearing equal. The RNA was subsequently electro-blotted and RNA cross-linked to nylon membrane using EDC.

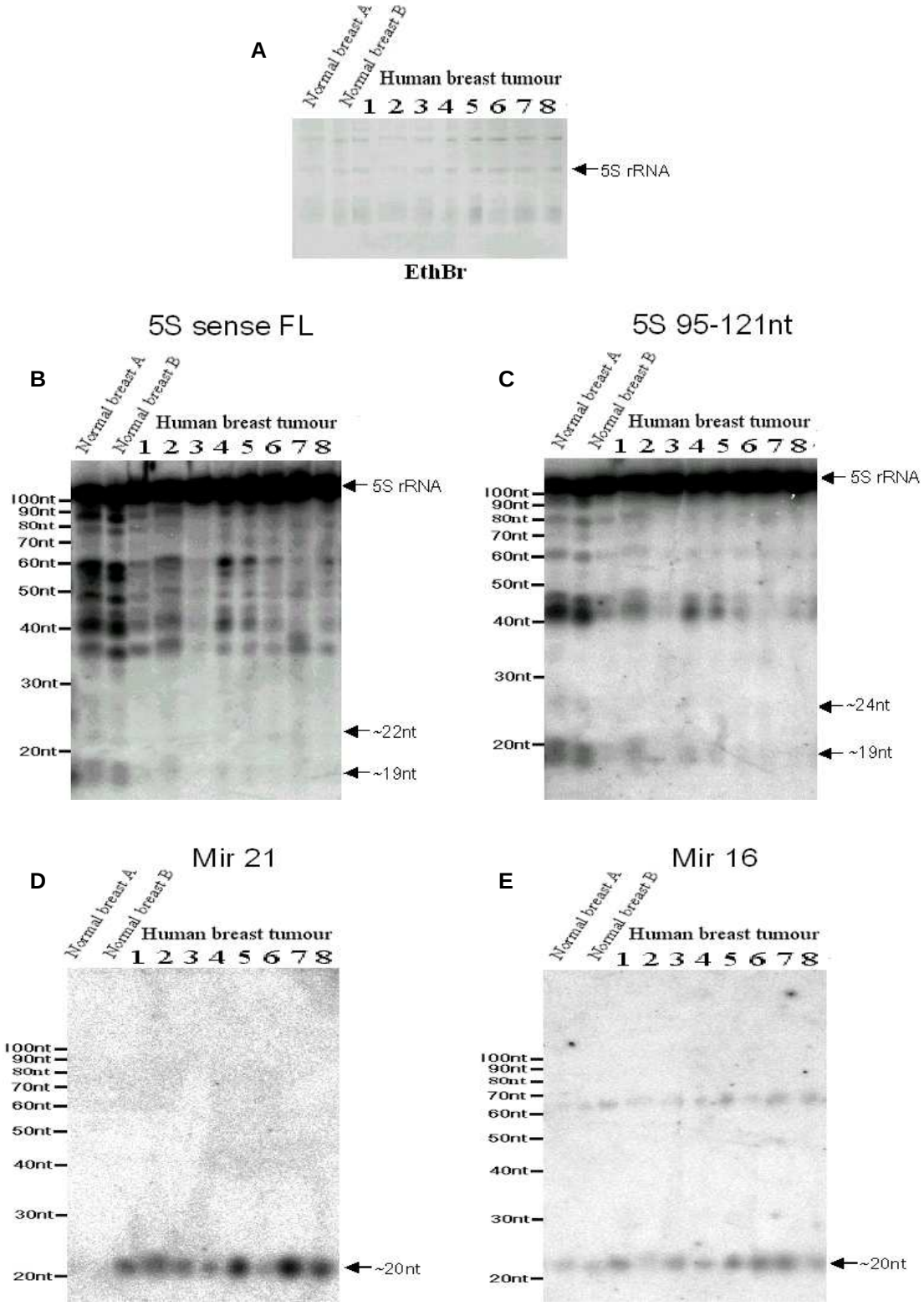


**Figure 60- Short 5S rRNA in human breast cancer.** RNA was extracted from various human breast tumours (numbered 1-8) and normal breast controls (normal breast A and B). Each RNA sample was measured by $UV^{260}$ light absorbance. 1µg/well of each sample was then separated by 10% denaturing-polyacrylamide gel electrophoresis. The RNA was stained with ethidium bromide (A) before electro-blotting and cross-linking to nylon membrane using EDC. The RNA was hybridised in 2X SSC 2% SDS buffer to the full-length 5S rRNA sense-specific (B), 5S rRNA (95-121nt) sense-specific (C), hsa-mir-21(D) and hsa-mir-16 (E) probes at 40℃ wash 0.2X SSC 0.2% SDS 40℃. The radioactive signal was recorded by phosp hor-imaging with 16 hour exposures shown. The RNA bands of interest are indicated with an arrows (right) and marker size annotated on the left of the images.

Hybridisation of the RNA to the full-length 5S rRNA showed short RNA of ~19nt, ~22nt in normal breast samples (Figure 60B). However, these short 5S rRNA were absent or substantially reduced in the breast tumour samples. The most abundant intermediate 5S rRNA bands (~40-100nt) were observed to correlate with short 5S rRNA expression. This pattern was consistent with previous observations made in cell line RNA. The probe was removed (method 2) and this verified by phosphor-imaging, before hybridisation to the 5S rRNA 95-121nt sense-specific probe. This identified short 5S rRNA of ~19nt and a faint band of ~24nt in the normal breast samples (Figure 60C).

To control for the transfer and accessibility of short RNA, the RNA was hybridised to probes specific for hsa-mir-16 and hsa-mir-21. Hybridisation of the RNA to the hsa-mir-21 specific probe showed a general up-regulation of hsa-mir-21 in breast tumour samples compared with normal breast controls (Figure 60D). This observation supports previous findings of hsa-mir-21 up-regulation in breast cancers (Iorio et al., 2005). Hybridisation of the hsa-mir-16 specific probe, which has not been documented as substantially differentially regulated in breast cancer, in contrast showed a more generalised expression across all samples (Figure 60E).

### *6.3.12 Transgenes containing sequences complementary to the short 5S rRNA identified were not observed to be down-regulated*

Because of the similarity in size of the short 5S rRNA to siRNA and miRNA it was thought possible that they may target gene expression in the same way. If this was the case, then any mRNA containing complementary sequence would be expected to display down regulation. Using the BLAST algorithm on the ensemble website (http://www.ensembl.org/index.html) to search the mouse expression sequence tag (EST) database, a number of hypothetical mRNA targets for the short 5S rRNA were identical. Of these, sialic acid acetylesterase (SIAE) was observed to contain significant complementarity to the 5` of the 5S rRNA transcript in its 3` UTR, making it a potential target for the 5` (~1-22nt) short 5S rRNA identified (Figure 63).

To assess if the short 5S rRNA identified were capable of targeting complementary sequences for destruction, firefly luciferase sensor-constructs were designed as described in chapter 5. These constructs contained 5S rRNA sequence complementarity to 1-25nt, 77-98nt and 92-113nt of the sense consensus, or the first 60nt of the 90nt 5S rRNA sequence from the SIAE mRNA. Negative controls for each of these constructs were generated by the introduction of polymorphisms as described in chapter 5 with the exception of the SIAE sensor which used the sensor sequence inserted in the opposite orientation. In addition, to test the accessibility of the target site and the effectiveness of the polymorphic negative controls, sensor and polymorphic controls were generated for the micro RNA, hsa-mir-16 and hsa-mir-21. Insertion of the target sequence in the correct orientation was established by PCR screening of colonies and confirmed by sequencing.

The short 5S rRNA were previously observed to accumulate in undifferentiated CaCo-2 cells. Therefore the sensor-constructs for the various 5S rRNA regions, miRNA positive controls and their individual polymorphic negative controls were co-transfected with the unmodified renilla luciferase plasmid (pHRL) into undifferentiated CaCo-2 cells under standard conditions. Additional controls of unmodified firefly luciferase co-transfected with pHRL and individual transfections of each were also included to assess the specificity of the dual luciferase assay. The transfected cells were harvested after 48 hours and divided into two pellets. One pellet was used to produce the cell lysate for the luciferase assays to assess the targeting of firefly luciferase. The second pellet was used to extract RNA, to assess the presence of the short 5S rRNA in the cells.

If the luciferase mRNA contains a region which is targeted by the short RNA, that mRNA would be expected to undergo accelerated degradation and/or translational repression if the short RNA is functioning in a similar manner to an siRNA or miRNA. The cell lysate for each transfection was examined by a single read of firefly and renilla luciferase levels by luminometery. The values obtained from this analysis were then used to produce a ratio of the luciferase activities. Targeting would therefore be observed by an increase of the ratio of renilla luciferase to firefly luciferase relative to the control samples because of the relative decline in firefly luciferase activity. This experiment was repeated three

times and the means with standard deviation shown in Figure 61 and Figure 62. Figure 61 displays the Renilla to firefly ratios of the 5S rRNA and Hsa-mir-16 sensor-constructs to their polymorphic negative controls. From this graph there does appear to be a slight increase of the 1-25nt and Hsa-mir-16 sensor-construct ratios in relation to their controls.

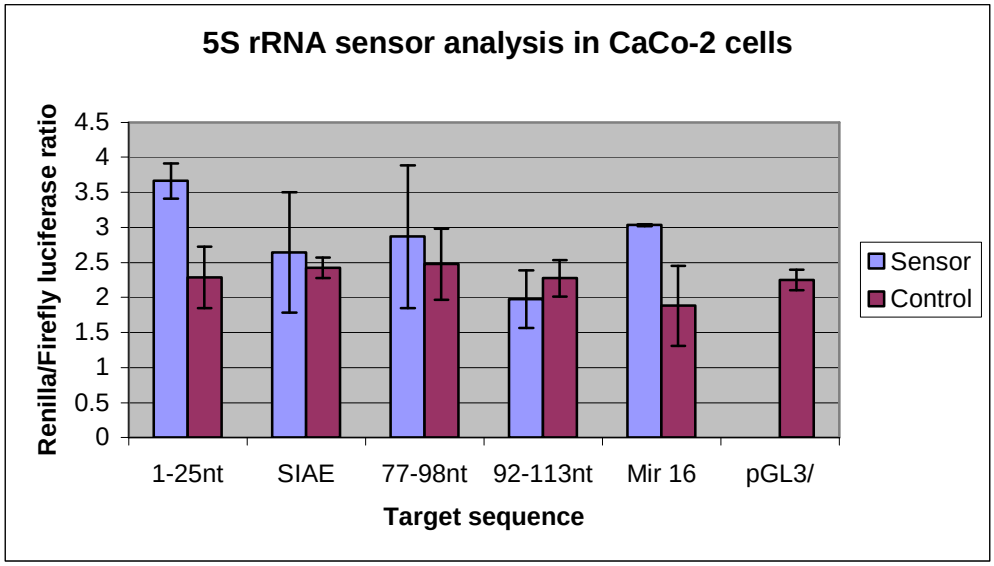


**Figure 61- The renilla/firefly luciferase ratios for the transfection of 5S rRNA sensor constructs into CaCo-2 cells without the Hsa-mir-21 control.** CaCo-2 cells were grown in standard conditions for 24 hours before transfection with 1ng pHRL plasmid (renilla) and 90ng of pGL3 (firefly) construct containing the target sequences for the 5S rRNA sense 1-25nt, 77-98nt and 92-113nt short RNA. The positive control containing hsa-mir-16 (Mir 16) target sequence and unmodified pGL3 were also included. Negative controls for the transfection included polymorphic target sequences for each of the sensor constructs and the transfection of the native pGL3 plasmid. Cells were lysed after 24 hours of transfection with the lysate used for fluorescence luminometery. The dual luciferase kit was used to assess levels of renilla and firefly protein. The renilla to firefly ratios were calculated for three separate transfections and combined to generate the above graph. The sensor constructs target sequence is indicated in the x-axis with the sensor (blue) and their specific polymorphic and native pGL3 controls (purple).

These increases are comparable to those of (O'Donnell et al., 2005) that are claimed to indicate targeting. However, they appear insignificant when compared to the targeting of the Hsa-mir-21 sensor construct (Figure 62). The hsa-mir-21 sensor-construct is targeted with almost a 200-fold decrease in firefly luciferase protein relative to renilla luciferase protein. This experiment was repeated in Hela cells where the strong targeting of the Hsa-mir-21 sensor-construct was also observed. However, no comparable change to the Hsa-mir-16 or 5S rRNA sensor-constructs as identified (data not shown).

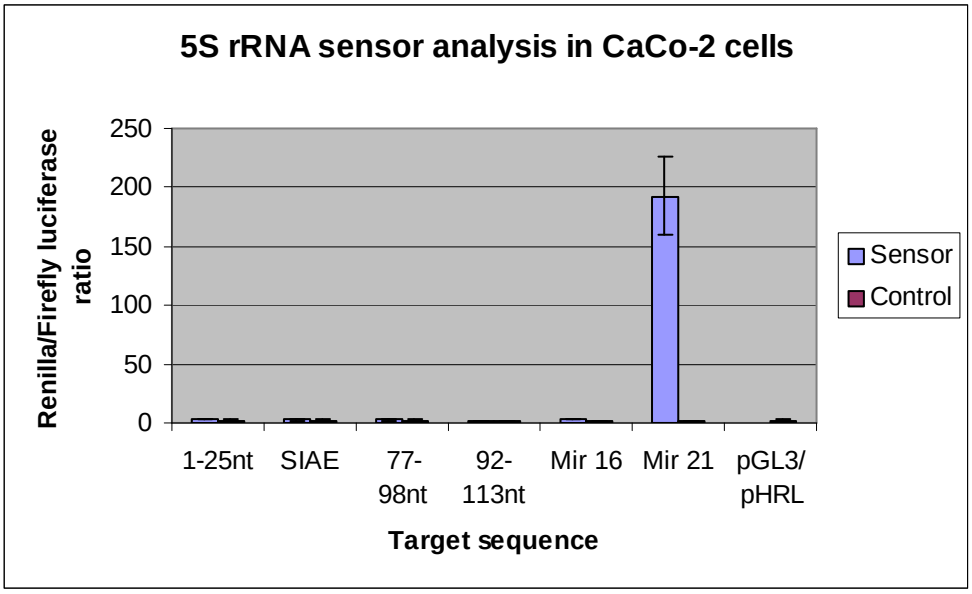


**Figure 62- The renilla/firefly luciferase ratios for the transfection of 5S rRNA sensor constructs into CaCo-2 cells with the Hsa-mir-21 control.** The renilla to firefly luciferase ratios calculated for three separate transfections in Figure 61 have been re-plotted with the addition of the hsa-mir-21 (Mir 21) values. The sensor constructs target sequence is indicated in the x-axis with the sensor (blue) and their specific polymorphic and native pGL3 controls (purple).

The hsa-mir-21 sensor result confirms that target sequences inserted into the XbaI site of pGL3 are accessible to small RNA targeting. However, the second positive control of hsa-mir-16 and the 5S rRNA sensors do not show comparable targeting. This may indicate other aspects such as the relative abundance of short RNA to target sequence or the sequence dependent accessibility of the site may play a major role in targeting efficiency. Therefore sensor experiments can provide evidence of short RNA targeting when positive. The inconsistency of known miRNA to elicit targeting in this assay makes the interpretation of negative results difficult.

### 6.3.13 RT-PCR confirms the SIAE mRNA to be reduced in cells expressing short 5S rRNA

The following section contains data largely generated in collaboration with Dr Jane Byrne. This material has been included because it is a direct progression of the work discussed in the previous section and provides additional data and insight into possible roles of short 5S rRNA.

The targeting ability of the short 5S rRNA was also investigated using RT-PCR to assess the transcript levels of putative human mRNA targets. The identification of possible human short 5S rRNA targets was carried out using the BLAST tool to search the human EST database on the ensemble web server (http://www.ensembl.org/index.html). If short RNA were targeting mRNA, then there would be an inverse correlation between short RNA accumulation and the expression levels of these mRNA. Several cDNA were identified with significant +/- match to the short 5S rRNA. SIAE was the most significant match to the ~1-22nt short 5S rRNA. The sialic acid acetylesterase (SIAE) mRNA contains a 90nt 5S rRNA like sequence in antisense orientation within its 3´ UTR. The alignment of the 5S consensus to this region of SIAE is shown in Figure 63.

```
5S       1  GTCTACGGCCATACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGG  60
            |||||||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||
SIAE  2177  GTCTACGGCCATATCACCCTGAACGCGCCTGATCTCATCTGATCTCGGAAGCTAAGCAGG  2118

5S      61  GTCGGGCCTGGTTAGTACTTGGATGGGAGA  90
            ||||||||||||||||||||||||||||||
SIAE  2117  GTCGGGCCTGGTTAGTACTTGGATGGGAGA  2088
```

**Figure 63- SIAE/5S rRNA alignment.** Plus/minus alignment of the 5S rRNA consensus to the 5SrRNA-like sequence within the 3´ UTR of SIAE mRNA (accession NM_170601).

Using the same RNA preparations from Hela and MCF-7 cell line preparations previously confirmed to express the short 5S rRNA (Figure 59), cDNA was synthesised. End-point RT-PCR was then used to estimate SIAE and GAPDH transcript levels (personal communication Jane Byrne). This revealed lower levels of SIAE transcripts in Hela cells when compared with MCF-7 samples. The GAPDH levels however remained constant (Figure 64B). This expression correlates with what would be expected if short 5S rRNA were targeting mRNA transcripts. However, the differential expression of mRNA in these two different cell lines could have been the result of unknown features unrelated to these 5S rRNA.

To resolve the problem associated with comparing different cell lines, RNA from the CaCo2 cells analysed in Figure 58 was use to create cDNA and the end-point RT-PCR was then repeated to estimate SIAE and GAPDH transcript levels (personal communication Jane Byrne). This showed the untreated CaCo-2 sample which expressed the short 5S rRNA to have the lowest SIAE transcript levels with GAPDH levels remaining constant (Figure 64D). This result was again consistent with the hypotheses of 5S rRNA targeting.

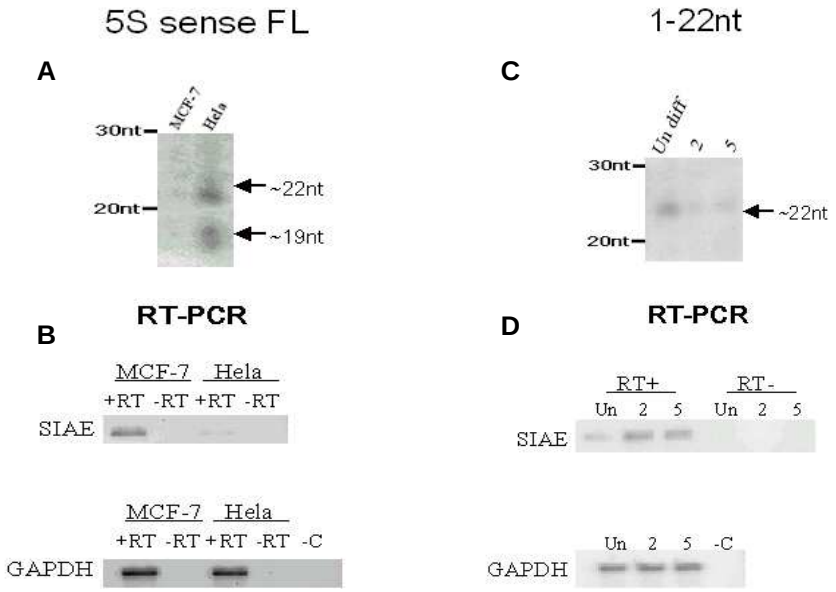


**Figure 64- End-point RT-PCR of SIAE mRNA in human cancer cell lines.** Hybridisation of the total RNA from the MCF-7 and Hela samples to the full-length sense-specific (A) and the CaCo-2 RNA samples to the 1-25nt sense-specific (C) 5S rRNA probes has been included to confirm the presence or absence of short 5S rRNA. 100ng/20µl of total RNA from MCF-7, Hela and three CaCo-2 samples was reverse transcribed using random hexamers. Equal amounts of the first strand cDNA for all samples was used to carried out PCR with SIAE and GAPDH specific primers (+RT). PCR of the non-reverse transcribed samples with the same primers was carried out in parallel as a negative control for any DNA contamination (-RT). Equal amounts of the MCF-7, Hela (B) and the three CaCo-2 (D) plus and minus PCR products were separated by agarose gel electrophoresis and stained with ethidium bromide (personal communication Jane Byrne). Comparison of the band intensity was used to indicate SIAE and GAPDH mRNA levels.

However, the RT-PCR of RNA obtained from cultured cells may not reflect what is happening *in vivo*. To observe if the SIAE mRNA transcripts are reduced in human tissue with elevated short 5S rRNA, the normal breast and breast tumour sample RNA (Figure 60) were used to synthesise cDNA. RT-PCR was then carried out to assess SIAE and GAPDH transcript levels (personal communication Jane Byrne). These RT-PCR data showed GAPDH levels to fluctuate across the samples making comparison of the SIAE mRNA levels difficult (Figure 65C). However, by

comparing the normal breast samples to tumour samples with a similar GAPDH levels a reduction in SIAE was observed to correlate with short 5S rRNA expression (Figure 65B).

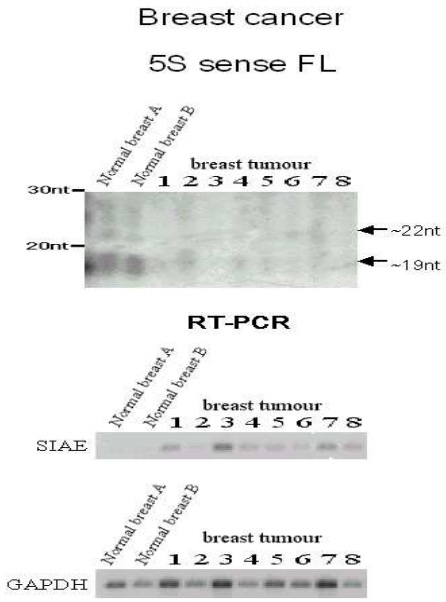


**Figure 65- End-point RT-PCR of SIAE mRNA in human breast tissue.** Hybridisation of the total RNA from the normal breast and breast tumours to the full-length sense-specific (A) has been included to confirm the short 5S rRNA expression. 100ng/20µl of total RNA from each sample were reverse transcribed using random hexamers. Equal amounts of the first strand cDNA for all of the samples were used to carry out PCR with SIAE and GAPDH specific primers. Equal amounts of the PCR products were then separated by agarose gel electrophoresis and stained with ethidium bromide (personal communication Jane Byrne). Comparison of the band intensity was used to indicate SIAE and GAPDH mRNA levels.

Following this analysis, quantitative RT-PCR of the CaCo2 samples was carried out (personal communication Jane Byrne). A 2.5 and 2 fold increase of SIAE mRNA was observed for the 2µM and 5µM butyrate treatments respectively in comparison to the untreated sample. This was a low but significant variation and was consistent with the observations made with non-quantitative RT-PCR.

## 6.4 Discussion

### *6.4.1 The unexpected detection of short 5S rRNA identified in embryonic cell lines.*

Despite the consistent sequencing of short 5S rRNA during investigations into 18-25nt RNA using MPSS and 454 sequencing, these sequences are consistently disregarded as unimportant degradation fragments (Lu et al., 2006, Mineno et al., 2006). This chapter describes the serendipitous detection of several discrete 5S rRNA, smaller in size than the ~120nt full-length transcript that occurs upon differentiation of ES and EC cells in culture. It particularly focuses on those RNA with the characteristic size of short interfering RNA (~20-26nt).

The ~20-26nt RNA were originally observed in differentiated mouse ES cells (male and female), when using a full-length 5S rRNA sense-specific probe as a positive control, for the hybridisation efficiency of blotted RNA (data not shown). Subsequent short 5S rRNA were also detected in the differentiated RNA of two other embryonic cell lines (mouse P19 EC (Figure 45) and human NT2 (Figure 46) cell lines). In all three embryonic cell lines examined, the short 5S rRNA were predominantly observed in the two-day differentiated samples (Figure 44, Figure 45 and Figure 46). These findings indicated the short 5S rRNA to be a general feature of mammalian embryonic cell lines that are differentiated in culture.

In addition to the 20-26nt 5S rRNA, RNA of an intermediate size (~90-30nt) were consistently observed in each of the embryonic cell lines analysed. For the mouse embryonic cell lines (ES and P19 EC) these intermediate 5S rRNA were predominantly observed in the two-day differentiated samples. These included prominent RNA of ~60nt and ~40nt derived from the 5` and 3` of the full-length 5S rRNA transcript respectively (Figure 44 and Figure 45). However, a quantifiable reduction in the full-length 5S rRNA transcript was not consistently observed to coincide with the induction of these shorter RNA. Intermediate RNA of a very similar size were also observed for the human NT2 embryonic carcinoma cell line. Unlike the intermediate RNA of the mouse embryonic cell lines, intermediate 5S rRNA were detected in all of the NT2 samples analysed in the differentiation time course (Figure 46). However, the banding pattern of

these intermediate 5S rRNA was slightly altered in the two-day differentiated sample, coinciding with the induction of the short RNA of ~22nt. This alteration to the banding pattern included the reduction of a ~70nt RNA, with the appearance of several smaller discrete RNA bands (including an RNA of ~22nt). This was of particular interest because ~70nt is the typical size of micro RNA precursors. Therefore the reduction of this ~70nt 5S rRNA, exclusively in the two-day differentiated sample, may indicate a miRNA-like processing of these 5S rRNA.

The observation of 5S rRNA, smaller in size than the ~120nt full-length transcript, especially those of a size characteristic of short interfering RNA was unexpected. However, this was not completely without precedent, because 5S rRNA have been reported to undergo specific nuclease cleavage *in vitro*, resulting in the fractionation and expulsion of 5S rRNA transcripts from the ribosome (Holmberg and Nygard, 2000). Also, the fragmentation of other rRNA has been reported in various eukaryotes undergoing apoptosis (specific cleavage) and necrosis (generalised cleavage) and cellular stresses such as heat-shock (generalised cleavage) (Hoat et al., 2006, Houge et al., 1995). However, it is unknown what role RNAi and non-RNAi pathways may have in the processing or degradation of these sequences.

To assess if the highly abundant short RNA processing observed for 5S rRNA was specific or general to all rRNA, the most abundant 18S rRNA and 28S rRNA sequences were identified in the database of short RNA present in mouse embryos and hybridised to ES and P19 EC cell RNA. This resulted in the detection of short 28S rRNA of ~19nt and ~28nt at every time point of embryonic differentiation (Figure 53). This suggests that the processing of ribosomal RNA may not be restricted to 5S rRNA. However, peak accumulation in two-day differentiated ES cell RNA was not a general feature of rRNA processing. This differential expression of the short 28S rRNA and the 5S rRNA is more consistent with sequence specific processing, than widespread ribosome degradation being the source of the short 5S rRNA. The specific processing of the 5S rRNA is further supported by the absence of smearing in the NT2 samples and the discrete nature of the RNA observed with a conserved banding pattern in the two-day differentiated embryonic cell lines. Given that 5S rRNA is highly regulated in relation to cell growth, any process controlling 5S rRNA levels through general or

RNAi mediated degradation could be very important in the cell. Therefore the short 5S rRNA (~20-26nt) were further investigated.

The short 5S rRNA (~20-26nt), detected with a full-length 5S rRNA probe in the two-day differentiated embryonic cells could have been from multiple sequences with a common size or a single abundant RNA. To define the 5S rRNA regions represented in the short RNA bands, mouse ES cell RNA was hybridised to overlapping oligonucleotide probes specific to both sense and antisense 5S rRNA consensus transcripts. From this analysis, two 5S rRNA of ~26nt and ~22nt were observed to be derived from the extreme 5´ and 3´ of the 5S rRNA sense transcript respectively (Figure 48) with a single 5S rRNA of ~26nt, derived from a central region of the 5S rRNA antisense transcript (Figure 49). It can be deduced from this, that the short RNA detected represented ssRNA from single regions of the 5S rRNA transcript, however it cannot be established whether these bands represent a single sequence or a set of identically sized but not completely overlapping RNA. Despite the short 5S rRNA appearing to be ssRNA, these could still be loaded into RISC and act through the RNAi pathway (Martinez et al., 2002). Alternatively the 5S rRNA may form dsRNA *in vivo*, however incorporation into the RISC may cause degradation of the complementary strand (Matranga et al., 2005) resulting in only the loaded strand being detected by hybridisation.

In addition to the short RNA, the oligonucleotide probes identified several associated intermediately sized RNA that are co-expressed and so may be precursors of the ~20-26nt RNA. Under the hypothesis that these RNA were processed to generate the short RNA, RACE sequencing was carried out on the sense-specific ~80nt RNA hybridised with the short 5S rRNA derived from the extreme 5` of the transcript. The sequence obtained indicated there to be no divergence from the 5S rRNA consensus sequence over this region. This information, although not able to specify the genomic location of the precise source 5S short rRNA transcripts, does suggest that the ~80nt RNA and, by inference, the short 5S RNA are not derived from diverged 5S rRNA pseudogenes. Also sequenced during the RACE analysis of the ~80nt 5S rRNA was a clone without the predicted short RNA of ~22nt identified with oligonucleotide 1 probe (Figure 48). Therefore a cleavage product of the ~80nt RNA may have been fortuitously sequenced, which represents the ~60nt RNA. The direct sequencing

of all the 5S rRNA bands (~20-90nt) would be required to confirm this and predict other cleavage points, during the processing of the 5S rRNA transcript.

Because of limitations in the hybridisation method used to identify the three short 5S rRNA regions in the ES cell RNA, it is possible that not all small 5S rRNA have been identified. In an attempt to assess the full extent of potential variation in the short 5S rRNA population of the embryonic cell lines assessed, the database of short RNA present in mouse embryos (Mineno et al., 2006) was screened. As discussed previously, this dataset was used because the short RNA expressed in mouse embryos may be similar to those expressed in the mouse ES cells. From this analysis many short 5S rRNA sense-specific sequences were identified, however unlike the ES RNA, no antisense-specific 5S rRNA were observed. These sequences, although observed to be derived from across the 5S rRNA sequence, were most abundant for specific regions. Like the short 5S rRNA detected in the ES cell RNA, the regions of highly abundant short 5S rRNA mapped to the extreme 5` and 3` of the 5S rRNA sense transcript. The sense-specific RNA of the other regions may also be expressed in the mouse ES cell line however this was not investigated further.

Despite various RNA sequences with different cleavage points mapping to the 3` of the sense transcript, it is likely that these would all have sufficient complementary to be detected by the sense-specific DNA oligonucleotide probe 8. Therefore this may indicate that the population of short 5S rRNA hybridised by the sense-specific oligonucleotide 8 probe (Figure 48) in the ES RNA may be quite heterogeneous. The identification of the short 5S rRNA sequences in the database of short RNA present in mouse embryos also confirmed the chemistry of the short RNA to be suitable for currently used techniques of short RNA detection. In addition the identification of 5S rRNA through the deep sequencing of RNA from mouse embryos argues against them being a cell culture artifact.

Subsequent analysis of the short 5S rRNA focused on their potential interaction with the RNAi pathway. First the potential processing of 5S rRNA into short RNA by the RNase III enzyme Dicer was assessed. Through the use of Human recombinant Dicer *in vitro,* it was shown that the intermediate 5S rRNA could be processed into short 5S rRNA. Interestingly the *in vitro* Dicer processing of these RNA was observed to generate some short RNA that were not detected by

northern blotting of RNA from ES cells. The *in vitro* processing of the 50-80nt 5S rRNA by Dicer supports the hypothesis that the intermediately sized RNA may be precursors of the smaller ~20nt RNA and validates the RACE analysis of the ~80nt sense-specific 5S rRNA. In addition to this they confirm the intermediate 5S rRNA to have dsRNA structure (Bernstein et al., 2001) but the precise sequence and structure of the dsRNA remains to be determined.

Dicer-null ES cells were assessed for the accumulation of short 5S rRNA during differentiation. From this analysis a substantial, but not complete reduction of short 5S rRNA and associated intermediate RNA was observed (Figure 57). Therefore this result suggests that 5S rRNA processing is not exclusively undertaken by Dicer. These results also implicate a possible function of Dicer in the generation of the intermediate 5S rRNA. However, despite having a 2nt 3´ overhang thought to be required for Dicer recognition through the PAZ domain (Zhang et al., 2004), the secondary structure of ribosome bound 5S rRNA does not have an extensive region of dsRNA, which would be expected in a typical Dicer substrate (Zhang et al., 2004). If 5S rRNA sense transcripts are processed by Dicer *in vivo*, the concept of what makes a Dicer substrate would require expansion. Alternatively Dicer may process double stranded 5S rRNA resulting from the association of complementary sense and antisense transcripts into the short 5S rRNA detected.

Due to the number of 5S rRNA copies in the mammalian genomes (Drouin, 2000, Sorensen and Frederiksen, 1991, Szymanski et al., 2002) and the widespread transcription reported in mammals (Katayama et al., 2005) antisense 5S rRNA transcripts would not be unexpected. Transcripts containing antisense 5S rRNA sequence could therefore theoretically form dsRNA with 5S rRNA sense transcripts. However, the hybridisation of a precisely sized ~120nt antisense 5S rRNA (oligonucleotide probe 5), identical in size to the 5S rRNA sense transcript in ES cell RNA was unexpected. The possibility that this may have resulted from the hybridisation of the antisense-specific probe to the sense-specific 5S rRNA probes remaining on the membrane was quickly eliminated by hybridising a previously unused membrane (Figure 50). The cross-hybridisation of the antisense-specific to the 5S rRNA sense transcript was also judged unlikely due to the lack of complementarity of the antisense-specific probe to the sense 5S rRNA consensus sequence (Figure 49D) and the hybridisation and wash conditions

tolerated (Figure 51). Thus leaving the genuine hybridisation of antisense-specific RNA as the most likely explanation of the ~120nt band. However, the size of the ~120nt band and the fact that it is identical in size to the 5S rRNA sense transcript, would suggest this not to be derived from read-through transcription.

The use of the sense 5S rRNA transcript as a template for RdRP could however generate a complementary antisense RNA of ~120nt. The hypothesis is speculative because no mammalian RdRP has yet been reported, despite its requirement in the RNAi pathway of other organisms (Lipardi et al., 2001, Martienssen, 2003, Sijen et al., 2001). However, the hybridisation of probes specific to the extreme 5` and 3` of the hypothetical full-length 5S rRNA indicated that the antisense molecule is not fully complementary with the sense RNA (Figure 51). This result could be explained in the context of a RdRP model where post-transcriptional modification carried out by ADAR that converts A to I in regions of dsRNA. Alternatively the lack of significant hybridisation to the extreme 5` and 3` of the hypothetical full-length 5S rRNA antisense RNA may suggest the transcription of a ~120nt RNA only partly containing 5S rRNA antisense sequence. The 5S small nucleolar RNA (snoRNA) provide a known source of RNA, which are known to contain antisense 5S rRNA sequence (Chang et al., 2002). Typically snoRNA range in size from 60-300nt are involved in the post-transcriptional modification of many sequences including tRNA and rRNA. They mediated this by forming direct base-pairing interactions with the target sequence and so contain sequence antisense to that of the target sequence for review see (Kiss, 2002).

There have been few studies on 5S snoRNA, however five 5S snoRNA have been described in the human genome (Chang et al., 2002). Upon aligning these sequences to the 5S rRNA consensus it was observed that nucleotides complementary to 1-90 of the sense transcript were largely retained with 87-92% complementarity. This region contains sequence complementarity to the antisense-specific DNA oligonucleotide probe 5, thus hybridisation of a 5S snoRNA may explain the ~120nt antisense 5S rRNA band. The detection of a 5S snoRNA at ~120nt could also explain the lack of significant hybridisation to probes specific to the extreme 5` and 3` of the hypothetical full-length 5S rRNA antisense consensus transcript (Figure 51). However, of the five human snoRNA

identified, none were reported to be of ~120nt. The ~120nt RNA may therefore represent a currently unidentified 5S snoRNA, which may or may not be conserved between human and mouse.

To definitively establish the nature of the ~120nt RNA with antisense character the sequence must be determined. This could be achieved by 5` and 3` RACE of the ~120nt RNA using the antisense-specific oligonucleotide 5 probe as the primer. However, even if this RNA contains genuine antisense 5S rRNA sequence, its detection at each point of ES cell differentiation examined, would indicate that its presence alone is not sufficient to cause short RNA accumulation. Therefore other factors that control the ability of this sequence to form dsRNA or access of the dsRNA to processing enzymes.

### *6.4.2 The detection of short 5S rRNA identified in human cancer cell lines and breast tissue.*

Following the observations in embryonic cells, discrete short 5S rRNA of ~19nt and 22nt were also detected in the three human cancer cell lines CaCo-2, MDA-MB-435 and Hela (Figure 58 and Figure 59). It was also observed that in these human cancer cell lines expressing the short 5S rRNA, the predominant ~60nt and ~40nt intermediate RNA previously detected in the embryonic cell lines (Figure 44, Figure 45 and Figure 46) were conserved. The cancer cell line RNA accumulating the short 5S rRNA also had no signs of increased RNA degradation (smearing), further supporting the specific processing of these RNA.

The detection of short 5S rRNA in a breast cancer cell line prompted the analysis of breast cancer tissue RNA, to assess if short 5S rRNA also accumulated *in vivo*. From this analysis short 5S rRNA were observed. However, unexpectedly the short 5S rRNA were detected predominantly in the control samples representing "normal" breast tissue. It was unknown if this result indicated genuine differences in the accumulation of the short 5S rRNA in normal and cancer breast tissue. Because there were several differences between these samples, other than their cancer status. First there was a significant age difference between the "normal" and cancer tissue donors. In addition to this there was uncertainty about the time elapsed between the harvest and freezing of tissue samples in liquid nitrogen of cancerous and non cancerous material. Thus the detection of

the short 5S rRNA in the control samples could be explained by delayed freezing of "normal" samples in liquid nitrogen in comparison to tumour samples, resulting in the partial degradation of the control RNA by nucleases contained within the cells. The ethidium bromide stained image would argue against the excessive degradation of RNA in these samples, however this cannot distinguish subtle differences in RNA quality. The indications from the miRNA hybridisations would however support short RNA integrity with the reported up-regulation of hsa-mir-21 in human breast cancer observed (Volinia et al., 2006) in comparison to the control samples (Figure 60).

The potential targeting of the short 5S rRNA could have major implications in cellular regulation because 5S rRNA is critical to the growth rate of the cell. In addition to this there are many other sequences expressed in the human and mouse genome containing sequences complementary 5S rRNA. Therefore the potential recruitment of short 5S rRNA into the RNAi pathway was tested. For this, the method of short RNA validation by sensor constructs was used. Several sensor constructs were generated placing sequence that were complementary to the mapped 5S short RNA in the 3´UTR of the firefly luciferase gene. This approach did indicate potential targeting of the 1-25nt 5S rRNA and hsa-mir-16 miRNA sensor constructs, however a less than 2 fold difference was observed when compared with the control samples (Figure 61). Results similar to these have been published in support of short RNA targeting (O'Donnell et al., 2005). However, these variations were much smaller than the extent of hsa-mir-21 targeting observed (Figure 62). In fact the results obtained with the hsa-mir-21 sensor construct indicates hsa-mir-21 to be a very effective repressor. This observation may therefore have implications when analysing the function of hsa-mir-21, in situations where it is found to be up regulated e.g. breast cancer.

The mRNA encoding SIAE has a near perfect 90nt insertion of the 5S rRNA sequence in the antisense orientation of it's 3`UTR. This is conserved in chimpanzees but not other primates or less closely related mammals. Such an insertion would make this mRNA a very likely target of short 5S rRNA if they were at all capable of directing sequence specific down-regulation of gene expression. Using RT-PCR, the mRNA levels of hypothetical naturally occurring target SIAE were compared in RNA samples known to contain different amounts of short 5S rRNA. In each case there was a reciprocal relationship of the amount

of short 5S rRNA to the SIAE mRNA. This is consistent with targeting but more evidence is required.

# 7

# Discussion

# 7.1 Discussion

## *7.1.1 The identification of repeat-associated short RNA of ~19-26nt*

The identification of short (~19-26nt) transposon-specific RNA has been associated with the repression of transposable elements in *D.melanogaster, C.elgans* and plants (Aravin et al., 2001, Lippman et al., 2003, Sijen and Plasterk, 2003). However, prior to the commencement of this project there had been no such RNA reported in mammals. Transposons account for almost half of the DNA contained within the human (Lander et al., 2001) and mouse (Waterston et al., 2002) genomes and therefore represent the largest group of sequences in these genomes. As in other species almost all mammalian transposons become condensed into transcriptionally repressed heterochromatin (Martens et al., 2005), however the mechanism by which this is achieved is currently poorly characterised. With the current support for the RNAi mediated suppression of transposon sequences in other species and the emerging importance of RNAi in mammals the possible interaction of these processes was investigated.

Using a candidate-lead approach, four defined transposon-specific short RNA (~20-26nt) were detected specifically in differentiated mouse embryonic cell lines using two differentiation protocols (Figure 24, Figure 35 and Figure 36). These RNA were specific to the murine L1 $T_F$-monomer (hybridised with a ~200nt probe) and B2 transposon sequences (hybridised with 20nt probes). A discrete RNA band of ~19nt that could be visualised simply by ethidium bromide staining was frequently observed, specifically in the differentiated embryonic samples expressing the short RNA. Serendipitously, use of the differentiated mouse embryonic cell samples also lead to the identification of three highly abundant short 5S rRNA (hybridised with 25nt probes) (Figure 48 and Figure 49). The 5S rRNA were also subsequently identified in differentiated human embryonic carcinoma cells (NT2) (Figure 46), some cancer cell lines (Figure 58 and Figure 59) normal and cancerous breast tissue (Figure 60).

These previously unreported repeat-associated RNA were all observed in various embryonic cell samples, predominantly after two-days of cultured embryonic cell differentiation. Throughout this thesis the embryonic cells have been

referred to as undifferentiated and differentiated. This terminology has been adopted because the various treatments used on the embryonic cell lines (see materials and methods) were intended to and had previously been reported to induce differentiation of cells in culture (Freemantle et al., 2002, Niwa, 2001). However, markers of embryonic cell differentiation were not assessed as this technique was primarily used to induce the accumulation of the short RNA and not for comparison to *in vivo* development. Therefore it is unknown if the induction of the short repeat-associated RNA is as a result of genuine differentiation or another process induced by the treatment of the cells in culture. The identification of the short B2 and 5S rRNA in a dataset derived from the MPSS of 18-25nt 9.5, 10.5 and 11.5 day mouse embryo RNA (Mineno et al., 2006) supports the idea of genuine expression of these short RNA. If the short repeat-associated RNA were confirmed to be expressed as in the embryonic cell lines, this would support a potential role of these RNA in the repression of transposable elements and other repeat sequences, during mammalian development.

RNA quality is critical to the reliability of any RNA-based analysis and can be negatively affected by the presence of nuclease, which can easily contaminate samples from laboratory equipment and surroundings. The discrete size of the ~19nt RNA band made it unlikely to be the result of random RNA degradation by nucleases (cellular or external) because these degrade RNA into single nucleotides unless protected by dsRNA structure (Lung et al., 2006). However, the RNA samples expressing the short repeat-associated RNA were also observed to have a more general smeared appearance in some instances, which could indicate RNA degradation.

There are several other reasons to think that the ~19nt RNA are not a result of random contaminating nuclease degradation of these samples. The observation of smearing predominantly in a specific sample (two-day differentiated) would be unlikely if external contamination of, for example plasticware was the cause of smearing as all samples would be exposed to equal risk. Also the discrete appearance of the short repeat-associated RNA (Figure 24, Figure 35, Figure 36, Figure 48, Figure 49) and the conserved banding patterns observed for the intermediate RNA detected in different cell lines and tissue types (Figure 44, Figure 45, Figure 46, Figure 58, Figure 59 and Figure 60) argue against their

production by generalised nuclease degradation. In addition the short 5S rRNA is observed to be co-induced with the other repeat-associated short RNA in a number of samples which do not have a smeared appearance (Figure 46 and Figure 58), supporting the specific processing of these RNA.

The reproducible pattern of RNA fragmentation, associating with a particular treatment (differentiation) and production of discrete RNA argue for some degree of specificity in the action. An alternative to the generalised nuclease degradation of these samples as mentioned previously, could be the induction of specific cellular RNA processing by the differentiation protocols used, resulting in the somewhat smeared appearance of the RNA sample when separated by gel electrophoresis. There are several cellular processes, which could result in increased RNA turnover by specific processing in the cell. These include the induction of cellular stress, apoptosis and necrosis. The accumulation of the short repeat-associated RNA by these cellular processes could be easily assessed by their specific induction followed by analysis of the RNA for the short repeat-associated RNA. This possibility was not investigated at the time. Instead priority was given to defining the short RNA accumulating with the differentiation treatments and assessment of any gene silencing properties. However, further knowledge of the cellular processes involved in their processing would assist in further understanding under what conditions these RNA are expressed in mammals and how they may function.

Evidence was obtained supporting the involvement of the RNAi enzyme Dicer, in the processing of the B2 (Figure 37) and 5S rRNA (Figure 56 and Figure 57). Firstly, human recombinant Dicer processed gel purified intermediate transcripts of the short 5S rRNA into ~20nt RNA *in vitro*. Dicer is a dsRNA specific nuclease III enzyme (Zhang et al., 2004), therefore it can be deduced from this result that the intermediate 5S RNA have some double stranded character and can be processed in a manner similar to known inducers of RNAi . Further to this, it was observed that both the 5S rRNA and B2 short RNA accumulated to substantially lower levels in the Dicer-null ES cells compared to the wild type ES cells. Similar reductions of the associated intermediate RNA were also observed. This further supports the hypothesis that these RNA may be involved in RNAi-like processes.

If Dicer were entirely responsible for the processing of the short repeat-associated RNA, complete elimination of the RNA would have been expected as was observed for the micro RNA control used. The reduction of mmu-mir-292as to below detectable levels suggests the mutant cells were functionally null for Dicer. However, both of the repeat-associated short RNA appeared to be more abundant than that micro RNA in the control cell lines. Therefore perhaps the lack of micro RNA signal in the Dicer-null samples was due to the micro RNA levels falling below the detection threshold of the hybridisation method and not the complete elimination of micro RNA processing. This possibility could be excluded if a micro RNA of comparable expression to the repeat-associated RNA in the control embryonic cell line was used as a positive control. However, because mmu-mir-292as is reported to be one of the highest expressed micro RNA in these cells this may prove difficult.

An alternative theory is that Dicer-dependant processing is only partly responsible for the generation of the short RNA, indicating there to be a low level of Dicer-independent processing of these RNA. However, this would require the Dicer-independent processing to be capable of generating a cleavage pattern identical to that of Dicer. An obvious candidate for this potential Dicer-independent processing would be another member of the RNase III family. This is because the similarity of the substrates cleaved by this family of enzymes (dsRNA) may create some redundancy in the processing of the repeat-associated RNA (MacRae and Doudna, 2007). However, if another RNase III was capable of this, it would be equally possible that the other RNase III enzyme alone processed the short repeat-associated RNA. In the latter case the reduction of 5S and B2 short RNA in Dicer null cells would then be the result of reduced RNase III activity as a consequence of it now also carrying out reactions normally catalysed by Dicer.

Despite the indication that the short repeat-associated RNA could be processed by Dicer, conflicting evidence for the targeting ability of the short repeat associated RNA has been obtained (Figure 61,Figure 62,Figure 63,Figure 64 and Figure 65). However, the extent of this analysis has been limited by time constraints. Therefore the main focus for the progression of this work would involve a more extensive investigation of the potential targeting and interaction of these RNA with components of the RNAi pathway. This could involve using a

protein pull-down assay of the various argonaute proteins to assess if the repeat-associated RNA identified are found to associated with these key components of the RNAi pathway, as carried out by (Aravin et al., 2007) as well as more extensive use of sensor constructs.

### *7.1.2 The identification of ~70-90nt repeat associated RNA.*

Because of the increased understanding of the diversity and function of small RNA, this thesis has predominantly focused on the identification of mammalian short repeat-associated RNA ~20-25nt in size. However, recent work has indicated there to be an even more extensive RNA world (Aravin et al., 2006, Girard et al., 2006). In the course of this investigation, discrete RNA of ~70-90nt of several types of retrotransposon, including those not found to express short RNA of ~20-26nt were unexpectedly detected. These were expressed in both mouse and human, from a range of RNA extracted from cell lines and tissue and did not correspond to any form of conventional transcript of such elements.

The retrotransposons that express these ~70-90nt RNA include members of the SINE (Alu, B1 and B2) (Figure 10, Figure 11, Figure 30 and Figure 35), LINE (L1 A, $G_F$ and $T_F$ monomers) (Figure 22 and Figure 23 and Figure 24) and LTR (IAP and MuERV) (Figure 28 and Figure 29) classes. These are summarised in Table 16 along with the 20-25nt RNA identified in this study. For some of the transposable elements both sense and antisense ~70-90nt RNA were detected (B1, B2, L1 A and L1 $T_F$) with others only being clearly detected as a single polarity (Alu, L1$G_F$, IAP and MuERV). In each case these RNA were too short to be explained by the full-length transcript, so are thought not to represent standard transcription products of the elements. Additionally RNA of this size were detected during the hybridisation of 5S rRNA probes to various RNA samples (Figure 44, Figure 45, Figure 46, Figure 58, Figure 59 and Figure 60), although these unlike, the transposon RNA, were exclusively detected in samples also containing short ~20-26nt RNA.

| Transposon type | Sub type | Size | Strand | Organism |
|---|---|---|---|---|
| LINE | L1-A | ~65-70nt | Sense and antisense | Mouse |
| | $L1\text{-}G_F$ | ~70-80nt | Antisense | Mouse |
| | $L1\text{-}T_F$ | ~70-80nt | Sense and antisense | Mouse |
| | | ~19nt | Antisense | Mouse |
| SINE | Alu | ~70nt | Antisense | Human |
| | B1 | ~90nt | Sense and antisense | Mouse |
| | | ~70nt | Antisense | Mouse |
| | B2 | ~90nt | Sense and antisense | Mouse |
| | | ~30-70nt | Sense | Mouse |
| | | ~18-20nt | Sense and antisense | Mouse |
| LTR | IAP | ~70nt | Antisense | Mouse |
| | MuERV-L | ~70nt | Antisense | Mouse |

**Table 16- Table of transposon RNA identified by northern blot.** This table summarises the various transposon specific short RNA identified by northern blot that have been described in chapters 3,4 and 5.

The majority of the retrotransposon derived RNA detected were ~70nt. Therefore it could be argued that the retrotransposon RNA detected were as a result of spurious cross-hybridisation to the highly abundant tRNA in this region and not genuine hybridisation with ~70nt retrotransposon sequence. However, this seems improbable, as it would require the cross-hybridisation of all of the retrotransposon probes (which have highly varied sequences) to the tRNA under the hybridisation conditions used. In addition, the $L1G_F$ antisense-specific probe (Figure 23) was found to differentially hybridise an ~80nt RNA in the male ES cell samples compared with the ~70nt RNA hybridised in the female ES cell samples. This differential hybridisation can be simply explained by polymorphism because the male and female cell lines were of different mouse strains. Polymorphism of transposon sequences within species is well documented (Roy-Engel et al., 2001) however as tRNA is expected to be, under strong purifying selection against polymorphism formation, undermining cross-hybridisation to the tRNA as a potential explanation of these RNA. Nevertheless sequencing is required to unequivocally identify these RNA as transposon derived.

Based on the hypothesis that the ~70-90nt RNA identified were as a result of the genuine hybridisation of retrotransposon sequences, attempts were made to observe if these RNA acted like miRNA precursors. In the case of the B2 ~90nt similarities to the miRNA were observed, with it becoming reduced during the production of short RNA of ~20nt (Figure 35), thus indicating that this may well act as short RNA precursor. However, several of the other RNA identified did not to associate with defined short RNA (Figure 10, Figure 11, Figure 22, Figure 23, Figure 28, Figure 29 and Figure 30) under the conditions used. The RNase

protection and *in vitro Dicer* digestion of these RNA would be pursued in future work to obtain more information on the potential structure and function of these RNA.

The RNA, which did not associate with short RNA of ~20-26nt were overlooked in this thesis in favour of investigating the possible RNAi regulation of mammalian transposable elements. However, in retrospect this may have been an error because it was the ~70-90nt RNA and not the ~20-26nt RNA that were consistently identified with the retrotransposons-specific probes (see Table 16). One possibility that has not yet been addressed is whether these longer ~70-90nt RNA may be important in the repression of mammalian transposable elements. RNA of ~70-90nt have not been observed to be involved in the repression of transposable elements in other species. Despite this, RNA of this size range have been indicated to be involved in RNA dependant heterochromatin formation in *D.melanogaster* (Wang et al., 2005). This is proposed to be mediated through the association of vigilin proteins (which are conserved in mammals) with ~75-100nt RNA of low secondary structure which is often inosine modified.

Whether the RNA identified in this study associate with the RNAi pathway to mediate the repression of transposable elements in mammals (as was the original hypothesis) has yet to be determined. However, this study has made significant progress in the field through consistently identifying RNA of ~70-90nt, despite few RNA of the predicted size (~20-25nt) being identified. The inability to consistently identify transposon-specific RNA of 20-25nt is not consistent with the original hypothesis of RNAi mediated suppression of transposons in mammals. However, there are caveats in the approach used, which may have prevented the observation of such RNA, if indeed they do exist. These include the use of generalised consensus probes and the limited sources of RNA examined.

Given the developing importance placed on the role of non-coding RNA in gene regulation, the identification of any small transposon-specific RNA accumulating in mammalian cells does raise the question as to their function, if not to suppress transposable elements. This is particularly true of the ~70-90nt transposon-specific RNA which could not directly act through the hypothesised RNAi pathway, but were identified for every transposon analysed (mouse and human). Could it therefore be the case that this study has lead to the

identification of a new class of regulatory RNA ~70nt in size, which are themselves directly involved in the regulation of mammalian transposon RNA? If this were the case perhaps there are several other, yet to be identified pathways through which non-coding RNA of varying sizes act to regulate genes adding further complexity to the system of gene regulation.

The progression of this work requires the exact sequences of these RNA to be known. This could be achieved by cloning or RACE sequencing. Knowledge of the exact sequences of these RNA could assist in further understanding their processing and allow them to be specifically antisense targeted for knockdown to assess their effect on transposon transcript levels.

### *7.1.3 Conclusion*

This thesis aimed to explore the potential involvement of the RNAi pathway in the silencing of mammalian transposons, through the detection of short transposon-associated RNA of ~20-26nt. The principle strategy adopted for this investigation involved using a modified gel blotting technique with a candidate lead approach. This was justified by the identification of several novel small RNA from repetitive elements.

It is currently unknown if the discrete short repeat-associated RNA described here are involved in the suppression of repeat elements, transcriptionally (promoter specific) and/or post-transcriptionally (transcript specific) through an RNAi mediated mechanism. However, what is apparent is that they are among the first of an increasing number of short repeat-associated RNA being identified in mammals. In addition to this, the ability to specifically induce their expression by the differentiation of embryonic cell lines in both mouse and human may prove useful for the identification of many more short repeat-associated RNA. Other recently described discrete short RNA identified in mammals include those of the human LINE 1 transposon (Yang and Kazazian, 2006) and the repeat-associated piRNA (Aravin et al., 2007). The association of these RNA with proteins of the RNAi pathway also support the original hypothesis that mammalian transposable elements may be silenced through an RNAi mediated pathway.

In addition to the identification of novel small RNA, this work validates the use of the modified gel blotting technique. This showed it to be sufficiently sensitive for the detection of endogenously expressed short RNA in mammalian cells, with the best results being obtained with short high specific activity probes. Given the benefit of gel blotting over other short RNA detection methods, this technique may become more widely used in the field of short RNA validation.

# Appendices

## Appendix I: L1 related RNA identified in the database of short RNA present in mouse embryos

| L1 sequence | Regions of identity & % | | | | Embryonic day (TPM) | | |
|---|---|---|---|---|---|---|---|
| | Strand | L1 | Sequence | % | E 9.5 | E10.5 | E11.5 |
| GAAGAGAGCTTGCCTGCAGAGA | Sense | 1689-1710 | 1-22 | 100 | 2 | 14 | 0 |
| AGAAGAGAGCTTGCCTGCAGAG | Sense | 1688-1709 | 1-22 | 100 | 9 | 0 | 0 |
| ATGGAAATGAACAAAACCATAC | Sense | 2098-2119 | 1-22 | 100 | 9 | 0 | 0 |
| CTGCACCTGAGATTCTCTCTTC | Antisense | 2245-2266 | 1-22 | 100 | 0 | 11 | 7 |
| AATCTCAGGTGCAGAAGATTCC | Sense | 2253-2274 | 1-22 | 100 | 0 | 10 | 0 |
| AGTTTGATTATTTCCTGCCATC | Antisense | 3981-3999 | 1-22 | 100 | 0 | 11 | 0 |
| CGAGTTTGATTATTTCCTGCCG | Antisense | 3980-4000 | 1-22 | 100 | 0 | 8 | 0 |
| AACATATCCTGAAGAAATCCAA | Sense | 4174-4195 | 1-22 | 95 | 9 | 0 | 0 |
| CAACTTTGGTACCTGGTATCTG | Antisense | 4273-4294 | 1-22 | 100 | 0 | 10 | 0 |
| TCCTCAATAAAATTCTTGCTAA | Sense | 4596-4617 | 1-22 | 95 | 0 | 0 | 9 |
| AAAAATTCCACCAGAGAACTCC | Sense | 5105-5125 | 1-22 | 100 | 13 | 0 | 0 |
| CAAGTCAATGGCCTTTCTCTAC | Sense | 5179-5199 | 1-22 | 100 | 7 | 8 | 4 |
| TCCCATGCTCATGGATTGGCA | Sense | 5376-5396 | 1-21 | 100 | 0 | 0 | 10 |
| GGCACAGGGGAAAAATTCCTGA | Sense | 5936-5957 | 1-22 | 100 | 19 | 0 | 2 |
| AATTCCTGAACAGAACAGCAAT | Sense | 5949-5970 | 1-22 | 100 | 0 | 8 | 0 |
| CAGCACAAGCCATTGCTGTTCT | Antisense | 5960-5981 | 1-22 | 100 | 9 | 0 | 0 |
| AAATGGGACCTAATGAAACTCC | Sense | 5999-6020 | 1-22 | 100 | 11 | 0 | 0 |
| CTCAGGTGAGAATTCTTTGTTC | Antisense | 6199-6220 | 1-22 | 100 | 0 | 9 | 0 |
| CCTCAGGTGAGAATTCTTTGTT | Antisense | 6200-6221 | 1-22 | 100 | 0 | 13 | 0 |
| CTGCCATTCGGTATTCCTCAGG | Antisense | 6215-6236 | 1-22 | 100 | 9 | 5 | 18 |
| TCTGCCATTCGGTATTCCTCAG | Antisense | 6216-6237 | 1-22 | 100 | 0 | 14 | 5 |
| CTCTGCCATTCGGTATTCCTCA | Antisense | 6217-6238 | 1-22 | 100 | 4 | 8 | 6 |
| CTCAGCCATTCGGTATTCCTCA | Antisense | 6217-6238 | 1-22 | 95 | 4 | 9 | 0 |
| TCTGCCATTCGGTATTCCTCGG | Antisense | 6218-6237 | 1-22 | 100 | 0 | 8 | 0 |
| CTTCTCAGCCATTCGGTATTCC | Antisense | 6220- | 1-22 | 95 | 0 | 13 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 6241 | | | | | |
| CTTCTCTGCCATTCGGTATTCC | Antisense | 6220-6241 | 1-22 | 100 | 9 | 0 | 0 |
| TTTGATTTGCATTTCCCTGA | Antisense | 6273-6292 | 1-20 | 100 | 0 | 9 | 0 |
| GAATCTCAGGGTTGTTTTGATT | Antisense | 6286-6307 | 1-22 | 100 | 11 | 14 | 0 |
| CTCTGGAAATCAGTCTGGCGGT | Sense | 6428-6449 | 1-22 | 100 | 0 | 0 | 9 |
| GAAATCAGTCTGGCGGTTCCTC | Sense | 6433-6454 | 1-22 | 100 | 13 | 13 | 0 |
| CAGCAATACCTCTCCTGGGCAC | Sense | 6486-6506 | 1-22 | 100 | 2 | 6 | 14 |
| CTATGTTCATAGCAGCCTTATT | Sense | 6552-6573 | 1-22 | 100 | 0 | 13 | 0 |
| CATTCCTCTGTTGAGGGGCATC | Antisense | 6603-6624 | 1-22 | 100 | 9 | 13 | 13 |
| CCTCTGTTGAGGGGCATCT | Antisense | 6602-6620 | 1-19 | 100 | 4 | 7 | 11 |
| TTCCTCTGTTGAGGGGCATCTG | Antisense | 6602-6622 | 1-21 | 100 | 13 | 2 | 6 |
| ATTCCTCTGTTGAGGGGCATCT | Antisense | 6602-6623 | 1-22 | 100 | 0 | 11 | 2 |
| TCCTCTGTTGAGGGGCATCT | Antisense | 6602-6621 | 1-20 | 100 | 13 | 0 | 0 |
| ATTCCTCTGTTGAGGGACATCT | Antisense | 6602-6623 | 1-22 | 95 | 0 | 10 | 0 |
| TAGTCTCACAAGAGACAGCTAT | Antisense | 7067-7087 | 1-21 | 100 | 4 | 8 | 4 |
| TAGTCTCACAAGAGACAGCTA | Antisense | 7067-7087 | 1-21 | 100 | 0 | 8 | 0 |
| GATGCTCACAGTCAGCTATTGG | Sense | 7115-7136 | 1-22 | 100 | 11 | 13 | 0 |
| TGCTCACAGTCAGCTATTGGAT | Sense | 7117-7138 | 1-22 | 100 | 0 | 11 | 5 |
| ATCCATCCATTAGCTGACTGTG | Antisense | 7121-7142 | 1-22 | 100 | 0 | 8 | 0 |
| CCCCCAATGGAGGAGCTAGAGA | Sense | 7152-7171 | 3-22 | 100 | 0 | 6 | 10 |

**Figure 66- L1 related RNA identified in the database of short RNA present in mouse embryos.** The sequences are displayed in the column labelled “L1 sequence” with nucleotides polymorphic to the L1 consensus sequence indicated in red. Sequences are arranged by alignment to the L1 consensus with the region of alignment indicated (L1) from 5´ to 3´ the monomer, 5´ UTR ORF and 3´ UTR regions have also been defined by different colours. The region of sequence identity to L1 transcript is also shown in the “sequence” column, with the percentage identity between the sequence and alignments indicated in the “%” column. The TPM value of each sequence at the 9.5 day, 10.5 day and 11.5 day (9.5,10.5 and 11.5) embryonic time points is also shown.

# References


ABRUSAN, G. & KRAMBECK, H. J. (2006) Competition may determine the diversity of transposable elements. *Theor Popul Biol,* 70**,** 364-75.

ADEY, N. B., SCHICHMAN, S. A., GRAHAM, D. K., PETERSON, S. N., EDGELL, M. H. & HUTCHISON, C. A., 3RD (1994) Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Mol Biol Evol,* 11**,** 778-89.

ADEY, N. B., TOLLEFSBOL, T. O., SPARKS, A. B., EDGELL, M. H. & HUTCHISON, C. A., 3RD (1994) Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc Natl Acad Sci U S A,* 91**,** 1569-73.

AGRAWAL, N., DASARADHI, P. V., MOHMMED, A., MALHOTRA, P., BHATNAGAR, R. K. & MUKHERJEE, S. K. (2003) RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev,* 67**,** 657-85.

ALLEGRUCCI, C., THURSTON, A., LUCAS, E. & YOUNG, L. (2005) Epigenetics and the germline. *Reproduction,* 129**,** 137-49.

ARAVIN, A., GAIDATZIS, D., PFEFFER, S., LAGOS-QUINTANA, M., LANDGRAF, P., IOVINO, N., MORRIS, P., BROWNSTEIN, M. J., KURAMOCHI-MIYAGAWA, S., NAKANO, T., CHIEN, M., RUSSO, J. J., JU, J., SHERIDAN, R., SANDER, C., ZAVOLAN, M. & TUSCHL, T. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*.

ARAVIN, A. A., NAUMOVA, N. M., TULIN, A. V., VAGIN, V. V., ROZOVSKY, Y. M. & GVOZDEV, V. A. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol,* 11**,** 1017-27.

ARAVIN, A. A., SACHIDANANDAM, R., GIRARD, A., FEJES-TOTH, K. & HANNON, G. J. (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science,* 316**,** 744-7.

BARTEL, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* 116**,** 281-97.

BENTWICH, I. (2005) A postulated role for microRNA in cellular differentiation. *Faseb J,* 19**,** 875-9.

BERALDI, R., PITTOGGI, C., SCIAMANNA, I., MATTEI, E. & SPADAFORA, C. (2006) Expression of LINE-1 retroposons is essential for murine preimplantation development. *Mol Reprod Dev,* 73**,** 279-87.

BERNSTEIN, E., CAUDY, A. A., HAMMOND, S. M. & HANNON, G. J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature,* 409**,** 363-6.

BERNSTEIN, E., KIM, S. Y., CARMELL, M. A., MURCHISON, E. P., ALCORN, H., LI, M. Z., MILLS, A. A., ELLEDGE, S. J., ANDERSON, K. V. & HANNON, G. J. (2003) Dicer is essential for mouse development. *Nat Genet,* 35**,** 215-7.

BOEKE, J. D., GARFINKEL, D. J., STYLES, C. A. & FINK, G. R. (1985) Ty elements transpose through an RNA intermediate. *Cell,* 40**,** 491-500.

BORCHERT, G. M., LANIER, W. & DAVIDSON, B. L. (2006) RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol,* 13**,** 1097-101.

BOVIA, F. & STRUB, K. (1996) The signal recognition particle and related small cytoplasmic ribonucleoprotein particles. *J Cell Sci,* 109 (Pt 11)**,** 2601-8.

BROUDE, N. E. & BUDOWSKY, E. I. (1971) The reaction of glyoxal with nucleic acid components. 3. Kinetics of the reaction with monomers. *Biochim Biophys Acta,* 254**,** 380-8.

BUHLER, M., VERDEL, A. & MOAZED, D. (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell,* 125**,** 873-86.

CALIN, G. A., DUMITRU, C. D., SHIMIZU, M., BICHI, R., ZUPO, S., NOCH, E., ALDLER, H., RATTAN, S., KEATING, M., RAI, K., RASSENTI, L., KIPPS, T., NEGRINI, M., BULLRICH, F. & CROCE, C. M. (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A,* 99**,** 15524-9.

CALIN, G. A., SEVIGNANI, C., DUMITRU, C. D., HYSLOP, T., NOCH, E., YENDAMURI, S., SHIMIZU, M., RATTAN, S., BULLRICH, F., NEGRINI, M. & CROCE, C. M. (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A,* 101**,** 2999-3004.

CARMELL, M. A., XUAN, Z., ZHANG, M. Q. & HANNON, G. J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev,* 16**,** 2733-42.

CHANG, L. S., LIN, S. Y., LIEU, A. S. & WU, T. L. (2002) Differential expression of human 5S snoRNA genes. *Biochem Biophys Res Commun,* 299**,** 196-200.

CHENDRIMADA, T. P., GREGORY, R. I., KUMARASWAMY, E., NORMAN, J., COOCH, N., NISHIKURA, K. & SHIEKHATTAR, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature,* 436**,** 740-4.

CHRISTY, R. J. & HUANG, R. C. (1988) Functional analysis of the long terminal repeats of intracisternal A-particle genes: sequences within the U3 region determine both the efficiency and direction of promoter activity. *Mol Cell Biol,* 8**,** 1093-102.

CLOUGH, J. E., FOSTER, J. A., BARNETT, M. & WICHMAN, H. A. (1996) Computer simulation of transposable element evolution: random template and strict master models. *J Mol Evol,* 42**,** 52-8.

COGONI, C., ROMANO, N. & MACINO, G. (1994) Suppression of gene expression by homologous transgenes. *Antonie Van Leeuwenhoek,* 65**,** 205-9.

COSTAS, J. (2003) Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L. *J Mol Evol,* 56**,** 181-6.

DAHARY, D., ELROY-STEIN, O. & SOREK, R. (2005) Naturally occurring antisense: transcriptional leakage or real overlap? *Genome Res,* 15**,** 364-8.

DEBERARDINIS, R. J., GOODIER, J. L., OSTERTAG, E. M. & KAZAZIAN, H. H., JR. (1998) Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet,* 20**,** 288-90.

DEBERARDINIS, R. J. & KAZAZIAN, H. H., JR. (1999) Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics,* 56**,** 317-23.

DECERBO, J. & CARMICHAEL, G. G. (2005) Retention and repression: fates of hyperedited RNAs in the nucleus. *Curr Opin Cell Biol,* 17**,** 302-8.

DEWANNIEUX, M., DUPRESSOIR, A., HARPER, F., PIERRON, G. & HEIDMANN, T. (2004) Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. *Nat Genet,* 36**,** 534-9.

DEWANNIEUX, M., ESNAULT, C. & HEIDMANN, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet,* 35**,** 41-8.

DEWANNIEUX, M. & HEIDMANN, T. (2005) L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J Mol Biol,* 349**,** 241-7.

DOENCH, J. G. & SHARP, P. A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev,* 18**,** 504-11.

DROUIN, G. (2000) Expressed retrotransposed 5S rRNA genes in the mouse and rat genomes. *Genome,* 43**,** 213-5.

EICKBUSH, T. H. & FURANO, A. V. (2002) Fruit flies and humans respond differently to retrotransposons. *Curr Opin Genet Dev,* 12**,** 669-74.

ELBASHIR, S. M., HARBORTH, J., LENDECKEL, W., YALCIN, A., WEBER, K. & TUSCHL, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature,* 411**,** 494-8.

ELBASHIR, S. M., LENDECKEL, W. & TUSCHL, T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev,* 15**,** 188-200.

ELBASHIR, S. M., MARTINEZ, J., PATKANIOWSKA, A., LENDECKEL, W. & TUSCHL, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J,* 20**,** 6877-88.

EMERSON, B. M. & ROEDER, R. G. (1984) DNA sequences and transcription factor interactions of active and inactive forms of mammalian 5 S RNA genes. *J Biol Chem,* 259**,** 7926-35.

ESPINOZA, C. A., ALLEN, T. A., HIEB, A. R., KUGEL, J. F. & GOODRICH, J. A. (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol,* 11**,** 822-9.

EVSIKOV, A. V., DE VRIES, W. N., PEASTON, A. E., RADFORD, E. E., FANCHER, K. S., CHEN, F. H., BLAKE, J. A., BULT, C. J., LATHAM, K. E., SOLTER, D. & KNOWLES, B. B. (2004) Systems biology of the 2-cell mouse embryo. *Cytogenet Genome Res,* 105**,** 240-50.

FAGARD, M., BOUTET, S., MOREL, J. B., BELLINI, C. & VAUCHERET, H. (2000) AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc Natl Acad Sci U S A,* 97**,** 11650-4.

FERRIGNO, O., VIROLLE, T., DJABARI, Z., ORTONNE, J. P., WHITE, R. J. & ABERDAM, D. (2001) Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nat Genet,* 28**,** 77-81.

FIRE, A., ALBERTSON, D., HARRISON, S. W. & MOERMAN, D. G. (1991) Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. *Development,* 113**,** 503-14.

FIRE, A., XU, S., MONTGOMERY, M. K., KOSTAS, S. A., DRIVER, S. E. & MELLO, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature,* 391**,** 806-11.

FLAVELL, R. B. (1994) Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc Natl Acad Sci U S A,* 91**,** 3490-6.

FREEMANTLE, S. J., KERLEY, J. S., OLSEN, S. L., GROSS, R. H. & SPINELLA, M. J. (2002) Developmentally-related candidate retinoic acid target genes regulated early during neuronal differentiation of human embryonal carcinoma. *Oncogene,* 21**,** 2880-9.

GENDREL, A. V., LIPPMAN, Z., YORDAN, C., COLOT, V. & MARTIENSSEN, R. A. (2002) Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. *Science,* 297**,** 1871-3.

GIBBONS, R., DUGAICZYK, L. J., GIRKE, T., DUISTERMARS, B., ZIELINSKI, R. & DUGAICZYK, A. (2004) Distinguishing humans from great apes with AluYb8 repeats. *J Mol Biol,* 339**,** 721-9.

GIRARD, A., SACHIDANANDAM, R., HANNON, G. J. & CARMELL, M. A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*.

GOGOLEVSKAYA, I. K., KOVAL, A. P. & KRAMEROV, D. A. (2005) Evolutionary history of 4.5SH RNA. *Mol Biol Evol,* 22**,** 1546-54.

GOODIER, J. L., OSTERTAG, E. M., DU, K. & KAZAZIAN, H. H., JR. (2001) A novel active L1 retrotransposon subfamily in the mouse. *Genome Res,* 11**,** 1677-85.

GRAFF, J. R., HERMAN, J. G., MYOHANEN, S., BAYLIN, S. B. & VERTINO, P. M. (1997) Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. *J Biol Chem,* 272**,** 22322-9.

GREWAL, S. I. & ELGIN, S. C. (2007) Transcription and RNA interference in the formation of heterochromatin. *Nature,* 447**,** 399-406.

HAGAN, C. R., SHEFFIELD, R. F. & RUDIN, C. M. (2003) Human Alu element retrotransposition induced by genotoxic stress. *Nat Genet,* 35**,** 219-20.

HALEY, B. & ZAMORE, P. D. (2004) Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol,* 11**,** 599-606.

HAMILTON, A., VOINNET, O., CHAPPELL, L. & BAULCOMBE, D. (2002) Two classes of short interfering RNA in RNA silencing. *Embo J,* 21**,** 4671-9.

HAMILTON, A. J. & BAULCOMBE, D. C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science,* 286**,** 950-2.

HAMMOND, S. M., BERNSTEIN, E., BEACH, D. & HANNON, G. J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature,* 404**,** 293-6.

HAMMOND, S. M., BOETTCHER, S., CAUDY, A. A., KOBAYASHI, R. & HANNON, G. J. (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science,* 293**,** 1146-50.

HAN, J., LEE, Y., YEOM, K. H., KIM, Y. K., JIN, H. & KIM, V. N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev,* 18**,** 3016-27.

HAN, J., LEE, Y., YEOM, K. H., NAM, J. W., HEO, I., RHEE, J. K., SOHN, S. Y., CHO, Y., ZHANG, B. T. & KIM, V. N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell,* 125**,** 887-901.

HANNON, G. J. (2002) RNA interference. *Nature,* 418**,** 244-51.

HARADA, F. & KATO, N. (1980) Nucleotide sequences of 4.5S RNAs associated with poly(A)-containing RNAs of mouse and hamster cells. *Nucleic Acids Res,* 8**,** 1273-85.

HARRISON, L. E., WANG, Q. M. & STUDZINSKI, G. P. (1999) Butyrate-induced G2/M block in Caco-2 colon cancer cells is associated with decreased p34cdc2 activity. *Proc Soc Exp Biol Med,* 222**,** 150-6.

HIGUCHI, M., MAAS, S., SINGLE, F. N., HARTNER, J., ROZOV, A., BURNASHEV, N., FELDMEYER, D., SPRENGEL, R. & SEEBURG, P. H. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature,* 406**,** 78-81.

HOAT, T. X., NAKAYASHIKI, H., TOSA, Y. & MAYAMA, S. (2006) Specific cleavage of ribosomal RNA and mRNA during victorin-induced apoptotic cell death in oat. *Plant J,* 46**,** 922-33.

HOLMBERG, L. & NYGARD, O. (2000) Release of ribosome-bound 5S rRNA upon cleavage of the phosphodiester bond between nucleotides A54 and A55 in 5S rRNA. *Biol Chem,* 381**,** 1041-6.

HORNSTEIN, E. & SHOMRON, N. (2006) Canalization of development by microRNAs. *Nat Genet,* 38 Suppl**,** S20-4.

HOUBAVIY, H. B., MURRAY, M. F. & SHARP, P. A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev Cell,* 5**,** 351-8.

HOUGE, G., ROBAYE, B., EIKHOM, T. S., GOLSTEIN, J., MELLGREN, G., GJERTSEN, B. T., LANOTTE, M. & DOSKELAND, S. O. (1995) Fine mapping of 28S rRNA sites specifically cleaved in cells undergoing apoptosis. *Mol Cell Biol,* 15**,** 2051-62.

IORIO, M. V., FERRACIN, M., LIU, C. G., VERONESE, A., SPIZZO, R., SABBIONI, S., MAGRI, E., PEDRIALI, M., FABBRI, M., CAMPIGLIO, M., MENARD, S., PALAZZO, J. P., ROSENBERG, A., MUSIANI, P., VOLINIA, S., NENCI, I., CALIN, G. A., QUERZOLI, P., NEGRINI, M. & CROCE, C. M. (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res,* 65**,** 7065-70.

JONES, L., HAMILTON, A. J., VOINNET, O., THOMAS, C. L., MAULE, A. J. & BAULCOMBE, D. C. (1999) RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell,* 11**,** 2291-301.

JONES, P. A. & BAYLIN, S. B. (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

JURKA, J., KAPITONOV, V. V., PAVLICEK, A., KLONOWSKI, P., KOHANY, O. & WALICHIEWICZ, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res,* 110**,** 462-7.

KANELLOPOULOU, C., MULJO, S. A., KUNG, A. L., GANESAN, S., DRAPKIN, R., JENUWEIN, T., LIVINGSTON, D. M. & RAJEWSKY, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev,* 19**,** 489-501.

KASS, D. H., RAYNOR, M. E. & WILLIAMS, T. M. (2000) Evolutionary history of B1 retroposons in the genus Mus. *J Mol Evol,* 51**,** 256-64.

KATAYAMA, S., TOMARU, Y., KASUKAWA, T., WAKI, K., NAKANISHI, M., NAKAMURA, M., NISHIDA, H., YAP, C. C., SUZUKI, M., KAWAI, J., SUZUKI, H., CARNINCI, P., HAYASHIZAKI, Y., WELLS, C., FRITH, M., RAVASI, T., PANG, K. C., HALLINAN, J., MATTICK, J., HUME, D. A., LIPOVICH, L., BATALOV, S., ENGSTROM, P. G., MIZUNO, Y., FAGHIHI, M. A., SANDELIN, A., CHALK, A. M., MOTTAGUI-TABAR, S., LIANG, Z., LENHARD, B. & WAHLESTEDT, C. (2005) Antisense transcription in the mammalian transcriptome. *Science,* 309**,** 1564-6.

KAWAHARA, Y. & NISHIKURA, K. (2006) Extensive adenosine-to-inosine editing detected in Alu repeats of antisense RNAs reveals scarcity of sense-antisense duplex formation. *FEBS Lett,* 580**,** 2301-5.

KAWAHARA, Y., ZINSHTEYN, B., SETHUPATHY, P., IIZASA, H., HATZIGEORGIOU, A. G. & NISHIKURA, K. (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science,* 315**,** 1137-40.

KAWASAKI, H. & TAIRA, K. (2004) Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature,* 431**,** 211-7.

KETTING, R. F., FISCHER, S. E., BERNSTEIN, E., SIJEN, T., HANNON, G. J. & PLASTERK, R. H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev,* 15**,** 2654-9.

KHVOROVA, A., REYNOLDS, A. & JAYASENA, S. D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell,* 115**,** 209-16.

KIGAMI, D., MINAMI, N., TAKAYAMA, H. & IMAI, H. (2003) MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod,* 68**,** 651-4.

KIM, D. D., KIM, T. T., WALSH, T., KOBAYASHI, Y., MATISE, T. C., BUYSKE, S. & GABRIEL, A. (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res,* 14**,** 1719-25.

KIM, D. H., VILLENEUVE, L. M., MORRIS, K. V. & ROSSI, J. J. (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol,* 13**,** 793-7.

KISS, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell,* 109**,** 145-8.

KLECKNER, N. (1977) Translocatable elements in procaryotes. *Cell,* 11**,** 11-23.

KO, M. S., KITCHEN, J. R., WANG, X., THREAT, T. A., WANG, X., HASEGAWA, A., SUN, T., GRAHOVAC, M. J., KARGUL, G. J., LIM, M. K., CUI, Y., SANO, Y., TANAKA, T., LIANG, Y., MASON, S., PAONESSA, P. D., SAULS, A. D., DEPALMA, G. E., SHARARA, R., ROWE, L. B., EPPIG, J., MORRELL, C. & DOI, H. (2000) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. *Development,* 127**,** 1737-49.

KRAFT, R., KADYK, L. & LEINWAND, L. A. (1992) Sequence organization of variant mouse 4.5 S RNA genes and pseudogenes. *Genomics,* 12**,** 555-66.

KRUTZFELDT, J., POY, M. N. & STOFFEL, M. (2006) Strategies to determine the biological function of microRNAs. *Nat Genet,* 38 Suppl**,** S14-9.

KUMAR, M. S., LU, J., MERCER, K. L., GOLUB, T. R. & JACKS, T. (2007) Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet,* 39**,** 673-7.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. (2001) Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LEAH, R., FREDERIKSEN, S., ENGBERG, J. & SORENSEN, P. D. (1990) Nucleotide sequence of a mouse 5S rRNA variant gene. *Nucleic Acids Res,* 18**,** 7441.

LEE, R. C., FEINBAUM, R. L. & AMBROS, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75**,** 843-54.

LEE, Y., AHN, C., HAN, J., CHOI, H., KIM, J., YIM, J., LEE, J., PROVOST, P., RADMARK, O., KIM, S. & KIM, V. N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature,* 425**,** 415-9.

LEE, Y., KIM, M., HAN, J., YEOM, K. H., LEE, S., BAEK, S. H. & KIM, V. N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *Embo J,* 23**,** 4051-60.

LEI, H., OH, S. P., OKANO, M., JUTTERMANN, R., GOSS, K. A., JAENISCH, R. & LI, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development,* 122**,** 3195-205.

LEINWAND, L. A., WYDRO, R. M. & NADAL-GINARD, B. (1982) Small RNA molecules related to the Alu family of repetitive DNA sequences. *Mol Cell Biol,* 2**,** 1320-30.

LEVANON, E. Y., EISENBERG, E., YELIN, R., NEMZER, S., HALLEGGER, M., SHEMESH, R., FLIGELMAN, Z. Y., SHOSHAN, A., POLLOCK, S. R., SZTYBEL, D., OLSHANSKY, M., RECHAVI, G. & JANTSCH, M. F. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol,* 22**,** 1001-5.

LI, T., SPEAROW, J., RUBIN, C. M. & SCHMID, C. W. (1999) Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene,* 239**,** 367-72.

LI, T. H., KIM, C., RUBIN, C. M. & SCHMID, C. W. (2000) K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res,* 28**,** 3031-9.

LIPARDI, C., WEI, Q. & PATERSON, B. M. (2001) RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell,* 107**,** 297-307.

LIPPMAN, Z., GENDREL, A. V., BLACK, M., VAUGHN, M. W., DEDHIA, N., MCCOMBIE, W. R., LAVINE, K., MITTAL, V., MAY, B., KASSCHAU, K. D., CARRINGTON, J. C., DOERGE, R. W., COLOT, V. & MARTIENSSEN, R. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature,* 430**,** 471-6.

LIPPMAN, Z., MAY, B., YORDAN, C., SINGER, T. & MARTIENSSEN, R. (2003) Distinct Mechanisms Determine Transposon Inheritance and Methylation via Small Interfering RNA and Histone Modification. *PLoS Biol,* 1**,** E67.

LIU, W. M., CHU, W. M., CHOUDARY, P. V. & SCHMID, C. W. (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res,* 23**,** 1758-65.

LOEB, D. D., PADGETT, R. W., HARDIES, S. C., SHEHEE, W. R., COMER, M. B., EDGELL, M. H. & HUTCHISON, C. A., 3RD (1986) The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol Cell Biol,* 6**,** 168-82.

LU, C., KULKARNI, K., SOURET, F. F., MUTHUVALLIAPPAN, R., TEJ, S. S., POETHIG, R. S., HENDERSON, I. R., JACOBSEN, S. E., WANG, W., GREEN, P. J. & MEYERS, B. C. (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res,* 16**,** 1276-88.

LU, J., GETZ, G., MISKA, E. A., ALVAREZ-SAAVEDRA, E., LAMB, J., PECK, D., SWEET-CORDERO, A., EBERT, B. L., MAK, R. H., FERRANDO, A. A., DOWNING, J. R., JACKS, T., HORVITZ, H. R. & GOLUB, T. R. (2005) MicroRNA expression profiles classify human cancers. *Nature,* 435**,** 834-8.

LUNG, B., ZEMANN, A., MADEJ, M. J., SCHUELKE, M., TECHRITZ, S., RUF, S., BOCK, R. & HUTTENHOFER, A. (2006) Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Res,* 34**,** 3842-52.

MA, J., SVOBODA, P., SCHULTZ, R. M. & STEIN, P. (2001) Regulation of zygotic gene activation in the preimplantation mouse embryo: global activation and repression of gene expression. *Biol Reprod,* 64**,** 1713-21.

MACHATTIE, L. A. & SHAPIRO, J. A. (1978) Chromosomal integration of phage lambda by means of a DNA insertion element. *Proc Natl Acad Sci U S A,* 75**,** 1490-4.

MACRAE, I. J. & DOUDNA, J. A. (2007) Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol,* 17**,** 138-45.

MAKSAKOVA, I. A. & MAGER, D. L. (2005) Transcriptional regulation of early transposon elements, an active family of mouse long terminal repeat retrotransposons. *J Virol,* 79**,** 13865-74.

MARTENS, J. H., O'SULLIVAN, R. J., BRAUNSCHWEIG, U., OPRAVIL, S., RADOLF, M., STEINLEIN, P. & JENUWEIN, T. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *Embo J,* 24**,** 800-12.

MARTIENSSEN, R. A. (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat Genet,* 35**,** 213-4.

MARTINEZ, J., PATKANIOWSKA, A., URLAUB, H., LUHRMANN, R. & TUSCHL, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell,* 110**,** 563-74.

MATRANGA, C., TOMARI, Y., SHIN, C., BARTEL, D. P. & ZAMORE, P. D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell,* 123**,** 607-20.

MAYOROV, V. I., ROGOZIN, I. B., ELISAPHENKO, E. A. & ADKISON, L. R. (2000) B2 elements present in the human genome. *Mamm Genome,* 11**,** 177-9.

MCCAFFREY, A. P., MEUSE, L., PHAM, T. T., CONKLIN, D. S., HANNON, G. J. & KAY, M. A. (2002) RNA interference in adult mice. *Nature,* 418**,** 38-9.

MCCARTHY, E. M. & MCDONALD, J. F. (2004) Long terminal repeat retrotransposons of Mus musculus. *Genome Biol,* 5**,** R14.

MCCLINTOCK, B. (1956) Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol,* 21**,** 197-216.

MEARS, M. L. & HUTCHISON, C. A., 3RD (2001) The evolution of modern lineages of mouse L1 elements. *J Mol Evol,* 52**,** 51-62.

METTE, M. F., AUFSATZ, W., VAN DER WINDEN, J., MATZKE, M. A. & MATZKE, A. J. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *Embo J,* 19**,** 5194-201.

METTE, M. F., MATZKE, A. J. & MATZKE, M. A. (2001) Resistance of RNA-mediated TGS to HC-Pro, a viral suppressor of PTGS, suggests alternative pathways for dsRNA processing. *Curr Biol,* 11**,** 1119-23.

MINENO, J., OKAMOTO, S., ANDO, T., SATO, M., CHONO, H., IZU, H., TAKAYAMA, M., ASADA, K., MIROCHNITCHENKO, O., INOUYE, M. & KATO, I. (2006) The expression profile of microRNAs in mouse embryos. *Nucleic Acids Res,* 34**,** 1765-71.

MORAN, J. V., HOLMES, S. E., NAAS, T. P., DEBERARDINIS, R. J., BOEKE, J. D. & KAZAZIAN, H. H., JR. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell,* 87**,** 917-27.

MORRIS, K. V., CHAN, S. W., JACOBSEN, S. E. & LOONEY, D. J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science,* 305**,** 1289-92.

MORSE, D. P. & BASS, B. L. (1997) Detection of inosine in messenger RNA by inosine-specific cleavage. *Biochemistry,* 36**,** 8429-34.

MURCHISON, E. P., PARTRIDGE, J. F., TAM, O. H., CHELOUFI, S. & HANNON, G. J. (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci U S A.*

NAPOLI, C., LEMIEUX, C. & JORGENSEN, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell,* 2**,** 279-289.

NIELSEN, J. N., HALLENBERG, C., FREDERIKSEN, S., SORENSEN, P. D. & LOMHOLT, B. (1993) Transcription of human 5S rRNA genes is influenced by an upstream DNA sequence. *Nucleic Acids Res,* 21**,** 3631-6.

NISHIHARA, H., SMIT, A. F. & OKADA, N. (2006) Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res,* 16**,** 864-74.

NIWA, H. (2001) Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct Funct,* 26**,** 137-48.

O'DONNELL, K. A., WENTZEL, E. A., ZELLER, K. I., DANG, C. V. & MENDELL, J. T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature,* 435**,** 839-43.

OHTA, T. (1986) Population genetics of an expanding family of mobile genetic elements. *Genetics,* 113**,** 145-59.

OKAMURA, K., ISHIZUKA, A., SIOMI, H. & SIOMI, M. C. (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev,* 18**,** 1655-66.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

ONODERA, Y., HAAG, J. R., REAM, T., NUNES, P. C., PONTES, O. & PIKAARD, C. S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell,* 120**,** 613-22.

OSTERTAG, E. M., GOODIER, J. L., ZHANG, Y. & KAZAZIAN, H. H., JR. (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet,* 73**,** 1444-51.

PAL-BHADRA, M., LEIBOVITCH, B. A., GANDHI, S. G., RAO, M., BHADRA, U., BIRCHLER, J. A. & ELGIN, S. C. (2004) Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science,* 303**,** 669-72.

PALL, G. S., CODONY-SERVAT, C., BYRNE, J., RITCHIE, L. & HAMILTON, A. (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res,* 35**,** e60.

PALLADINO, M. J., KEEGAN, L. P., O'CONNELL, M. A. & REENAN, R. A. (2000) A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. *Cell,* 102**,** 437-49.

PEASTON, A. E., EVSIKOV, A. V., GRABER, J. H., DE VRIES, W. N., HOLBROOK, A. E., SOLTER, D. & KNOWLES, B. B. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell,* 7**,** 597-606.

PONTIER, D., YAHUBYAN, G., VEGA, D., BULSKI, A., SAEZ-VASQUEZ, J., HAKIMI, M. A., LERBS-MACHE, S., COLOT, V. & LAGRANGE, T. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev,* 19**,** 2030-40.

QUENTIN, Y. (1994) A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res,* 22**,** 2222-7.

RANA, T. M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol,* 8**,** 23-36.

RIVAS, F. V., TOLIA, N. H., SONG, J. J., ARAGON, J. P., LIU, J., HANNON, G. J. & JOSHUA-TOR, L. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol,* 12**,** 340-9.

ROY-ENGEL, A. M., CARROLL, M. L., VOGEL, E., GARBER, R. K., NGUYEN, S. V., SALEM, A. H., BATZER, M. A. & DEININGER, P. L. (2001) Alu insertion polymorphisms for the study of human genomic diversity. *Genetics,* 159**,** 279-90.

SCADDEN, A. D. (2005) The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol,* 12**,** 489-96.

SCADDEN, A. D. & SMITH, C. W. (1997) A ribonuclease specific for inosine-containing RNA: a potential role in antiviral defence? *Embo J,* 16**,** 2140-9.

SCADDEN, A. D. & SMITH, C. W. (2001) RNAi is antagonized by A-->I hyper-editing. *EMBO Rep,* 2**,** 1107-11.

SCHOENIGER, L. O. & JELINEK, W. R. (1986) 4.5S RNA is encoded by hundreds of tandemly linked genes, has a short half-life, and is hydrogen bonded in vivo to poly(A)-terminated RNAs in the cytoplasm of cultured mouse cells. *Mol Cell Biol,* 6**,** 1508-19.

SCHULZ, W. A. (2006) L1 retrotransposons in human cancers. *J Biomed Biotechnol,* 2006**,** 83672.

SEMPERE, L. F., FREEMANTLE, S., PITHA-ROWE, I., MOSS, E., DMITROVSKY, E. & AMBROS, V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol,* 5**,** R13.

SEVERYNSE, D. M., HUTCHISON, C. A., 3RD & EDGELL, M. H. (1992) Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon. *Mamm Genome,* 2**,** 41-50.

SHAPIRO, J. A. (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci U S A,* 76**,** 1933-7.

SHEN, M. R., BATZER, M. A. & DEININGER, P. L. (1991) Evolution of the master Alu gene(s). *J Mol Evol,* 33**,** 311-20.

SIGUIER, P., FILEE, J. & CHANDLER, M. (2006) Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol,* 9**,** 526-31.

SIJEN, T., FLEENOR, J., SIMMER, F., THIJSSEN, K. L., PARRISH, S., TIMMONS, L., PLASTERK, R. H. & FIRE, A. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell,* 107**,** 465-76.

SIJEN, T. & PLASTERK, R. H. (2003) Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature,* 426**,** 310-4.

SMIT, A. F. (1996) The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev,* 6**,** 743-8.

SMITH, C. J., WATSON, C. F., BIRD, C. R., RAY, J., SCHUCH, W. & GRIERSON, D. (1990) Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants. *Mol Gen Genet,* 224**,** 477-81.

SMITH, H. B. (1999) Interfering with viral infection. Plants Do it too. *Plant Cell,* 11**,** 1191-4.

SOIFER, H. S., ZARAGOZA, A., PEYVAN, M., BEHLKE, M. A. & ROSSI, J. J. (2005) A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res,* 33**,** 846-56.

SORENSEN, P. D. & FREDERIKSEN, S. (1991) Characterization of human 5S rRNA genes. *Nucleic Acids Res,* 19**,** 4147-51.

SPEEK, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol,* 21**,** 1973-85.

SUZUKI, H., MORIWAKI, K. & SAKURAI, S. (1994) Sequences and evolutionary analysis of mouse 5S rDNAs. *Mol Biol Evol,* 11**,** 704-10.

SVOBODA, P., STEIN, P., ANGER, M., BERNSTEIN, E., HANNON, G. J. & SCHULTZ, R. M. (2004) RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol,* 269**,** 276-85.

SZYMANSKI, M., BARCISZEWSKA, M. Z., ERDMANN, V. A. & BARCISZEWSKI, J. (2002) 5S Ribosomal RNA Database. *Nucleic Acids Res,* 30**,** 176-8.

SZYMANSKI, M., BARCISZEWSKA, M. Z., ERDMANN, V. A. & BARCISZEWSKI, J. (2003) 5 S rRNA: structure and interactions. *Biochem J,* 371**,** 641-51.

TABARA, H., SARKISSIAN, M., KELLY, W. G., FLEENOR, J., GRISHOK, A., TIMMONS, L., FIRE, A. & MELLO, C. C. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell,* 99**,** 123-32.

TOLIA, N. H. & JOSHUA-TOR, L. (2007) Slicer and the argonautes. *Nat Chem Biol,* 3**,** 36-43.

TONKIN, L. A. & BASS, B. L. (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science,* 302**,** 1725.

TONKIN, L. A., SACCOMANNO, L., MORSE, D. P., BRODIGAN, T., KRAUSE, M. & BASS, B. L. (2002) RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans. *Embo J,* 21**,** 6025-35.

VAGIN, V. V., SIGOVA, A., LI, C., SEITZ, H., GVOZDEV, V. & ZAMORE, P. D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science,* 313**,** 320-4.

VAN DER KROL, A. R., MUR, L. A., BELD, M., MOL, J. N. & STUITJE, A. R. (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell,* 2**,** 291-9.

VASSETZKY, N. S., TEN, O. A. & KRAMEROV, D. A. (2003) B1 and related SINEs in mammalian genomes. *Gene,* 319**,** 149-60.

VERDEL, A., JIA, S., GERBER, S., SUGIYAMA, T., GYGI, S., GREWAL, S. I. & MOAZED, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science,* 303**,** 672-6.

VOLINIA, S., CALIN, G. A., LIU, C. G., AMBS, S., CIMMINO, A., PETROCCA, F., VISONE, R., IORIO, M., ROLDO, C., FERRACIN, M., PRUEITT, R. L., YANAIHARA, N., LANZA, G., SCARPA, A., VECCHIONE, A., NEGRINI, M., HARRIS, C. C. & CROCE, C. M. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A,* 103**,** 2257-61.

VOLIVA, C. F., JAHN, C. L., COMER, M. B., HUTCHISON, C. A., 3RD & EDGELL, M. H. (1983) The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. *Nucleic Acids Res,* 11**,** 8847-59.

VOLPE, T. A., KIDNER, C., HALL, I. M., TENG, G., GREWAL, S. I. & MARTIENSSEN, R. A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science,* 297**,** 1833-7.

WALSH, C. P., CHAILLET, J. R. & BESTOR, T. H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet,* 20**,** 116-7.

WANG, Q., ZHANG, Z., BLACKWELL, K. & CARMICHAEL, G. G. (2005) Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. *Curr Biol,* 15**,** 384-91.

WASSENEGGER, M., HEIMES, S., RIEDEL, L. & SANGER, H. L. (1994) RNA-directed de novo methylation of genomic sequences in plants. *Cell,* 76**,** 567-76.

WATANABE, T., TAKEDA, A., TSUKIYAMA, T., MISE, K., OKUNO, T., SASAKI, H., MINAMI, N. & IMAI, H. (2006) Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev,* 20**,** 1732-43.

WATERSTON, R. H., LINDBLAD-TOH, K., BIRNEY, E., ROGERS, J., ABRIL, J. F., AGARWAL, P., AGARWALA, R., AINSCOUGH, R., ALEXANDERSSON, M., AN, P., ANTONARAKIS, S. E., ATTWOOD, J., BAERTSCH, R., BAILEY, J., BARLOW, K., BECK, S., BERRY, E., BIRREN, B., BLOOM, T., BORK, P., BOTCHERBY, M., BRAY, N., BRENT, M. R., BROWN, D. G., BROWN, S. D., BULT, C., BURTON, J., BUTLER, J., CAMPBELL, R. D., CARNINCI, P., CAWLEY, S., CHIAROMONTE, F., CHINWALLA, A. T., CHURCH, D. M., CLAMP, M., CLEE, C., COLLINS, F. S., COOK, L. L., COPLEY, R. R., COULSON, A., COURONNE, O., CUFF, J., CURWEN, V., CUTTS, T., DALY, M., DAVID, R., DAVIES, J., DELEHAUNTY, K. D., DERI, J., DERMITZAKIS, E. T., DEWEY, C., DICKENS, N. J., DIEKHANS, M., DODGE, S., DUBCHAK, I., DUNN, D. M., EDDY, S. R., ELNITSKI, L., EMES, R. D., ESWARA, P., EYRAS, E., FELSENFELD, A., FEWELL, G. A., FLICEK, P., FOLEY, K., FRANKEL, W. N., FULTON, L. A., FULTON, R. S., FUREY, T. S., GAGE, D., GIBBS, R. A., GLUSMAN, G., GNERRE, S., GOLDMAN, N., GOODSTADT, L., GRAFHAM, D., GRAVES, T. A., GREEN, E. D., GREGORY, S., GUIGO, R., GUYER, M., HARDISON, R. C., HAUSSLER, D., HAYASHIZAKI, Y., HILLIER, L. W., HINRICHS, A., HLAVINA, W., HOLZER, T., HSU, F., HUA, A., HUBBARD, T., HUNT, A., JACKSON, I., JAFFE, D. B., JOHNSON, L. S., JONES, M., JONES, T. A., JOY, A., KAMAL, M., KARLSSON, E. K., et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature,* 420**,** 520-62.

WEINBERG, M. S., VILLENEUVE, L. M., EHSANI, A., AMARZGUIOUI, M., AAGAARD, L., CHEN, Z. X., RIGGS, A. D., ROSSI, J. J. & MORRIS, K. V. (2006) The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *Rna,* 12**,** 256-62.

WHITE, R. J., STOTT, D. & RIGBY, P. W. (1989) Regulation of RNA polymerase III transcription in response to F9 embryonal carcinoma stem cell differentiation. *Cell,* 59**,** 1081-92.

WOODCOCK, D. M., LAWLER, C. B., LINSENMEYER, M. E., DOHERTY, J. P. & WARREN, W. D. (1997) Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J Biol Chem,* 272**,** 7810-6.

XIE, Z., JOHANSEN, L. K., GUSTAFSON, A. M., KASSCHAU, K. D., LELLIS, A. D., ZILBERMAN, D., JACOBSEN, S. E. & CARRINGTON, J. C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol,* 2**,** E104.

YANG, D., LU, H. & ERICKSON, J. W. (2000) Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in Drosophila embryos. *Curr Biol,* 10**,** 1191-200.

YANG, N. & KAZAZIAN, H. H., JR. (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol,* 9, 758-9.

YANG, S., TUTTON, S., PIERCE, E. & YOON, K. (2001) Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. *Mol Cell Biol,* 21**,** 7807-16.

YATES, P. A., BURMAN, R. W., MUMMANENI, P., KRUSSEL, S. & TURKER, M. S. (1999) Tandem B1 elements located in a mouse methylation center provide a target for de novo DNA methylation. *J Biol Chem,* 274**,** 36357-61.

YE, K., MALININA, L. & PATEL, D. J. (2003) Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature,* 426**,** 874-8.

YODER, J. A., WALSH, C. P. & BESTOR, T. H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet,* 13**,** 335-40.

ZAMORE, P. D., TUSCHL, T., SHARP, P. A. & BARTEL, D. P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell,* 101**,** 25-33.

ZHANG, H., KOLB, F. A., JASKIEWICZ, L., WESTHOF, E. & FILIPOWICZ, W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell,* 118**,** 57-68.

ZILBERMAN, D., CAO, X. & JACOBSEN, S. E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science,* 299**,** 716-9.