Regulation of microRNAs and analysis of the splicing inhibitor Spliceostatin A

Priti Rupa Sarkar

Imperial College London, Division of Surgery and Oncology, Institute of Reproduction and Developmental Biology

Thesis submitted to Imperial College London for the degree of Doctor of Philosophy

Statement of originality and copyright declaration

I performed all experiments included in this thesis unless otherwise stated.

The copyright of this thesis rests with the author and is made available under a Creative Commons. Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

I'd like to thank my supervisor Dr Nick Dibb for all the help and support he has given me throughout my PhD. Without his encouragement and patience; this project would not have been possible.

Many thanks to Dr Elcie Chan for her work on the generation of stem cell miRNA libraries which helped formed the basis of my project. I am also very grateful to Chin for his advice on the lentiviral experiments. I'd like to thank Alice for all her help with the confocal imaging and also Dr Navaratnam for all the reagents he has kindly loaned to me over the years. Thanks to all IRDB members, especially the 5th floor, past and present, for all their help and advice and for making the last few years some of the best.

Special thanks to my family for their continuous encouragement and best wishes. Finally I would like to thank Tom for all his support and patience, especially during the last few months, and also for making the polyadenylation mining tool for me.

Publications

Sarkar R, Tan GC, Winston RW, Dibb NJ. Prevalence of post-transcriptional regulation of microRNAs in embryonic stem cells. In preparation.

Sarkar R, Tan GC, Chandrashekran A, Tatusova T, Winston RW, Vorechovsky I, Kasputin Y, Dibb NJ. Noisy splicing and genetic disease. Revisions in preparation.

Kapustin Y, Chan E, **Sarkar R,** Wong F, Vorechovsky I, Winston RM, Tatusova T, Dibb NJ. Cryptic splice sites and split genes. Nucleic Acids Res. 2011 Aug;39(14):5837-44. Epub 2011 Apr 5.

Tan GC, Chan E, Molnar A, Ellis P, Robinson S, Isa IM, Chauhan R, **Sarkar R**, Guillot P, Castellano L, Langford C, Cui W, Winston RM, Meister G, Baulcombe D, Dibb NJ. IsomiRs have specific cell and tissue expression patterns and can target different mRNAs. In preparation.

ABSTRACT

The advent of deep sequencing has revealed marked differences in microRNA (miRNA) expression profiles between cell types. We focus on how miRNAs are regulated in embryonic stem cells (ESCs) and their differentiated neural progenitor stem cells (NPSCs). We discovered that 60% of miRNAs are regulated by transcription. The remaining 40% of miRNAs were transcribed but not processed in ESCs. Furthermore, systematic analysis of miRNAs on chromosome 1 and 2 helped to validate our results. We show that 43% of these miRNAs, which are highly expressed in differentiated cell types but poorly expressed in ESCs, were transcribed but not processed in ESCs. Therefore we infer that post-transcriptional regulation of miRNAs is prevalent in ESCs and could potentially be a characteristic of this cell type.

Post-transcriptional regulation of miRNAs is also evident in miRNA clusters. Clusters of miRNAs are transcribed together however individual mature miRNA expression often differs between tissues and even within the same tissue type. Following investigation into post-transcriptional regulation, we discovered extensive alternative polyadenylation in three different miRNA clusters. We suggest that alternative polyadenylation has the potential to influence regulation of individual miRNAs within a cluster contributing to their differential expression levels.

We have also established an improved assay for the novel anti-tumour splicing inhibitor, Spliceostatin A (SSA). Prior to our work, there was a disparity between the effects of SSA *in vitro* versus *in vivo.* In agreement with this, our initial results showed that only <25% of genes were affected by SSA, as assayed by RT-PCR. However, we demonstrate that the splicing of nascent RNAs was more globally inhibited by SSA, indicating that the previous assay had been obscured by the large amounts of spliced mRNA already present in cells. The new assay allowed us to uncover the effects of SSA on nuclear localization of RNA and to verify these effects with another inhibitor, TG003.

TABLE OF CONTENTS

FIGURE AND TABLE LIST 11
ABBREVIATIONS 11

CHAPTER 1 - INTRODUCTION 15

1.1 RNA 16
1.1.1 Non-coding RNA 17
1.1.2 Small non-coding RNA 18
1.2 MICRORNA 20
1.2.1 The discovery of microRNAs 20
1.2.2 Stages of miRNA biogenesis 21
1.2.3 MicroRNA regulation of messenger RNA 25
1.2.4 MicroRNA evolution and conservation 29
1.2.5 Transcriptional regulation of miRNAs 32
1.2.6 Post-transcriptional regulation of miRNAs 33
1.2.7 MicroRNAs in stem cells and disease 42
1.3 SPLICING AND THE SPLICEOSOME 49
1.3.1 RNA splicing and the spliceosome 49
1.3.2 Splice site selection and alternative splicing 52
1.3.3 Roles of splicing in disease 56
1.3.4 Splicing inhibitors 59
1.3.5 Splicing and microRNAs 61
1.4 PROJECT AIMS 63
1.4.1 MicroRNA regulation in ESCs 63
1.4.2 Alternative polyadenylation of miRNA genes 63
1.4.3 Establishing a working assay for novel splicing inhibitor, Spliceostatin A 63

CHAPTER 2 - MATERIALS AND METHODS 64

2.1 CELL CULTURE 65
2.1.1 Cell culture of HEK293, HT1080 and HeLa cells 65
2.1.2 Human embryonic stem cells and neuroprogenitor stem cells 65
2.1.3 Freezing cell lines 66
2.2 GENERAL RNA AND DNA PROTOCOLS 67
2.2.1 Total RNA extraction 67
2.2.2 Small RNA enrichment 67
2.2.3 cDNA synthesis 68
2.2.4 Polymerase Chain Reaction (PCR) 68
2.3 DNA EXTRACTION AND SEQUENCING 69

2.3.1 DNA gel elution....69
2.3.2 Vector cloning....69
2.3.3 Colony PCR....70
2.3.4 Sequencing....71
2.4 RACE....72
2.4.1 5'rapid amplification of cDNA ends (5'RACE)....72
2.4.2 3'rapid amplification of cDNA ends (3'RACE)....73
2.5 NORTHERN HYBRIDISATION....75
2.5.1 Labelling of DNA probes ^{32}P γATP....75
2.5.2 Northern blot gel....75
2.5.3 Hybridisation....76
2.5.4 Northern blots with double-DIG labelled LNA probes....76
2.5.5 Hybridisation with LNA probes....77
2.5.6 Washing and detection....77
2.5.7 Northern blotting of primary miRNAs by agarose-formaldehyde gel....78
2.6 ISOLATING NASCENT RNA TRANSCRIPTS....80
2.6.1 Thiouridine labelling....80
2.6.2 Total RNA extraction....81
2.6.3 Biotinylation of 4sU labelled RNA transcripts....81
2.6.4 Biotin affinity purification of 4sU labelled RNA (nascent RNA)....83
2.7 IMMUNOCYTOCHEMISTRY OF NASCENT RNAS....83
2.8 ISOLATION OF NUCLEAR AND CYTOPLASMIC RNAS....84
2.9 LENTIVIRUS PRODUCTION AND INFECTION....85
2.10 REAGENTS AND PRIMERS....87
2.10.1 PCR, Northern blotting and Cloning reagents....87
2.10.2 Other commonly used reagents....87
2.10.3 Conditions for commonly used chemicals and drugs....88
2.10.4 MicroRNA PCR primers....88
2.10.5 3'RACE primers....90
2.10.6 5'RACE primers....91
2.10.7 Northern blotting probes....91
2.10.8 Primers used for checking lentivirus transfection miR-let 7a and miR-let 7f....91
2.10.9 Primers used in Chapter 5....92
2.11 BIOINFORMATICS PROGRAMS....93

CHAPTER 3 - EXTENSIVE POST-TRANSCRIPTIONAL REGULATION OF MICRORNAS IN ESCS....94

3.1 INTRODUCTION....95
3.1.1 MicroRNA signature in human stem cells....95
3.1.2 Post-transcriptional regulation of microRNAs....97
3.2 RESULTS....100
3.2.1 RT-PCR methods cannot control for detection of confounding transcripts that overlap miRNAs of interest....100

3.2.2 Problems with RT-PCR analysis of miRNA expression are confirmed when investigating precursor miRNAs 101
3.2.3 3'RACE successfully overcomes detection of confounding overlapping transcripts allowing the identification of primary miRNAs regulated by transcription and post-transcriptionally 103
3.2.4 Many miRNAs in ESCs are regulated post-transcriptionally 107
3.2.5 Post-transcriptionally regulated miRNA are actively transcribed and turned over 108
3.2.6 Over-active expression of primary miRNAs in ESCs suggest a prevalence of post-transcriptional regulation of miRNAs 111
3.3 DISCUSSION 114
3.3.1 Transcriptional regulation of miRNAs 114
3.3.2 Post-transcriptional regulation of microRNAs is prevalent in ESCs 115
3.3.3 The role of post-transcriptional regulation of miRNAs 119

CHAPTER 4 - EXTENSIVE ALTERNATIVE POLYADENYLATION OF MICRORNAS 120

4.1. INTRODUCTION 121
4.1.1 Alternative polyadenylation 121
4.1.2 Alternative polyadenylation and stem cells 124
4.1.3 Regulation of microRNA biogenesis by alternative polyadenylation 126
4.2 RESULTS 128
4.2.1 Annotation of the miR-302 cluster with evidence of extensive alternative polyadenylation 128
4.2.2 Annotation of the transcription start site of the microRNA 302 cluster 132
4.2.3 Alternative polyadenylation of a singular miR-99a 132
4.2.4 MicroRNA let 7a/f cluster forms different primary transcripts in ESCs and NPSCs. 135
4.2.5 Each alternatively polyadenylated let 7a/f primary transcript found in ESCs and NPSCs can be processed into precursor miRNAs in HepG2s. 138
4.3 DISCUSSION 141
4.3.1 MicroRNA gene clusters 141
4.3.2 Alternative polyadenylation in miRNA clusters 142

CHAPTER 5 - DEVELOPING A WORKING ASSAY FOR THE NOVEL ANTI-TUMOUR DRUG SPLICEOSTATIN A 148

5.1 INTRODUCTION 149
5.1.1 Spliceostatin A and TG003 149
5.1.2 The importance of splicing inhibitors in disease 152
5.1.3 Aims 153

5.2 RESULTS....154
5.2.1 The effect of Spliceostatin A is obscured by existing stable RNAs....154
5.2.2 Effects of Spliceostatin A on nascent RNAs....161
5.2.3 Inhibition of the spliceosome causes retention of RNA in the nucleus....166
5.3 DISCUSSION....171
5.3.1 The effects of splicing inhibitors, SSA and TG003 can be seen more clearly in nascent RNAs....172
5.3.2 The effects of SSA and TG003 differ....173
5.3.3 Splicing inhibition causes nuclear retention of RNA transcripts....175

CHAPTER 6 - GENERAL DISCUSSION....176

6.1 POST-TRANSCRIPTIONAL REGULATION OF MICRORNAS IN ESCS....177
6.2 ALTERNATIVE POLYADENYLATION OF MIRNA CLUSTERS....178
6.3 EFFECT OF SPLICEOSTATIN A ON NASCENT RNAS....179
6.4 SPLICING AND MICRORNAS....181
6.5 CONCLUSION....183

REFERENCES....185
SUPPLEMENTARY MATERIAL....202

FIGURE AND TABLE LIST

Figure 1.1	MicroRNA biogenesis	**25**
Figure 1.2	Illustration of microRNA seed sequence	**27**
Figure 1.3	Schematic of Lin28-mediated control of let-7 biogenesis	**37**
Figure 1.4	Phase-contrast images of ESCs and neural differentiation	**44**
Figure 1.5	Northern blot and sequencing data	**44**
Figure 1.6	Schematic of pre-mRNA splicing by the spliceosome	**51**
Figure 1.7	Splice site recognition	**53**
Figure 2.1	Overview of the 5'RACE protocol	**73**
Figure 2.2	Overview of the 3'RACE protocol	**74**
Figure 2.3	Capillary transfer arrangement	**79**
Figure 2.4	pTRIPz inducible lentiviral vector used in the cloning of the three miR-let-7a/miR-let-7f	**86**
Figure 3.1.1	Northern blot and Sequencing data	**97**
Figure 3.1.2	Many stages of miRNA biogenesis is open to post-transcriptional regulation	**99**
Figure 3.2.1	RT-PCR of miRNA is confounded by overlapping transcripts	**101**
Figure 3.2.2	RT-PCR analysis of precursor microRNAs	**103**
Figure 3.2.3	Transcriptional regulation of miRNAs specific to ES and NPS cells	**106**
Figure 3.2.4	Post-transcriptional regulation of miRNAs in ESCs	**107**
Figure 3.2.5	RNA turnover of primary miRNA indicates that post-transcriptionally regulated miRNA are consistently being processed	**109**
Figure 3.2.6	Detection of miRNAs in nascent RNA samples from ESCs and NPSCs indicates that post-transcriptionally regulated miRNAs are continuously being processed	**111**
Figure 3.2.7	Over-expression of primary miRNAs in ESCs implies that the processing of many miRNAs is blocked	**113**
Figure 4.1.1	Polyadenylation of pre-mRNA	**122**
Figure 4.1.2	Types of alternative polyadenylation	**124**
Figure 4.2.1	The 302 microRNA cluster has many 3' ends, potentially allowing individual miRNAs differential expression within the cluster	**131**
Figure 4.2.2	Identification of miR-302 transcription start site corroborating existing data	**132**
Figure 4.2.3	miR-99a shows extensive alternative polyadenylation	**134**
Figure 4.2.4	Primary miRNA let-7a and let-7f are differentially polyadenylated in ESCs and NPSCs	**137**

Figure 4.2.5	Lentivirus infection of HepG2 cells with primary let-7a and let-7f of different sequence lengths indicates that alternative polyadenylation alone cannot inhibit precursor miRNA production	**140**
Figure 4.3.1	Maps of the 5'end of primary let-7a-1/let-7f-1 cluster	**145**
Figure 5.1.1	Natural splicing inhibitors and their synthetic derivatives	**149**
Figure 5.2.1	SSA shows little visible effect on RNA by RT-PCR analysis	**155**
Figure 5.2.2	Spliceostatin A has a visible effect on genes induced by IFNα	**158**
Figure 5.2.3	Detection of splicing inhibition for MxA, OAS1 and 6-16 genes	**159**
Figure 5.2.4	Alternative splicing inhibitor TG003 has a similar effect on RNA splicing as SSA	**160**
Figure 5.2.5	Detection and integrity analysis of biotinylated and biotin affinity purified RNA	**163**
Figure 5.2.6	Spliceostatin A has a visible effect on newly synthesised RNAs	**165**
Figure 5.2.7	Spliceostatin A and TG003 splicing inhibitors cause retention of RNAs in the nucleus	**167**
Figure 5.2.8	Quantification of immunocytochemistry analysis of HepG2 cells	**168**
Figure 5.2.9	Titration of Spliceostatin A and TG003 concentration	**169**
Figure 5.2.10	Spliceostatin A causes a build-up of nascent pre-mRNAs in the nucleus	**171**
Supplementary figure S1	Alternative polyadenylation of miR-21	**205**
Supplementary figure S2	Human and mouse miR-302 cluster sequence alignment	**208**
Supplementary figure S3	PCR of miR let-7a primary transcript, digestion of this insert post ligation into pTRIPz and sequencing of the insert and vector	**209**
Supplementary figure S4	PCR of miR let-7f primary transcript, digestion of this insert post ligation into pTRIPz and sequencing of the insert and vector	**209**
Supplementary figure S5	PCR of miR let-7af primary transcript, digestion of this insert post ligation into pTRIPz and sequencing of the insert and vector	**210**
Table 3.1.1	MicroRNA expression data taken from ESC and NPSC sequencing libraries	**96**
Table 4.1.1	MicroRNAs within clusters	**121**
Supplementary table S1	ESC and NPSC miRNAs and their host mRNAs	**202**
Supplementary table S2	List of verified poly(A) signal motifs	**204**
Supplementary table S3	Genes upregulated by IFNα	**204**

ABBREVIATIONS

3' RACE assay 3'rapid amplification of cDNA ends assay
5' RACE assay 5'rapid amplification of cDNA ends assay
ADAR adenosine deaminases acting on RNA
Ago Argonaute protein
AMD Actinomycin D
APA Alternative polyadenylation
b bases
bp base pair
cDNA complementary deoxyribonucleic acid
RNAIP RNA immunoprecipitation
DAPI 4,6-diamidino-2-phenylindole
DGCR8 DiGeorge syndrome critical region 8
DIG-labelled LNA probe digoxigenin-labelled locked nucleic acid probe
DMF Dimethylformamide
DMSO Dimethylsulfoxide
DNA deoxyribonucleic acid
dNTP deoxynucleotide triphosphate
dsRNA double stranded RNA
EDTA Ethylenediaminetetraacetic acid
eIF4E eukaryotic translation initiation factor subunit 4E
endo-siRNA endogenous small interfering RNA
ESCs/ESs human embryonic stem cells
EST expressed sequence tag
HEK293 human embryonic kidney 293 cells
HEPG2 Hepatocellular carcinoma
HT1080 Fibrosarcoma cell line
IFNα interferon alpha
iPSCs induced pluripotent stem cells
kb kilobase
lnc RNA long non-coding RNA
miRNA/miR- microRNA
miRNP microribonucleoprotein complex
mRNA Messenger RNA
ncRNA non-coding RNA
NMD non-sense mediated mRNA decay
NPSCs/NSCs/NSs human neuronal progenitor stem cells
nt nucleotide
PACT protein activator of protein kinase R (PKR)
PAS Polyadenylation signal
PAGE polyacrylamide gel electrophoresis
PBS phosphate buffer saline
PBS-BT phosphate buffer saline supplemented with 0.1% BSA and 0.2% Tween20

piRNA PIWI interacting RNA
PLL polylysine
pre-miRNA precursor miRNA
pri-miRNA primary miRNA
RIPA radioimmunoprecipitate assay
RISC RNA-induced silencing complex
RLC RISC loading complex
RNA ribonucleic acid
RNAbp RNA binding protein
RT reverse transcription
RT-PCR Real-time polymerase chain reaction
SDS sodium dodecyl sulfate
siRNA small interfering RNA
SNPs Single Nucleotide Polymorphisms
SSA spliceostatin
SSC saline sodium citrate
ssRNA single stranded RNA
TAE Tri-acetate EDTA
TBE Tris-borate/EDTA
TBS-T Tris-Buffered Saline and Tween 20
TEMED Tetramethylethylenediamine
TG003 Cdc2-like kinase inhibitor
TRBP trans-activator RNA (Tar) binding protein
TSS transcription start site
UTR untranslated region

CHAPTER 1 - INTRODUCTION

1.1 RNA

Ribonucleic acid (RNA) is a nucleic acid polymer that plays a number of crucial roles in biological processes across all species. The central dogma of molecular biology states that DNA encodes genetic information that is transcribed into messenger RNA (mRNA) by RNA polymerase enzymes. Messenger RNAs are subsequently translated into proteins by ribosomes (Crick, 1970). This promotes the idea that the primary role of RNA is to facilitate the translation of protein. However, due to its diverse sequence and biochemical properties, RNAs are powerful molecules, with a vital role in cellular processes that extends beyond its ability to code for proteins.

RNAs have the capability to form secondary and tertiary structures allowing molecular binding interactions and catalytic reactions (Hiriart et al., 2012). Since the discovery of RNA in 1868 (Dahm, 2005) our understanding of RNA function has been constantly expanding. RNAs have been characterised according to function including catalytic and non-coding RNAs which have roles within gene regulation, viral transposon defence and epigenetics (Kaikkonen et al., 2011).

Owing to advances in technologies, an abundance of post-transcriptional RNA regulation has been discovered including alternative splicing, alternative polyadenylation and RNA editing (Du et al., 2009). In addition biochemical advancements have allowed analysis of protein-RNA interactions *in vivo* yielding further insights into RNA biogenesis and regulation. The human genome sequencing project and several recent high-throughput transcriptomic analyses

have revealed that eukaryotes transcribe 90% of their genomes but less than 10% of genes that encode RNA transcripts get translated into proteins (Martens-Uzunova et al., 2013, Kaikkonen et al., 2011). In addition, the coding exons of these genes account for only 1.5% of the genome (Esteller, 2011).

The functionality of the majority of transcribed non-coding RNAs is still under vigorous debate (Ponting and Belgard, 2010), however we are aware that a subset of RNAs are not "junk" nor are they simply the intermediary between DNA and protein but actually functional end products themselves (Baulcombe, 2003, Fire, 2007, Licatalosi and Darnell, 2010).

1.1.1 Non-coding RNA

Although protein-coding genes represent a small percentage of the human genome, they remain the most studied, and only recently has the importance of some non-coding types of RNAs become apparent.

Previously the function of ncRNA was limited to "infrastructural" roles in protein synthesis with ribosomal RNA (rRNA), essential for the structure of ribosomes; transfer RNA (tRNA), essential as adaptors between mRNA and amino acids; small nuclear RNAs (snRNA), involved in splicing and small nucleolar RNA (snoRNA); involved in the chemical modifications of these three precursor ncRNAs (Lodish and Porter, 1980, Alberts, 2002, Lodish, 2000).

Recently some species of the non-coding RNA fraction have been shown to be expressed dynamically during development and regulated by external stimuli (Ounzain et al., 2013). This suggested more of a regulatory role for ncRNAs and with further technological advancements, their functional capacity in a number biological process became obvious (Ounzain et al., 2013).

There are several types of regulatory ncRNA which can be further categorised according to size, function and tissue specificity. Long non-coding RNAs (lncRNAs) are an abundant class of RNAs, ranging in sizes longer than ~200 nucleotides. Long ncRNAs share structural characteristics with mRNA however they are typically less conserved and expressed less frequently. Current publications show lncRNAs to be key modulators of cell biology and some lnRNAs are misregulated in several human diseases (Harismendy et al., 2011, Hinterseher et al., 2011).

1.1.2 Small non-coding RNA

MicroRNAs (miRNAs), piwi RNAs (piRNAs) and endogenous small interfering RNAs (siRNAs) differ in origin, structure and cell specificity, however all play fundamental roles in biology and disease in multiple species (Liu and Paroo, 2010). What unifies these three small RNAs is their size of about 20-30nts, their association with Argonaute (Ago) family proteins and their role in RNAi mediated gene regulation (Kim et al., 2009). Germ cell-specific piRNAs are the longest in length (24-31nt), with roles in transposon silencing through heterochromatin formation or RNA destabilisation (Kim et al., 2009). Small

interfering RNAs are shorter in length (21nts) and derived from long double-stranded RNA. Small interfering RNA and miRNA share a primary function as post-transcriptional regulators of target RNAs.

MicroRNAs are small non-coding RNAs that regulate diverse biological functions including cell proliferation, apoptosis, organ development, signal transduction and differentiation (Esquela-Kerscher and Slack, 2006, Kloosterman and Plasterk, 2006). MicroRNAs control biological processes via post-transcriptional regulation, which facilitates fine tuning of gene expression. With over 2578 mature miRNA sequences identified in the human genome (miRBASE, http://www.mirbase.org/), miRNAs are the most highly sequenced of all small RNA types forming the largest fraction; currently over 60% of total small RNAs are miRNAs (Kuchenbauer et al., 2008). Although a lot is still unknown about miRNAs, they remain the best understood of the three small RNA classes. MicroRNAs are predicted to be encoded by only 1% of the mammalian genome therefore it is striking that miRNAs regulate nearly a third of all protein coding genes (Li et al., 2009). The global nature of miRNA-directed regulation on thousands of genes makes it unsurprising that they play an important role in various human diseases (Shah et al., 2009, Hudder and Novak, 2008, Barbato et al., 2009). The mechanisms and biogenesis of miRNAs must be better understood to acquire knowledge of gene networks and more effective strategies for genetic manipulation (Kim et al., 2009).

1.2 MICRORNA

1.2.1 The discovery of microRNAs

Victor Ambros, Gary Ruvkun and colleagues discovered the first microRNA, *Lin-4,* in *Caenorhabditis elegans* (*C.elegans*) twenty years ago (Wightman et al., 1993, Ambros, 2004, Lagos-Quintana et al., 2001). The *lin-4* gene was discovered to control timing of larval development of *C.elegans.* It does not code for protein but instead gets processed into two small RNAs (Lee et al., 1993). Lee et al., (1993) observed that one of the RNAs was ~22 nt in length and the other was ~60 nt; it was later concluded that the longer ~60 nt RNA was a precursor transcript that would fold into a stem loop in order to be processed into the smaller ~22 nt RNA. Furthermore it was observed that *lin-4* RNAs had antisense complementarity to sites in the 3'UTR of the *lin-14* mRNA (Lee et al., 1993, Wightman et al., 1993). Ruvkun and colleagues established the importance of the 3'UTR complementarity for the regulation of *lin-14* by *lin-4* showing a decrease in *Lin-14* protein without change in the mRNA levels (Wightman et al., 1993). This research provided one of the RNA interference models known today in which small RNA *lin-4* pairs with the 3'UTR of *lin-14* to cause translational repression (Lee et al., 1993, Wightman et al., 1993).

Over the past two decades since the discovery of small RNAs, this field of molecular biology has significantly advanced. Since the first miRNA, *lin-4,* was discovered there have been a further 21,264 miRNA annotated in over 100 species (miRBASE http://www.mirbase.org/). The importance of miRNAs in literally every biological process is becoming known, with roles in regulating

developmental timing and signalling pathways, cell death, metabolism and brain development (Esquela-Kerscher and Slack, 2006, Kloosterman and Plasterk, 2006). In addition miRNAs play a part in human pathologies like viral diseases, genetic disorders and cancer (Xue et al., 2009, Castellano et al., 2009).

Increasing numbers of mRNA targets of miRNAs are being identified and consequently miRNA functions are being experimentally determined helping to elucidate miRNA-induced gene regulation (Clancy et al., 2011, Kahai et al., 2009). MicroRNA-directed regulation is now known to occur at some of the most crucial levels of genome function including chromatin structure, chromosome segregation, transcription, RNA processing, RNA stability and translation (Kahai et al., 2009). MicroRNAs generally have an inhibitory role in gene expression however there have been recent reports of miRNAs causing upregulation of gene expression by direct RNA-binding (Kahai et al., 2009).

1.2.2 Stages of miRNA biogenesis

Approximately half of human miRNA genes are located within the intron of protein coding genes; another 40% are located within the intergenic regions of non-coding DNA and the remainder is suggested to be located within the exons of protein-coding or non-coding genes (Kim et al., 2009, Berezikov, 2011, Lu et al., 2008). It is estimated that over half of all human miRNAs are located in clusters of tandemly repeated pre-miRNAs, which are transcribed as single polycistronic units sharing a promoter. Despite regulation by the same transcriptional factors, each miRNA within a cluster can exhibit different cell-

type and developmental specific expression patterns (Sotillo and Thomas-Tikhonenko, 2011). This indicates that miRNAs within clusters can be processed individually, though the mechanism driving this remains elusive (Sotillo and Thomas-Tikhonenko, 2011).

As illustrated in figure 1.1 miRNA biogenesis begins in the nucleus where the majority of primary miRNA (pri-miRNA) transcripts are generated by RNA polymerase II (Du and Zamore, 2005). A few miRNAs that associate with Alu repeats can alternatively be transcribed by RNA polymerase III (Kim et al., 2009). Primary miRNAs are several kilobases long with a 5' cap structure and a 3' poly(A) tail and many secondary stem-loop structures (Kim et al., 2009). Stem-loop hairpin precursor miRNAs (pre-miRNA) are processed from the primary miRNA by an RNase III enzyme, Drosha and an RNA binding protein cofactor, DiGeorge syndrome critical region (DGCR8, also known as "Pasha") (Lee et al., 2003, Han et al., 2004, Gregory et al., 2004).

Together Drosha and DGCR8 form a large ~650 kDa microprocessor complex which is conserved in animals and essential for embryonic development (Kim et al., 2009). It has been reported that pri-miRNA processing by the microprocessor may be co-transcriptional, whereby Drosha processing occurs after the transcript associates with the splicing machinery, but before the miRNA-containing intron is removed (Kim and Kim, 2007, Morlando et al., 2008, Kim et al., 2009). Some miRNAs located within the introns of protein-coding genes can bypass Drosha processing and undergo spliceosome processing of short introns into pre-miRNA (mirtrons) (Figure 1.1) (Sotillo and Thomas-Tikhonenko, 2011).

Some mirtrons require exonucleolytic trimming at the 5' and 3' ends to be exported (Sotillo and Thomas-Tikhonenko, 2011, Ladewig et al., 2012).

All pre-miRNAs share a stem-loop structure of about 70 nts with a 2 nucleotide overhang at the 3' end (Sotillo and Thomas-Tikhonenko, 2011). Regardless of processing method, Exportin 5 recognises the >14 bp ds-RNA and the short 3'overhang of the miRNA, so that together with cofactor Ran-GTP they mediate the nuclear export of pre-miRNA into the cytoplasm (Kim et al., 2009). In the cytoplasm, pre-miRNAs are cleaved again near the terminal loop by another RNAse III endoribonuclease, Dicer, producing short (18 nt-25 nt) double-stranded RNA duplexes (Hutvagner et al., 2001, Sotillo and Thomas-Tikhonenko, 2011).

Dicer is highly conserved and in humans Dicer interacts closely with TRBP (TAR RNA-binding protein) and PACT (protein kinase RNA activator), which do not contribute to miRNA processing, but instead aid the formation of the RNA-induced silencing complex (RISC) (Figure 1.1) (Kim et al., 2009, Sotillo and Thomas-Tikhonenko, 2011). The RNA duplex unwinds into separate strands where one strand becomes the 18-25 nt mature "guide" miRNA and the other is the passenger miRNA "*" strand. It is predicted that the choice of mature miRNA is dependent on the themodynamic stability of base pairing of the 5' end of the duplex where the least stable strand unwinds enabling association with Ago proteins (Sotillo and Thomas-Tikhonenko, 2011, Khvorova et al., 2003, Kim et al., 2009). It is also possible that both strands can be processed as functional mature miRNAs in comparable amounts. Current literature is vague about the

mechanism behind strand selection however at present it is thought that Dicer frees the duplex after cleavage and the stable end of the miRNA duplex binds to TRBP and the other end associates with one of the Ago proteins (Preall and Sontheimer, 2005).

In *D. melanogaster* and humans it has been shown that multi-functioning Ago 2 is also able to remove the passenger miRNA strand by endonucleolytic cleavage (Matranga et al., 2005, Miyoshi et al., 2005). In humans, any of the four Ago proteins can be recruited by RISC with no obvious preference, however only Ago 2 can cleave target mRNAs resulting in post-transcriptional gene silencing (Sotillo and Thomas-Tikhonenko, 2011, Meister et al., 2004). In *C.elegans* and *D.melanogaster* the choice of Ago protein is potentially determined by the miRNA duplex structure, whereby perfect matches and central mismatches preferentially associate with different Ago proteins (Kim et al., 2009). In humans there seems to be little difference in the type of associated miRNA duplexes suggesting that Ago 1-4 may share similar roles (Kim et al., 2009).

It has recently come to light that not all miRNAs remain in the cytoplasm, miR-29b, miR-320 and miR-373 accumulate in the nucleus and target mRNA for silencing (Hwang et al., 2007, Kim et al., 2008, Place et al., 2008, Liao et al., 2010). This implies that miRNAs are able to translocate back into the nucleus after processing in the cytoplasm and shows that nuclear mRNAs do not escape regulation by miRNAs.

During miRNA processing, variations arise due to alternative RNAse III cleavage producing isomiRs which are defined as miRNAs with differing 5' and 3' ends (Lee et al., 2010). During miRNA biogenesis there are several opportunities for regulating miRNA expression. A level of control of miRNA expression is essential in maintaining normal cell function, as misregulation of mRNAs can result in fatal disease (Kim et al., 2009). Consequently studies into the regulation of miRNAs could prove to be hugely important in development and disease.

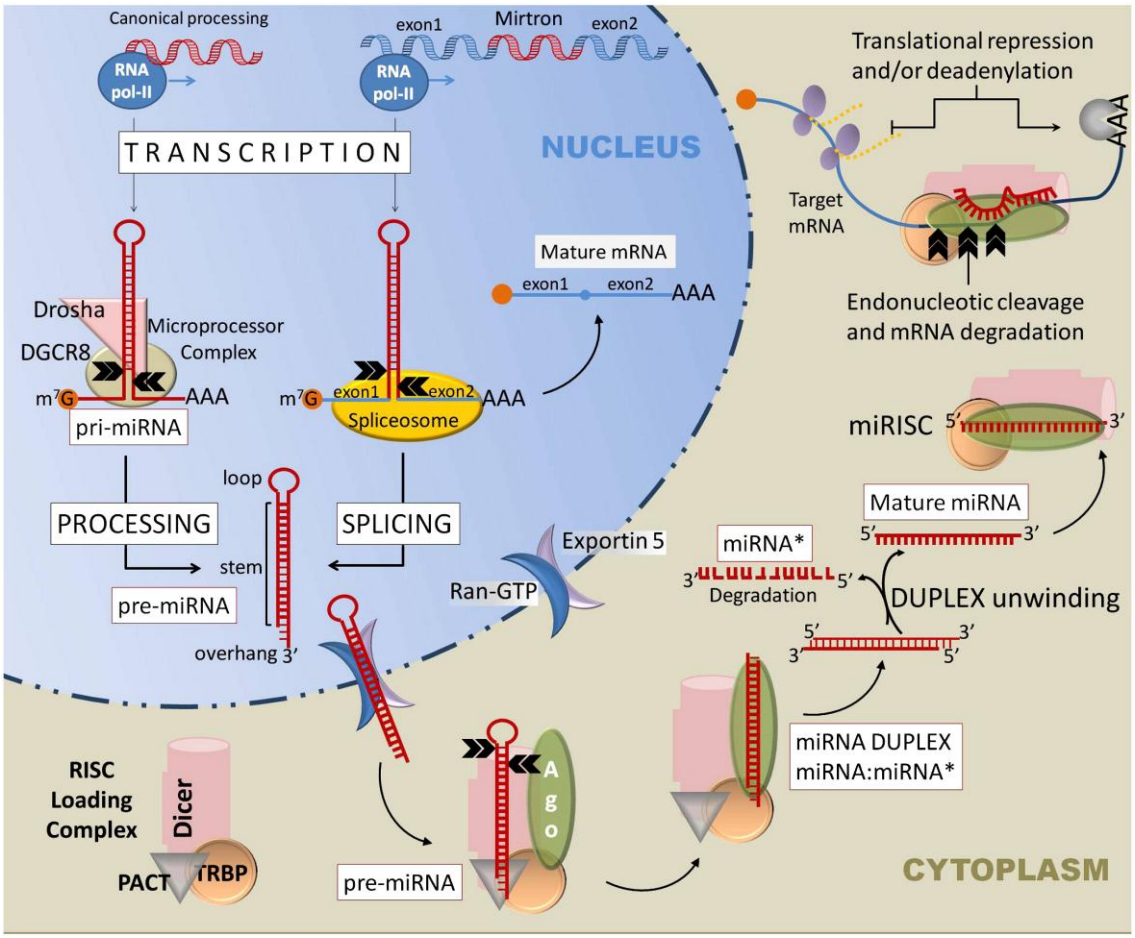


Figure 1.1 MicroRNA biogenesis. Schematic diagram of canonical miRNA biogenesis and mirtron processing. Double black arrows illustrate cleavage sites. (adapted from Sotillo and Thomas-Tikhonenko, 2011).

1.2.3 MicroRNA regulation of messenger RNA

The primary role of microRNAs is to regulate target mRNA, a process that is dictated by Watson-Crick complementary base pairing of the mature miRNA to the mRNA 3'UTR. In animal species miRNA targeting does not require a high degree of base-pairing to target mRNA. Therefore miRNAs that are not exactly complementary to their mRNA targets can act to inhibit protein synthesis while retaining the stability of the mRNA target (Guo et al., 2010, Ambros, 2004). However, the extent of complementarity can dictate the precise effect that the miRNA may exert on the target mRNA.

It is possible that miRNAs can also regulate target genes by direct cleavage of a single phosphodiester bond on the target mRNA between bases 10 and 11, this requires 100% complementarity between the miRNA and the target site of the mRNA (Elbashir et al., 2001). The most crucial feature for miRNA-mRNA targeting is the presence of a seed region which is normally the 2-8 nucleotides of the 5' end of mature miRNA, which matches with the 3'UTR sequence of the mRNA perfectly (figure 1.2). Although there are bulges and mismatches in the miRNA-mRNA pairing, perfect base pairing is required in the region that associates with Ago (seed region).

On the whole it is thought that perfect base pairing between the miRNA seed and the mRNA target can cause destabilisation by three different mechanisms; endonucleolytic target cleavage by Ago2, translational inhibition and mRNA decay (Sotillo and Thomas-Tikhonenko, 2011).

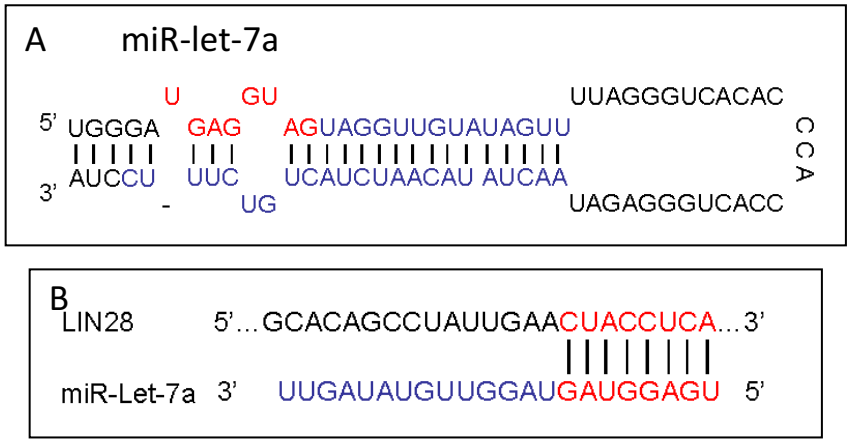


Figure 1.2 Illustration of microRNA seed sequence. A) Sequence and structure of miR-let-7a, highlighted in blue is the sequence of the mature miRNAs (guide strand on top and passenger below) and the red sequence is the 5' seed sequence of the guide miRNA. B) The 8 nt seed sequence (in red) of miR-let-7a binds the 3'UTR of Lin28 with perfect complementarity.

The RISC can inhibit the initiation of target mRNA translation by preventing cap-recognition by eIF4E (Kiriakidou et al., 2007). Messenger RNAs whose translation is cap-independent are reported to be immune to miRNA-mediated silencing (Humphreys et al., 2005). This RISC-mediated de-capping is a prerequisite to Poly(A) binding protein (PABP)-dependent deadenylation of the target mRNA by the deadenylation complex (Giraldez et al., 2006, Wu et al., 2006). Following deadenylation, the transcript is degraded by cytoplasmic 5'-3' exonuclease XRN1 (Wakiyama et al., 2007).

Translational repression can be induced by upregulating ribosome inhibitory proteins e.g. eIF6, which inhibits the joining of the 60S ribosomal units with the 40S and can also result in mRNA deadenylation (Chendrimada et al., 2007, Selbach et al., 2008). Furthermore, it is has been suggested that miRNAs can inhibit translational elongation by causing co-translational degradation (Nottrott et al., 2006) and/or cause premature ribosome dissociation from the mRNA

(Petersen et al., 2006). The mechanism behind miRNA-mediated translational repression is not yet fully clarified and further studies are required to confirm current hypotheses.

Global mRNA analyses have led to the discovery of widespread alternative polyadenylation (APA) that has functional significance to the regulation of mRNA by miRNA. An example of APA regulating miRNA-mRNA interactions occurs in muscle stem cells where Pax3, a myogenic regulator, is able to evade miR-206 regulation in specific cells by producing transcripts with shorter 3'UTRs (Boutet et al., 2012). In proliferating cells such as ESCs, mRNAs tend to have shorter 3'ends making them less susceptible to miRNA regulation conversely differentiated cells generate mRNA with long 3'UTRs resulting in higher levels of gene regulation. This phenomenon is particularly prevalent in cancer cells which have high expression of short 3'UTR mRNAs that are able to evade miRNA regulation (Clancy et al., 2011, Mayr and Bartel, 2009).

Bioinformatics and experimental data suggest that a single miRNA may have over a hundred mRNA targets and as a result could have a substantial effect on gene expression. It is characteristic for mRNA transcripts to be regulated by multiple miRNAs and multiple mRNA transcripts can also be targeted by a single miRNA (Barbato et al., 2009). Consequently expression profiling of miRNAs is a growing trend aiding the identification of miRNAs that regulate a range of processes in development, potentially showing promise as miRNA biomarkers of disease.

1.2.4 MicroRNA evolution and conservation

MicroRNA genes are evolutionarily conserved across a vast number of species and they reside in various locations, either singly or as single transcriptional units that encode several miRNAs (Lagos-Quintana et al., 2001, Lau et al., 2001). The discovery of conserved microRNA let-7 in worms, flies, and humans led to the realization that miRNAs are widespread in animals and evolutionarily conserved.

MicroRNAs in animals are possibly distinct from plant miRNAs because although there are some similarities, there is limited homology in sequence and structure (Zhang et al., 2006, Chapman and Carrington, 2007, Kim et al., 2009). Our knowledge about miRNA origin, evolution and conservation is increasing with the ever expanding databases of next-generation high throughput sequencing of small RNAs. The mirBASE data collection includes miRNAs from over 180 species including 368 miRNAs in *C.elegans*, 2042 in human and 338 in *Arabidopsis thaliana* making miRNAs the largest class of regulatory genes in animals. MicroRNAs are known to be present amongst a variety of animal species including miR-100 which is shared between eumetazoans indicating that miRNAs were present very early on in animal evolution (Berezikov, 2011). Conservation observed in 34 miRNAs between two bilaterian lineages, protostomes and deuterostomes, indicate that miRNA expansion occured at the bilaterian lineages (Berezikov, 2011). Bilaterian animal miRNAs are phylogenetically conserved and more than 55% of *C.elegans* miRNAs have human homologues reinforcing the importance of miRNAs throughout evolution (Kim et al., 2009, Pasquinelli et al., 2000). There are several theories about the

origin of miRNAs and it is known that novel miRNAs are constantly arising without evidence of miRNA loss (Berezikov, 2011). Consequently this may help explain the emergence of a direct correlation between the number of miRNAs and morphological complexity of a species (Berezikov, 2011).

There are several distinct origins of miRNAs however the major source of novel miRNAs is by gene duplication (Kim et al., 2009). Over one third of human miRNAs are found in "families" meaning they share similar mature miRNA sequences and identical (or virtually identical) seed sequences. MicroRNA gene families can arise by two types of duplication, local/tandem duplication where the duplicate gene remains adjacent to the original miRNA or non-local duplication where the new miRNA is located away from the original. Evidence of both tandem and non-local duplication can be seen in the human let-7 family which can be found in 12 different genomic loci (Kim et al., 2009).

More than half of all human miRNAs are located in the intron of protein-coding genes (Rodriguez et al., 2004) and transcribed either sense or antisense to the overlapping gene. Intronic miRNAs tend to have arisen later in evolution, being more species specific and less conserved (Berezikov, 2011). The hypothesis behind this observation is that those introns which are readily transcribed, and have an affinity towards forming secondary folding structures could potentially be recognised and processed by the miRNA machinery (Berezikov, 2011).

Other sources of miRNA genes can be derived from retrotransposons or DNA transposons as well as snoRNAs, tRNAs and pseudogenes (Yuan et al., 2010).

With the constant emergence of novel miRNAs it is likely that many novel miRNA targets will be determined by chance and often result in deleterious effects. Chen and Rajewsky, 2007 suggest that novel miRNAs are subject to transcriptional control. They theorise that miRNAs are initially expressed at low levels in restricted tissues to reduce random targeting allowing the transcriptome to gradually eliminate harmful consequences by natural selection (Chen and Rajewsky, 2007). In addition to transcriptional regulation miRNA genes are also under post-transcriptional regulation which can result in diversification that is important to the evolution of established miRNAs (Chen and Rajewsky, 2007, Berezikov, 2011).

Many mature miRNA transcripts can arise from the same hairpin as a result of differential processing by Drosha and Dicer enzymes and non-templated 3' nucleotide addition (Lee et al., 2010). These mature miRNAs can vary at the 3' or 5' ends producing isomiRs (Lee et al., 2010). 3' isomiRs are common to most miRNAs and although they share the same targets their production can be tissue specific and biologically relevant. 5' isomiRs can result in a change in the seed sequence resulting in different target predictions (Berezikov, 2011). The variation in expressions of dominant isomiRs during evolution is known as "seed shifting" and is often evident in distant species (Berezikov, 2011). After processing by Dicer the miRNA duplex unwinds resulting in the incorporation of only one miRNA strand into RISC. This is thought to be dependent on relative thermodynamic stability at the ends of the strands and potentially other accessory components. The star or passenger miRNA was thought to be degraded however although it is generally less abundant than the guide strand

they can still associate with the RISC complex, accumulate in cells to biologically relevant levels and regulate target mRNA. Therefore it is possible that the ratio of guide to star strand miRNA can be altered. This is known as “arm switching and frequently occurs in a variety of species and between tissues making this an important mechanism for miRNA selection (Berezikov, 2011).

1.2.5 Transcriptional regulation of miRNAs

MicroRNA expression is heavily regulated by transcription and like other RNA Pol II substrates, transcription factors are involved in the spatial-temporal regulation of miRNA gene transcription (Kim et al., 2009). So far 243 transcription factor-miRNA regulatory relationships have been identified bioinformatically and/or experimentally (Wang et al., 2010). Examples include the transcription factor and tumour suppressor p53, which is known to activate the miR-34 family (He et al., 2007). Furthermore during myogenesis, transcription factor myogenin and MYOD1 bind miR-1 and miR-133 loci and induce gene transcription (Chen et al., 2006, Rao et al., 2006).

MicroRNA genes are also exposed to epigenetic regulation, which is often altered in cancer cell lines. This can be seen in the miR-203 locus, where DNA methylation of this gene is common in T-cell lymphoma but not in normal T lymphocytes (Bueno et al., 2008). A study by Suzuki et al. 2011 combined miRNA expression data, chromatin signature and histone modifications in colorectal cancer cells to identify 47 miRNAs in 37 gene transcription units that are under epigenetic control (Suzuki et al., 2011).

1.2.6 Post-transcriptional regulation of miRNAs

Differential expression of miRNAs can be seen across different cell and tissue types as well as different developmental stages and there is an abundance of evidence that suggests transcriptional regulation alone does not account for the regulation of all miRNAs (Ryazansky et al., 2011, Siomi and Siomi, 2010, Davis and Hata, 2009, Thomson et al., 2006, Newman and Hammond, 2010, Obernosterer et al., 2006). Developmental timing and external stimuli can dictate miRNA expression resulting in a number of outcomes. This includes global down-regulation of miRNA expression in tumour cell lines; differential expression of individual miRNAs within clusters; expression of single miRNA strands and differential miRNA isomer expression (van Kouwenhove et al., 2011, Bernstein et al., 2003, Murchison et al., 2005). Post-transcriptional regulation is an important mechanism for the regulation of messenger RNA, as evident by the expanding repertoire of interfering RNAs. As a result it is not surprising that it would also play a large part in the regulation of miRNA themselves.

The complexity of miRNA biogenesis lends itself to post-transcriptional regulation at multiple stages, leading to either elevated or reduced miRNA expression. MicroRNA regulation may be governed by proteins involved in miRNA processing, nuclear export and variations in the miRNA transcript itself as a result of SNPs or processing irregularities (Wu et al., 2008).

Microprocessor levels can differ between cell types and in diseases such as prostate and ovarian cancer, where Drosha and Dicer expression is known to vary (Shiohama et al., 2003, Ambs et al., 2008, Merritt et al., 2008). This implies

that the miRNA processing machinery itself is under post-transcriptional control. The microprocessor is strictly regulated through a combination of negative and positive feedback loops. For example the microprocessor is able to cleave DGCR8 mRNA hairpins reducing DGCR8 expression, and DGCR8 protein is able to stabilise Drosha protein providing regulatory feedback loops (Han et al., 2009).

Drosha forms a complex with DEAD-box RNA helicases P68 and P72 and nuclear factors NF90 and NF45 (Gregory et al., 2004). These components have been proven to affect processing efficiency of specific miRNAs (Davis et al., 2008, Fukuda et al., 2007, Slezak-Prochazka et al., 2010, Suzuki et al., 2009, Yamagata et al., 2009). Endogenous p68 regulates miRNA expression under different stimuli by interaction with various proteins including SMAD transcription factors to facilitate the processing of pri-miRNA 21 and p53 to enhance expression of several pri-miRNAs in response to DNA damage (Slezak-Prochazka et al., 2010). It is known that p53 plays a large role in the regulation of miRNAs promoting recruitment of the microprocessor to the target pri-miRNAs (Suzuki et al., 2009). Regulation by p68/p72 can also cause down-regulation of a group of pri-miRNAs upon estrogen receptor alpha (ERα) stimulation (Yamagata et al., 2009). NF90/NF45 have also been shown to regulate miRNA biogenesis, as over expression of these proteins result in the accumulation of pri-let-7a-1, pre-miR-21 and pri-miR-15a~16-1, however the mature miRNA levels remain constant, suggesting a compensatory feedback mechanism is in place to restore mature miRNA expression (Sakamoto et al., 2009).

The low levels of pre-miRNA expression compared to mature and primary miRNA expression implies that Dicer efficiently processes pre-miRNA (Lee et al., 2008). Dicer associates with TRBP, PACT and Ago to facilitate RISC assembly. TRBP phosphorylation mediated by the mitogen-activated protein kinase (MAPK) signalling pathway results in stabilisation of RISC. This results in elevated expression of growth promoting miRNAs (e.g. miR-17 and miR-92) and down-regulation of growth-inhibiting miRNAs (e.g. miR-let-7a) (Paroo et al., 2009). Overexpression of Ago proteins is also known to upregulate the expression of specific mature miRNA (Diederichs and Haber, 2007).

Primary and precursor miRNAs may have multiple conformations caused by base pair mismatches of the stem structure and internal loops; the choice of conformation determines successful Drosha-DGCR8 processing (Ma and Huang 2010). The terminal loop of miRNA precursors contains key sites for several RNA binding proteins which stabilises a specific structure either aiding or diminishing Drosha-DGCR8 processing (Ma and Huang 2010).

During differentiation of mouse and human ESCs, mature miR-let-7 expression is enhanced whilst the primary transcript remains constant. The expression of the inhibitory RNA binding protein, Lin28 was discovered by biochemical purification of associated complexes and mass spectrometric analysis. In humans, two isoforms of Lin28 are generated, cytoplasmic Lin28 (or Lin28a) and Lin28b which is mainly localised to the nucleolus. Lin28 alone is able to inhibit the processing of both the primary and precursor miRNAs and Lin28b is able to bind the pri-miRNA in the nucleolus causing repression of mature miR-let-7

expression in ESCs and various cancer cell types (see figure 1.3). Both Lin28 and Lin28b bind to the terminal loop of precursor and/or primary miR-let-7 and block processing of these transcripts. Lin28 substrates are targeted by terminal uridyl transferase (TUTase) Zcchc11/TUT4 and subjected to 3'uridylation of the precursor and primary transcripts resulting in its degradation because Dicer cannot cleave the precursor miRNA with such an elongated 3' tail (Thornton and Gregory, 2012, Heo et al., 2008). Lin28b substrates are unable to be processed by the microprocessor complex by a mechanism that still remains unknown (Thornton and Gregory, 2012). To add to the complexity of miR-let-7 regulation, Lin28 itself is a target of miR-let-7, culminating in a complex double feedback loop (Thornton and Gregory, 2012).

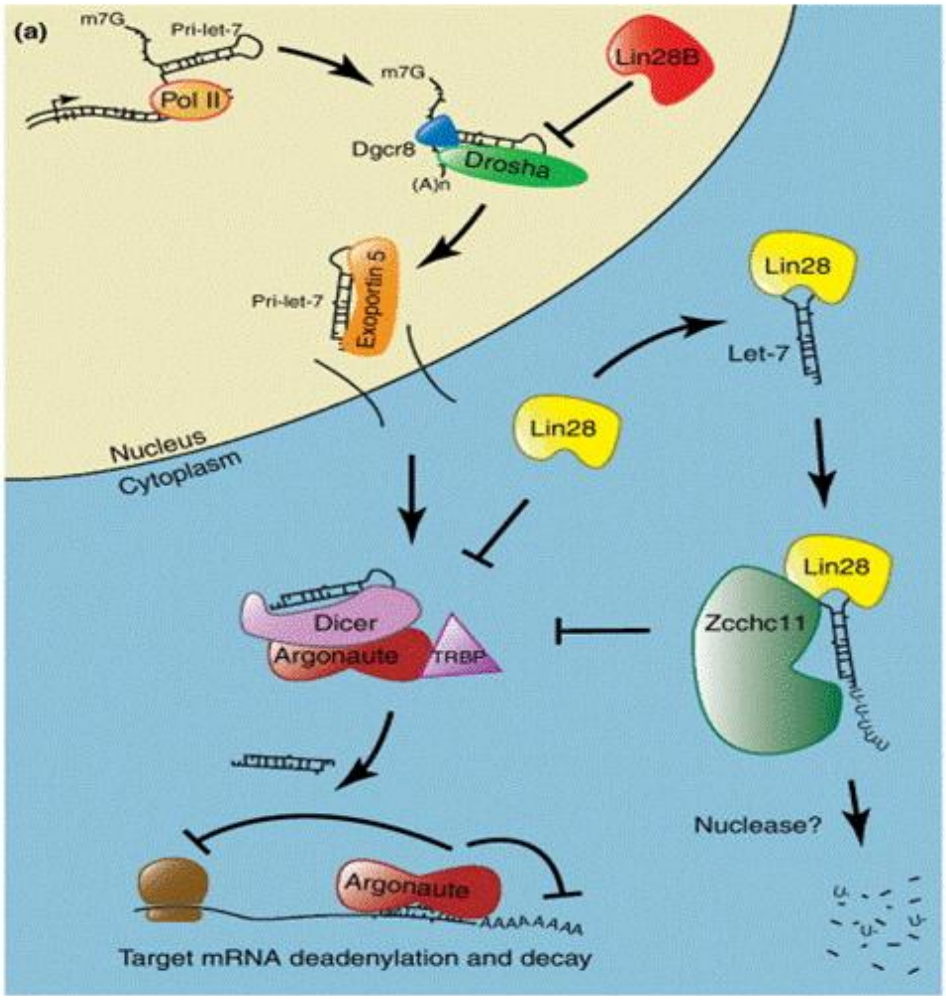


Figure 1.3 Schematic of Lin28-mediated control of let-7 biogenesis. Left: When Lin28 is absent miRNA-let-7 is processed through the canonical miRNA biogenesis pathway. Right: Lin28 blocks miR-let-7 processing by Dicer in the cytoplasm through direct binding with the terminal loop of pre-miR-let-7 and recruits the terminal uridyl transferase (TUTase) Zcchc11/TUT4 to catalyze the addition of a 3′ oligouridine tail, marking pre-let-7 for degradation. Lin28b is expressed primarily in the nucleolus and binds to pri-let-7, blocking the activity of the microprocessor complex through an unknown mechanism. Reproduced from (Thornton and Gregory, 2012)

Lin28 is the first of a rising number of RNA binding proteins which have been discovered to associate with the terminal loop of miRNAs and regulate their expression. This reveals a potent method for regulation of specific miRNAs. Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) is an RNA binding protein that assists Drosha-DGCR8 in the processing of miR-18a (Guil and Caceres, 2007). hnRNP A1 binds the terminal loop and the stem region resulting in the relaxation of the primary miRNA which facilitates processing by the microprocessor (Michlewski et al., 2010, Michlewski et al., 2008). Pri-miR-18a is located within the miR-17~19 cluster however hnRNP A1 does not affect the other members of the cluster which might explain the differences in expression of the individual miRNAs. In addition hnRNP A1 also binds to pri-let-7a-1 and pri-miR-101-1 suggesting that hnRNP A1 may regulate the expression of several miRNAs (Slezak-Prochazka et al., 2010).

KH-type splicing regulatory protein (KSRP) is a component of the microprocessor complex and RISC. KSPR binds to the terminal loop of primary or precursor transcripts and induces the expression of a cohort of miRNAs including miR-21, miR-let-7a and miR-16 (Trabucchi et al., 2009).

The expression of Arsenic Resistance protein 2 (Ars2) is linked to the proliferation of cells during mammalian embryonic development and Ars2 associates with Drosha and DGCR8 to regulate miRNA processing of miR-let-7 and miR-21 (Gruber et al., 2009). The mechanism behind Ars2 regulation is not fully known and its conservation is still to be confirmed in humans.

It is probable that there are many RNA binding proteins involved in the regulation of miRNAs by association with the terminal loop. Approximately 14% of terminal loop sequences across vertebrates are conserved (Michlewski et al., 2008). Michlewski et al. (2008) demonstrate that pri-miRNAs with conserved terminal loops are down-regulated by overexpression of antisense loop oligos conversely non-conserved pri-miRNA were unaffected by the antisense loop oligos (Michlewski et al., 2008). This suggests that there is a genuine role of conserved terminal loops in primary miRNA processing.

Exportin 5 interacts exclusively with pre-miRNAs recognising the >14 b double stranded region and the 2 nt 3' overhang. Knockdown of exportin 5 results in a decrease in both precursor and mature miRNA expression, indicating that exportin 5 has a role in stabilising the precursor transcript (Yi et al., 2003). Exportin 5 is a rate limiting factor for miRNA production as knockdown and over-expression coincides with reduction or elevation of miRNA expression respectively (Yi et al., 2005, Lund et al., 2004). It has been shown that export of pre-miR-31 is tissue specific and *in situ* hybridisation for pre-miR-31 showed cytoplasmic localisation in HS766T cells however in MCF7 cells pre-miR-31 accumulated in the nucleolus (Lee et al., 2008).

Changes in miRNA genes or miRNA transcripts can also result in altered miRNA processing. Single nucleotide polymorphisms (SNPs) and point mutations in the pri-miRNA sequence could alter the secondary structure inhibiting processing of the miRNA. Mutations in 5 out of 42 specific miRNA genes were found in 15% of patients with chronic lymphocytic leukemia (Calin et al., 2005). Moreover a

mutation in the miR-15a~16-1 gene, 7 bp downstream from pre-miR16-1 causes reduced mature miRNA expression (Calin et al., 2005). MicroRNA 146 and miR-125a are known to have reduced expression due to SNPs located on one of the strands in the double stranded stem of the pre-miRNA (Jazdzewski et al., 2008, Duan et al., 2007). This introduces a bulge in the pri-miRNA hairpin resulting in disruption of Drosha processing (Jazdzewski et al., 2008, Duan et al., 2007). These studies highlight the importance of secondary structure in the biogenesis of miRNAs however, it is important to note that changes to miRNA folding by mutations, are incredibly rare (Jazdzewski et al., 2008, Duan et al., 2007).

RNA editing plays a widespread role in miRNA sequence alteration and has a large effect on miRNA processing. RNA editing is carried out by adenosine deaminases acting on RNAs (ADARs) (Valente and Nishikura, 2005). ADAR editing is a well-known mechanism for post-transcriptional miRNA regulation in mammals where primary miRNAs are subject to adenosine – inosine conversion by adenosine deaminases. Inosines are similar to guanosine in structure and are recognised as such, therefore the sequence and the structure of the RNA is altered. (Valente and Nishikura, 2005) The sites subject to ADAR editing are often located on the double stranded stem of the pri-miRNA resulting in the suppression of processing of the pri-miRNA by Drosha and DGCR8 (Yang et al., 2006, Valente and Nishikura, 2005).

Primary miR-22 was the first miRNA discovered to undergo ADAR editing at sequence positions near the Drosha cleavage site however, the role of miR-22 editing is still unknown (Luciano et al., 2004). ADAR editing of pri-miR-142

resulted in reduced Drosha processing of the transcript and cleavage *in vitro* by Tudor-SN, which specifically degrades inosine-containing dsRNAs (Scadden, 2005, Yang et al., 2006). ADAR editing can occur after primary to precursor miRNA processing, causing inhibition of Dicer processing. Edited pre-miR-151 is bound by the Dicer complex, but cleavage and release of mature miRNA is prevented (Kawahara et al., 2007). Although ADARs are mainly localised to the nucleus some isoforms can shuttle between the cytoplasm and the nucleus. (Kawahara et al., 2007). ADARs are able to interfere with cytoplasmic pre-miRNA and edit the "seed" sequence resulting in inhibition of completely different mRNA targets as proven for miR-376 (Kawahara et al., 2007). Although ADAR editing is widespread amongst miRNA transcripts and its regulatory effects on miRNAs seem apparent, the biological significance of ADAR editing for a majority of miRNAs is not known.

It has recently been reported that 3' adenylation and uridylation of precursor and mature miRNAs, result in post-transcriptional enhancement or deregulation of the miRNA. MicroRNA-122 is stabilised by adenylation of its mature transcript by GLD-2, while uridylation by TUTase4 destabilises miR-let-7 at the precursor stage (Katoh et al 2009; Heo et al 2009). It is possible that adenylation and uridylation mechanisms compete for activation or degradation of the same miRNA transcript (Yu and Chen, 2010). Sequencing results have identified that 3' modifications are common amongst precursor and mature miRNAs implying that this may be a general mechanism for regulation of miRNAs (Newman et al., 2011).

Recently a large post-transcriptional regulatory network of conserved pseudogenes has been discovered, containing 3'UTRs with conserved miRNA binding sites (Sumazin et al., 2011, Treiber et al., 2012). These pseudogenes are actively transcribed and function as "sponges" to sequester mature miRNAs, preventing binding of genuine targets (e.g. PTEN and pseudogene PTENP1) (Sumazin et al., 2011, Treiber et al., 2012). In addition a smaller number of genes were discovered that affect miRNA-mRNA interactions through indirect sequestration of specific miRNAs, known as "non-sponge" regulators (Sumazin et al., 2011, Treiber et al., 2012).

The increasing number of publications on post-transcriptional regulation of miRNAs over the past decade emphasises the significance of understanding miRNA biogenesis. Despite this, there have been few reports of miRNAs regulated post-transcriptionally in human ESCs and ESC differentiation, the most well-known example being Lin28 regulation of the let-7 miRNA family.

1.2.7 MicroRNAs in stem cells and disease

Human embryonic stem cells (ESCs) are isolated from the inner cell mass of a blastocyst. The hallmark characteristic of ESCs, is their ability to remain indefinitely in an undifferentiated state of self-renewal, whilst simultaneously having the capacity to differentiate into all cell types (pluripotency) (Friel et al., 2005). Consequently, ESCs have the exciting potential for the treatment of disease through regenerative medicine and can also provide valuable model systems for drug development. ESCs' unique characteristics are under spatial-

temporal control regulated by cell signalling, epigenetic, transcriptional and post-transcriptional factors (Gangaraju and Lin, 2009). The switch from pluripotency to differentiation of stem cells requires reduced cell proliferation and transcriptional activation of cell lineage genes (Berardi et al., 2012).

In order to understand ESCs' great therapeutic potential for disease, we must expand our understanding of the mechanisms behind self-renewal and pluripotency. Consequently much focus has been placed on the generation of induced pluripotent stem cells (iPSCs) (Thorrez and Sampaolesi, 2011). A novel approach by Takahashi and Yamanaka to generate these cells involved the use of four defined transcription factors, Oct4, Sox2, Nanog, and Lin28 to generate iPSCs in mice and human fibroblast cells (Takahashi and Yamanaka, 2006, Takahashi et al., 2007). These pluripotency transcription factors and others e.g. Klf4, c-Myc and Esrrb have been discovered to be regulated by miRNAs at a post-transcriptional level (Takahashi and Yamanaka, 2006, Takahashi et al., 2007, Feng et al., 2009).

Neural differentiation of human ESCs *in vitro* occurs in three stages where more than 90% of the cell population becomes neural progenitor stem cells (NPSCs) after 3 weeks of differentiation (Gerrard et al., 2005). This was achieved by growing ESCs in neurobasal media supplemented with N-2, B-27 and mouse recombinant noggin which ensured differentiation into NPSCs by blocking the bone morphogenetic protein pathway (Gerrard et al., 2005). Cells at each of the 4 stages of neural differentiation have specific characteristics including differences in appearance (see Figure 1.4) and cell surface antigen expression. In

collaboration with Dr Wei Cui, our laboratory has made miRNA libraries from ESCs before and after differentiation into stage N3 NPSCs. Solexa sequencing of these libraries in Dr Nick Dibb's group by Dr Elcie Chan has identified a number of miRNAs that are preferentially expressed in ESCs or NSPCs (figure 1.5, and see the rest in supplementary table S1).

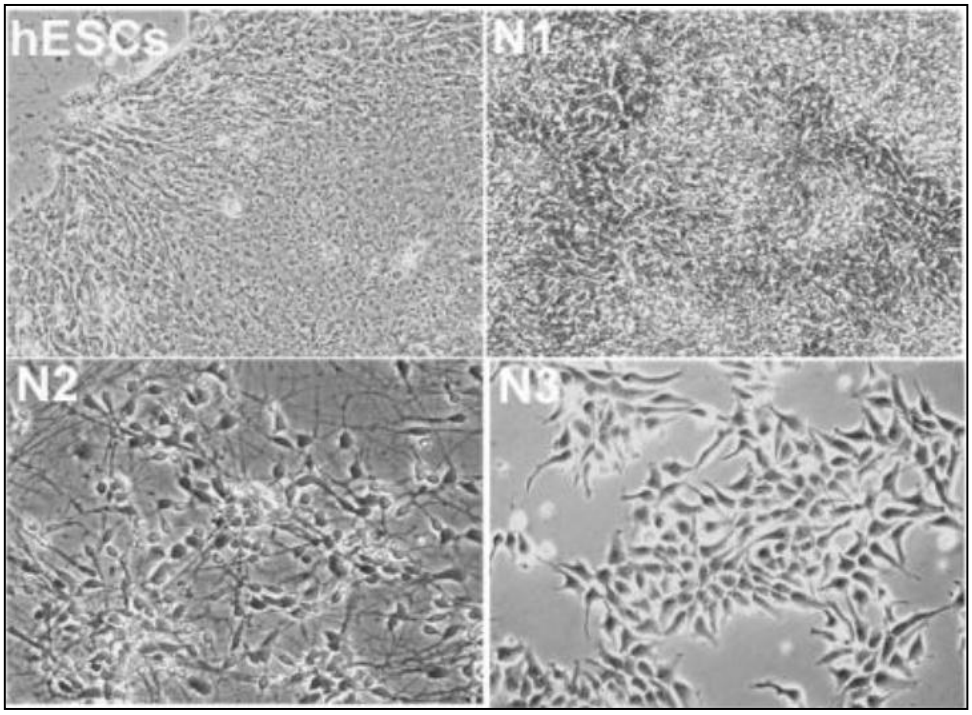


Figure 1.4 Phase-contrast images of ESCs and neural differentiation stages N1, N2 and N3 (courtesy of Dr Wei Cui (Gerrard et al., 2005).

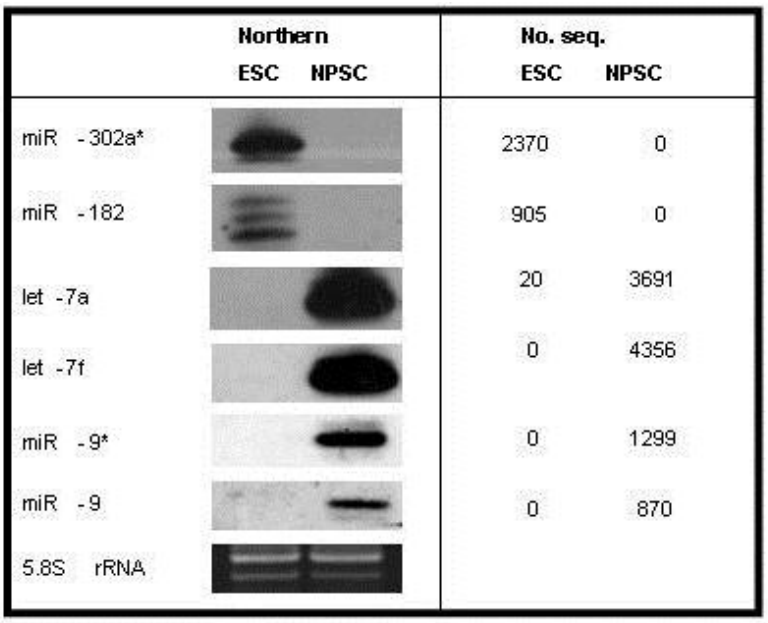


Figure 1.5 Northern blot and Sequencing data validating expression of miRNA specific to ESCs and NPSCs (Chan et al 2009 unpublished)

Dr Elcie Chan's sequencing data corroborates recent publications showing a distinct miRNA expression profile in stem cells, suggesting a miRNA-specific influence in maintaining pluripotency and influencing differentiation (Lakshmipathy et al., 2007, Laurent et al., 2008, Wilson et al., 2009). The preferential expression pattern of specific miRNA families in ESCs and the reduction of others is largely conserved in mice (Leonardo et al., 2012).

Since miRNAs are known to regulate over 50% of mammalian protein coding genes it is not surprising to learn of miRNA involvement in maintaining pluripotency and influencing differentiation (Chekulaeva and Filipowicz, 2009, Gangaraju and Lin, 2009). During development and disease miRNAs regulate cells by influencing self-renewal, differentiation, survival and apoptosis in addition to controlling cell response to external stimuli (Vidigal and Ventura, 2012).

Knockout of the microprocessor machinery (Drosha and/or DGCR8) in mice results in embryonic lethality proving miRNAs to be essential for development (Bernstein et al., 2003, Wang et al., 2007). Recently reports show that *in vitro* expression of ESC-specific miRNAs can greatly enhance the efficiency of somatic reprogramming to an ESC-like state as well as sustain pluripotency in ESCs (Li and He, 2012, Liao et al., 2011, Kuo and Ying, 2013). This verifies the role of highly expressed miRNAs in maintaining pluripotency in ESCs.

The miR-302 cluster was identified in Elcie's sequencing data and many others as being highly expressed in ESCs. Key transcriptional regulators Oct4 and Sox2

are required for the transcriptional regulation of the miR-302 family, and the miR-302 cluster is a known regulator of Oct4 and Sox2 forming an intricate feedback mechanism regulating ESC pluripotency (Marson et al., 2008, Card et al., 2008, Lin et al., 2008, Subramanyam et al., 2011, Liao et al., 2011, Judson et al., 2009). In 2008, Lin et al showed that by retroviral transduction of the miR-302 cluster in human skin cancer cells, they were able to recapitulate the gene expression profile of ESCs and iPSCs, however typical ESC morphology could not be visualised (Lin et al., 2008). Following this, it was discovered that the miR-302 cluster in humans, and a number of other miRNA clusters in mice, were able to enhance iPSC reprogramming of mouse and human fibroblasts, when used in conjunction with Oct4, Sox2 and Klf4 reprogramming factors (Lin et al., 2008, Subramanyam et al., 2011, Liao et al., 2011, Judson et al., 2009).

Recent reports claim that when performing lentiviral transduction, the miRNA-302 cluster alone can achieve full reprogramming of mouse and human somatic cells, with up to two times more efficiency than the transcription-factor based method (Anokye-Danso et al., 2011, Miyoshi et al., 2011). In addition to the reprogramming ability of the miR-302 cluster, it has been suggested that relative expression of miR-302b could serve as a biomarker for the developmental stage of ESCs and iPSCs, as miR-302b is strictly spatially and temporally regulated in ESCs (Berardi et al., 2012).

MYC also binds and up-regulates the ESC-specific miR-17 cluster, (Suh et al., 2004) and acts to repress the expression of miR-34 and miR-29, suggesting that ESC-specific transcription factors and miRNAs act mutually (Leonardo et al.,

2012). ESCs have a very short cell cycle which is essential to their pluripotency; their lack of G1/S checkpoint means a short G1 phase resulting in highly proliferative cells (Becker et al., 2006). Many ESC-specific miRNAs have mRNA targets involved in the cell-cycle pathway e.g. cyclin D1 and retinoblastoma-like 1 (Card et al., 2008, Leonardo et al., 2012, Subramanyam et al., 2011), and the majority of highly expressed miRNAs in ESC contribute to cell-cycle regulation by accelerating the G1 to S transition (Leonardo et al., 2012).

MicroRNA-let-7 was the first miRNA shown to play vital roles in development across a range of species (Bagga et al., 2005, Thornton and Gregory, 2012). Unlike miR-302, miR-let-7 is depleted in ESCs and upregulation of the miR-let-7 family promotes differentiation of ESCs by repressing "stemness factors" Lin28 and c-Myc (Thornton and Gregory, 2012). Careful regulation of miR-let-7 is evident by the extensive regulatory feedback mechanism with Lin28 (discussed in Ch 1.2.5). This suggests that fine tuning of miRNA biogenesis is important in the maintenance and differentiation of the ESC state. These studies demonstrate the importance of miRNAs in ESC differentiation as well as somatic cell reprogramming.

Comprehensive miRNA profiling of ESCs has produced distinctive expression patterns specific to ESCs. ESC-specific miRNAs are more likely to be arranged in genomic clusters, they have higher oncogenic miRNA gene expression, and highly expressed miRNAs share a consensus seed sequence indicating mutual regulation of specific target-mRNAs (Laurent et al., 2008). Although there is currently limited literature about post-transcriptional regulation, specifically in

human ESCs, it has been reported that expression of many primary miRNA transcripts during early mouse development is blocked post-transcriptionally at the Microprocessor stage (Thomson et al., 2006).

Given the significant impact of miRNAs in biological processes it is not surprising to learn that miRNAs are under significant control and are highly regulated. Despite this, misregulated miRNA processing does occur resulting in down-regulation of miRNAs in many cancers and other diseases highlighting the importance of research into regulation of miRNA (Thomson et al., 2006, Garzon et al., 2009).

My aim is to use the deep sequencing database generated by Dr Elcie Chan in conjunction with other available sequencing databases, to investigate miRNAs that are highly or lowly expressed in ESCs and differentiated NPSCs. Some of these specific miRNAs are known to play a clear role in maintaining pluripotency and promoting neural differentiation however, the regulatory mechanism responsible for the expression of the majority of miRNAs is not yet known. Currently, the methods of post-transcriptional regulation of only a few miRNAs are known. Consequently there is a severe lack of understanding about how the majority of miRNAs expressed in ESCs get downregulated upon differentiation. Likewise, the mechanism by which miRNAs that promote differentiation get upregulated is not completely understood. Therefore I intend to uncover any miRNA regulation patterns that are distinct to either ESCs or NPSCs in an attempt to uncover the mechanism behind miRNA-mediated ESC differentiation.

1.3 SPLICING AND THE SPLICEOSOME

1.3.1 RNA splicing and the spliceosome

In 1977 it was discovered that many eukaryotic genes contain non-coding intervening sequences (introns), that interrupted protein-coding sequences (exons) (Alberts, 2002). The entire gene undergoes transcription into precursor mRNA, requiring further processing into mature mRNA (Alberts, 2002). Precursor mRNA splicing can take place co-transcriptionally whilst the RNA is still bound to DNA by RNA polymerase II (Brody and Shav-Tal, 2011, Listerman et al., 2006). The spliceosome machinery is tightly regulated with an excision time *in vivo* of less than 3 minutes after transcription by RNA polymerase II (Beyer and Osheim, 1988). Pre-mRNA processing therefore takes place in the nucleus as do further distinct processing steps including 5' capping, 3' end processing and 3' polyadenylation (Alberts, 2002). Following the processing of pre-mRNA, the mature mRNA is exported into the cytoplasm where translation takes place (Alberts, 2002).

The spliceosome is an evolutionarily conserved multi-megadalton, macromolecular complex, that catalyzes the splicing reaction of pre-mRNA (Wahl et al., 2009, Rino and Carmo-Fonseca, 2009, Galej et al., 2013). A major and a minor form of the spliceosome have been characterised, designated to target U2- or U12-type introns (Will and Luhrmann, 2005). The U2-dependent spliceosome is the main form found in all eukaryotes and catalyses the excision of the most commonly found U2-class of introns. This spliceosome is composed of five small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4/U6 and U5), each

of which are composed of specific small nuclear RNAs (snRNAs) and associated proteins. In addition to snRNPs, the splicing machinery interacts with over 200 distinct proteins including RNA binding proteins and enzymes (Wahl et al., 2009, Rino and Carmo-Fonseca, 2009, Bonnal et al., 2012, Cooper et al., 2009).

Typically the splicing machinery recognises pre-mRNA substrates by "strength" (i.e. binding complementarity) of the 5' and 3' splice site consensus sequences, (A/C)AG/GUAAGU and (C/U)$N^{<10}$(C/T)AG/G respectively, situated at the ends of each intron (Buvoli et al., 2007, Roybal and Jurica, 2010, Kaida et al., 2007). Along with the splice site sequences, the branch site consensus sequence (UAUAAC) and other silencer and enhancer signals are required to regulate splicing. The branch site is located upstream of the 3' splice site flanking a specific adenosine base which is involved in catalyzing the two-step enzymatic pre-mRNA splicing reaction (Wahl et al., 2009, Alberts, 2002). Recognition of the 5' splice site by the U1 snRNP initiates gradual spliceosomal assembly on pre-mRNA transcripts. This is followed by the recognition of the branch site by U2 snRNP and subsequent binding at the polypyrimidine tract resulting in the recruitment of preassembled U5/U4/U6 snRNP (Wahl et al., 2009, Rino and Carmo-Fonseca, 2009). Subsequently catalytic activation of the spliceosome is achieved by conformational and compositional rearrangement resulting in the formation of a structured RNA network between the branch sequence of the pre-mRNA and U2, U5 and U6 snRNPs (Wahl et al., 2009). This initiates two transesterifcation reactions; initially the hydroxyl group of the adenosine flanking the branch site attacks the phosphodiester bond at the 5' splice site generating a free 5' exon and an intron lariat 3' exon (Wahl et al., 2009).

Consequently, the 5' exon hydroxyl group attacks the phosphodiester bond at the 3' splice site causing exon ligation and excision of the lariat intron (Wahl et al., 2009) (see figure 1.6). Following the 2nd catalytic reaction the spliceosome dissociates from the mRNA (and associated proteins) as well as releasing U2, U5 and U6 snRNPs to be reused in future splicing reactions (Wahl et al., 2009). This model of stepwise spliceosomal assembly on pre-mRNA substrates is conserved and has been derived from multiple *in vitro* and *in vivo* biochemical studies in yeast and mammalian systems (Rino and Carmo-Fonseca, 2009). It is possible that another model for spliceosome assembly exists based on evidence that the five snRNPs can associate and remain assembled in the absence of pre-mRNAs (Wahl et al., 2009, Rino and Carmo-Fonseca, 2009).

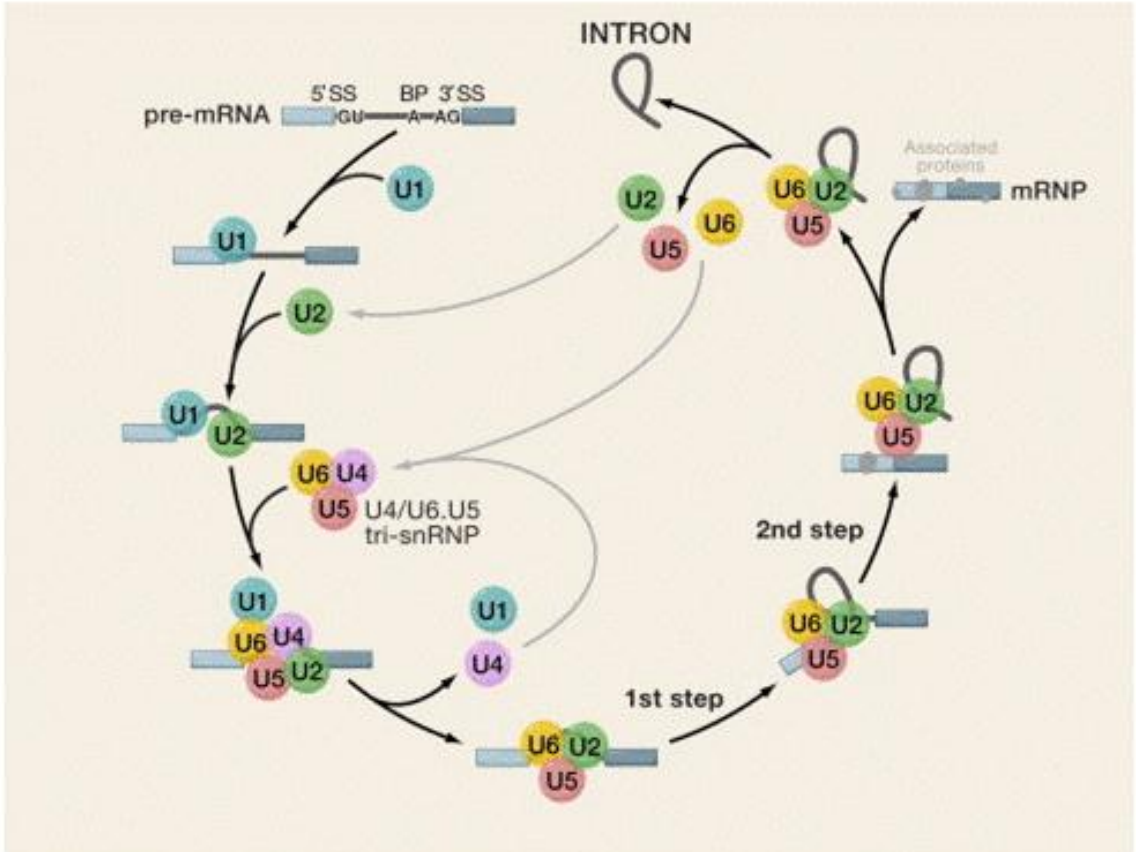


Figure 1.6 Schematic of pre-mRNA splicing by the spliceosome: Pre-mRNA is represented by two exons (blue boxes) separated by an intron (grey line) and the 5'splice site, branch site and 3' splice site are labelled. The sequence of the snRNP and pre-mRNA associations resulting in the removal of the intron from the pre-mRNA and the release of the mRNA with associated protein is displayed, adapted from (Wahl et al., 2009).

The less abundant U12-dependent spliceosome catalyses the removal of U12-type introns which are present in less than 1% of all human mRNAs and other species contain them including plants and animals (Will and Luhrmann, 2005, Koonin, 2006). Although both U2- and U12-dependent spliceosomes co-exist in cells and act upon pre-mRNAs with both intron types, the target branch sites and splice sites have different consensus sequences. The sequences for the U12-type 5' splice site (A/GUAUCCUUU) and the branch site (UCCUUAACU) are closer together in distance; longer in sequence and more highly constrained than U2-type sites. In addition contrary to the U2-type, the sequence of U12-type 3' splice site can be variable. The minor U12-dependent spliceosomes assembles and removes U12-type introns in an identical mechanism to the major U2-dependent spliceosome and parallel snRNPs are involved. Other than differences in splice site recognition and snRNP composition mechanistic variations are limited and the U12-dependent spliceosome remains relatively uncharacterized compared to the U2-type spliceosome (Will and Luhrmann, 2005).

1.3.2 Splice site selection and alternative splicing

The conformation and the composition of the spliceosome is highly dynamic, making the machinery accurate and flexible. The choice of 5' and 3' splice sites by the spliceosome, plays a major role in the resultant mature mRNA and subsequent protein expression. Splice site selection is regulated by a number of factors and the relative binding affinity of the splice site for U1 and/or U2 snRNP plays a significant part however, this alone is not capable of directing spliceosome assembly (Wahl et al., 2009).

As previously mentioned, the *cis*-acting flanking pre-mRNA enhancer and silencer sequences play a huge role in determining splice site usage by binding *trans*-acting factors that either recruit snRNPs to the splice site or prevent snRNP binding (Singh and Valcarcel, 2005). Serine/Arginine rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) are trans-acting factors that exhibit exonic splicing enhancing (ESEs) or exonic splicing silencing (ESSs) properties respectively (see figure 1.7). Furthermore some regulatory factors can regulate both ways depending on the sequence and position of target pre-mRNA (Zhu et al., 2001, Caceres et al., 1994, Kornblihtt et al., 2013).

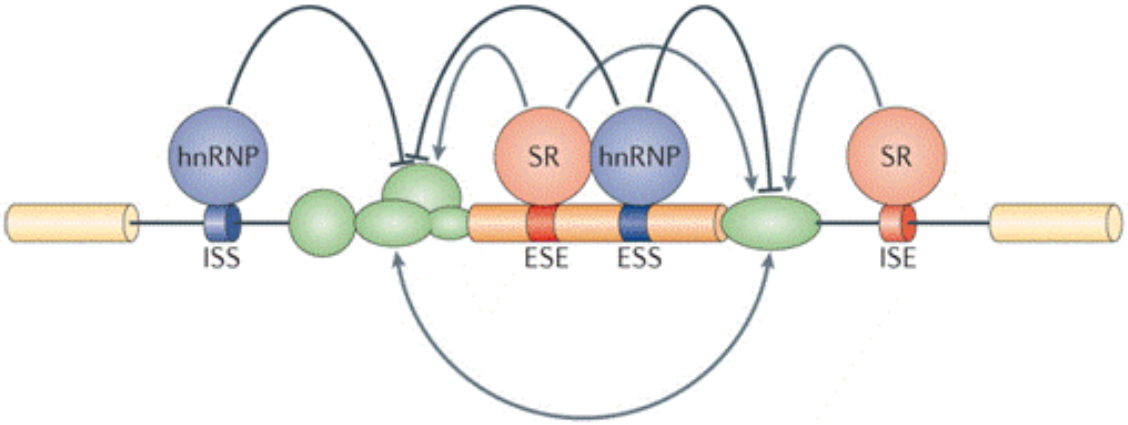


Figure 1.7 Splice site recognition. Splice site recognition is determined by *cis*-acting sequences in the pre-mRNA. Ser/Arg-rich protein (SRs) and heterogeneous nuclear ribonucleoproteins (hnRNPs) bind to 5' and 3' splice site regions flanking the exon and target components of the spliceosome (shown in green). This can cause activation or inhibition on the recognition and use of the specific splice sites by the spliceosome. (exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs),intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs)) Adapted from (Kornblihtt et al., 2013)

These RNA binding proteins act together in an opposing manner, exerting both silencing and enhancing effects that collectively directs splice site recognition by the spliceosome (Wahl et al., 2009) (see figure 1.7). This external regulatory feature of pre-mRNA splicing is not present in yeast genomes, where the 5'splice site and branch site consensus sequences bind with perfect complementarity to U1 and U2 snRNPs. In mammalian genomes perfect complementary binding of the branch sites and splice sites to the spliceosome components is not necessary for recognition allowing flexible selection of splice sites, forming the basis for alternative splicing.

Alternative splicing is an important and evolutionarily advantageous mechanism, responsible for the proteomic complexity of higher organisms allowing the human genome to create a huge variety of distinct mRNA transcripts from only 20,000 genes (Kornblihtt et al., 2013). Messenger RNA selection by the use of alternative splice sites, is a process that is often subject to tissue-specific or developmental control (Mayeda and Krainer, 1992). Alternative splicing is the primary mechanism for generating mRNA diversity although other mechanisms exist such as alternative transcription start sites, alternative polyadenylation and RNA editing (Nilsen and Graveley, 2010, Licatalosi and Darnell, 2010). Production of variable forms of mRNA from a single pre-mRNA transcript, by alternative splicing, is predicted to generate proteins with differing and sometimes even opposing functionality (Nilsen and Graveley, 2010, Licatalosi and Darnell, 2010). Furthermore alternative splicing is able to regulate specific mRNA downregulation, by introducing premature termination codons, resulting in nonsense mediated decay or regulation of *cis*-acting elements (Nilsen and

Graveley, 2010, Licatalosi and Darnell, 2010). Results of genome-wide RNA sequencing show that alternative splicing occurs in nearly 95% of human genes, with utilization of the varying splice sites dependent on physiological development and external stimuli, potentially impacting all aspects of biology (Wang et al., 2008, Kalsotra and Cooper, 2011, Kornblihtt et al., 2013, Mironov et al., 1999)

Alternative splicing patterns may change in response to external stimuli, implying functional significance. Yet alternative splicing may not always result in development modifications, rather it may cause a shift in ratios of mRNA isoforms, allowing "fine-tuning" of protein expression. Consequently the extent of functional impact of alternative splicing is not entirely understood, and the diverse mRNA profile includes thousands of low abundance transcripts that are not conserved. This indicates that a large number of alternative splicing events are potentially stochastic and unregulated (Pickrell et al., 2010).

Database analysis and deep RNA sequencing have verified that alternative splicing is widespread and that there is evidence for a large class of low level intraexonic splicing, and alternative 5' and 3' splicing, that uses non-canonical splice sites, which do not necessarily conform to the GT-AG splice site consensus of most introns (Ng et al., 2004, Pickrell et al., 2010). It is not clear whether non-canonical splicing is catalysed by the splicing machinery and for this reason the extent of non-canonical splicing remains uncharacterised. The splicing machinery is highly conserved and known to be present and functional in a variety of yeast species whose genomes contain very few introns. This suggests

extended functionality of the spliceosome beyond just splicing and corroborating evidence supports that the macromolecular complex is involved in gene transcription, mRNA biogenesis and nonsense mediated decay (Roy and Gilbert, 2006, Le Hir et al., 2000, Pleiss et al., 2007).

1.3.3 Roles of splicing in disease

Splicing is essential to understanding gene expression and is also of medical relevance. Translation of erroneously spliced pre-mRNA leads to aberrant proteins, which forms the basis of many diseases (Will and Luhrmann, 2011, Nilsen and Graveley, 2010). Splicing is regulated by many factors including numerous sequencing codes, several snRNPs and RNA binding proteins. It is this complex set of interactions which cause increased chance of exposure to mutations that can cause mRNA misregulation.

Aberrant pre-mRNA splicing is one of the most common mechanisms underlying human genetic disease and cancer disorders (Haj Khelil et al., 2008, Cooper et al., 2009). A large number of diseases can be caused by missense, nonsense and frame-shift mutations in exons of protein coding genes, as well as mutations in mRNA UTRs that affect translation and mRNA stability (Cooper et al., 2009). It is predicted that more than a third of disease-causing SNPs are able to disrupt splicing (Singh and Cooper, 2012) and for some genes, up to 50% of point mutations that cause disease lie within splicing consensus sequences (Cooper et al., 2009).

To identify putative disease causing pre-mRNA splicing mutations, available bioinformatic tools provide estimates on potential disruption of 5' and 3' splice sites, branch sites, enhancer and silencer sites and other features such as protein binding sites and RNA secondary structure. The dbass database (http://www.dbass.org.uk/) contains a detailed list of splicing-associated diseases with 577 5' splice site mutations and 307 3' splice site mutations recorded to date. Mutations that target 5' or 3' splice sites, the polypyrimidine tract, branch sites, enhancers and/or silencers most frequently result in exon skipping however intron retention can also occur (Singh and Cooper, 2012). Additional splicing abnormalities can occur where 5', 3' or both splice sites can undergo mutation, weakening the canonical splice sites, resulting in the specific activation of alternative dormant splice sites known as cryptic splice sites or induce splicing of intronic sequences known as pseudoexons (Singh and Cooper, 2012). Both events lead to inclusion of additional sequence in the spliced mRNA, that may act in preference to existing splicing enhancer or silencer sequences, or disrupt the mRNA reading frame producing defective proteins (Singh and Cooper, 2012).

Trans-acting splicing factor genes responsible for regulating alternative splicing are also exposed to mutations which can affect a large set of genes, rather than singular mRNA targets by mutations in *cis*-regulatory elements (Singh and Cooper, 2012). Genome-wide approaches have linked changes in alternative splicing profiles to tumorgenesis, with alteration of approximately half of all alternative splicing events in ovarian and breast cancer tissue (Venables et al., 2009). Alternative splicing contributes to cancer progression with roles in cell

proliferation, cell death and angiogenesis (Bonnal et al., 2012). An example of alternative splicing regulation during angiogenesis is the choice between alternative 3' splice sites of vascular endothelial growth factor (VEGF) isoforms which assist or impede blood vessel formation essential for tumour growth and spreading (Harper and Bates, 2008).

Misregulation of RNA splicing due to disease-causing mutations is generating an array of new therapeutics targeting the splicing factors and the spliceosome. An abundance of data has highlighted the role of splicing components as targets of novel therapeutics in a variety of diseases (Bonnal et al., 2012). Efforts are underway to target abnormal pre-mRNA splicing. This includes nucleic acid based therapy, such as antisense oligonucleotides, that bind competitively to splicing enhancers to prevent recognition (Hua et al., 2010, Bonnal et al., 2012). This method is being developed for the treatment of spinal muscular atrophy (SMA) which is caused by the disruption of SMN1 (survival of motor neuron 1) (Hua et al., 2010, Bonnal et al., 2012). The human genome makes an almost identical gene, SMN2, except for a nucleotide change that causes skipping of exon 7 (Hua et al., 2010, Bonnal et al., 2012). The modified antisense oligos base pairs with intronic splicing silencer sequences that inhibit the binding of splicing repressors preventing skipping of exon 7 and restoring the expression of SMN1 mRNA (Hua et al., 2010, Bonnal et al., 2012). The splicing defect can be corrected in SMA mouse models and in cell-free splicing assays in cultured human cells (Kornblihtt et al., 2013, Hua et al., 2010, Bonnal et al., 2012).

1.3.4 Splicing inhibitors

Recently an abundance of novel inhibitors of splicing have been identified and purified from the bacterial species *Pseudomonas* and *Streptomyces* as well as other sources (Bonnal et al., 2012).

Spliceostatin A (SSA) is a synthetic analogue of FR901463 FR901464 and FR901465 originating from a fermentation broth of *Pseudomonas* spp (Kaida et al., 2007). SSA displays cytotoxic effects in multiple tumour cell lines resulting in arrest in the G1 and G2/M phases of cell cycle (Kaida et al., 2007). SSA is proven to have improved activity, stability and solubility *in vitro* compared with its naturally occurring counterparts (Kaida et al., 2007). Using biotinylated SSA, it was discovered that splicing factor 3B (SF3b), a sub complex of the spliceosomal U2 snRNP, most tightly bound to streptavidin beads linked to SSA (Kaida et al., 2007). In addition knockdown of SF3b1 expression, mimics many of the drug induced effects, confirming SSA inhibition of the SF3B complex (Kaida et al., 2007). SF3b is a 5-polypeptide complex and the exact binding location has not yet been determined and remains to be fully elucidated. (Bonnal et al., 2012). Although SF3b3 is the most highly enriched component bound to SSA in a complex, the SF3b1 displays the tightest binding to SSA in competition experiments (Bonnal et al., 2012). With this in mind it is possible that the binding pockets of SSA involve an interface between the two SF3B subunits (Kaida et al., 2007, Bonnal et al., 2012). SSA association with the SF3b subunit blocks binding to pre-mRNA, resulting in a conformational change that compromises RNA binding efficiency and disables the proofreading mechanism (Kaida et al., 2007, Bonnal et al., 2012). This results in destabilisation of the U2

snRNP and their complexes with pre-mRNA preventing subsequent spliceosomal assembly (Kaida et al., 2007).

Despite targeting basic steps in splice site recognition, SSA appears to have a cytostatic (not cytotoxic) effect on cells with an enhanced effect in cancer cells (Kaida et al., 2007). This could be due to a number of factors including cellular permeability; splicing demand and lack of effect on U12-dependent splicing (Bonnal et al., 2012).

Another more characterised splicing inhibitor, TG003, is a benzothiazole inhibitor of the serine/arginine rich (SR) protein kinases CLK1 and CLK4. During splicing, SR proteins are highly phosphorylated and by this method are able to catalyse splicing reactions and spliceosome formation. TG003 inhibits the CLK-mediated phosphorylation of SR proteins and consequently inhibits the SR protein regulation of pre-mRNA splicing (Muraki et al., 2004, Hagiwara, 2005).

Pre-mRNA splicing inhibitors are leading the field in therapies against disease caused by deregulation of splicing. For example TG003 is of potential relevance for the treatment of Duchenne muscular dystrophy, as it is able to promote skipping of a mutant exon in the dystrophin gene, but not of the wild-type exon (Bonnal et al., 2012). SSA and related compounds display anti-tumoural properties with some, e.g. cisplatin, that are currently used in chemotherapy (Bonnal et al., 2012). Thus SSA may potentially be used not only as a powerful

tool for analysing pre-mRNA splicing but also for developing new therapeutics (Kaida et al., 2007).

Kaida et al 2007 report that SSA effectively inhibits the U2 component SF3b from binding to pre-mRNA thereby inhibiting splicing completely. However Kaida et al 2010 and Younis et al 2010 report that although SSA was able to inhibit splicing fully *in vitro,* the observed inhibitory action of SSA *in vivo* appeared to be partial even at high concentration (Kaida et al., 2010, Younis et al., 2010). The focus of my research is to establish a working assay for SSA *in vivo* in order to further investigate the properties of the novel inhibitor and its effects on RNA and the spliceosome.

1.3.5 Splicing and microRNAs

Recent evidence suggests an evolutionary connection between the RNA splicing machinery and miRNA mediated gene silencing (Dumesic et al., 2013, Tabach et al., 2013). Splicing proteins and miRNAs co-localize to P bodies and Cajal bodies, supporting the idea of functional cross-talk between splicing and miRNA gene silencing (Tabach et al., 2013).

Tabach et al 2013 suggest that a large number of proteins involved in intron removal are also required for miRNA interference, and they claim that genomes with few introns do not need miRNA regulation either. In addition there is a strong correlation between the presence of Ago proteins (and other miRNA components) and the number of introns a species has (Tabach et al., 2013).

Tabach et al 2013 suggests that miRNA pathway components such as Argonaute, can engage with pre-mRNA to provide another layer of splicing surveillance. Without cross-talk between Ago proteins and pre-mRNA the selective advantage of introns could fade too (Tabach et al., 2013).

Dumesic et al 2013 reveals that stalled spliceosomes promote the expression of small silencing RNAs in *Cryptococcus neoformans,* reinforcing the concept of functional cross-talk between the spliceosome and RNAi machinery (Dumesic et al., 2013). Mutations of the pre-mRNA 3' splice site resulted in increased production of siRNA however mutation of the 5' splice site did not affect siRNA production indicating that spliceosome initiation was the major player in siRNA generation. Inefficient splicing of pre-mRNAs means a greater dwell time on spliceosomes which provides the substrate for siRNA synthesis (Dumesic et al., 2013). MicroRNA regulation together with splicing mechanisms form part of a mRNA surveillance pathway consisting of synonymous components that act as quality control of mRNA splicing.

1.4 PROJECT AIMS

1.4.1 MicroRNA regulation in ESCs

To investigate the regulatory mechanisms, which are responsible for the observed differences in microRNA expression, between ESCs and differentiated NPSCs.

1.4.2 Alternative polyadenylation of miRNA genes

Whilst investigating miRNA regulation, I discovered an extensive use of alternative polyadenylation sites in miRNA genes. This generated a second aim, to uncover the extent of alternative polyadenylation within miRNA genes, and to establish if this mechanism is a potential method for post-transcriptional regulation of miRNAs.

1.4.3 Establishing a working assay for novel splicing inhibitor, Spliceostatin A

To establish a better assay of the effect of splicing inhibitors, *in vivo,* in order to establish the global effects of spliceosome inhibition upon splicing in human cells and to facilitate the development of splicing inhibitors for therapeutic purposes.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Cell Culture

2.1.1 Cell culture of HEK293, HT1080 and HeLa cells

Cells were cultured in D10 media which contains Dulbecco's modified eagle's medium (DMEM) with 10% (v/v) heat inactivated fetal bovine serum (FBS), supplemented with 2 mM L-glutamine, 50 U/ml penicillin and 50 μg/ml streptomycin, at 37°C in 10% CO_2 (Invitrogen, Gibco). Generally, cells were passaged twice a week or once they reached ~80% confluency by washing in PBS and then incubating with 0.25% trypsin (Invitrogen, Gibco) for ~5 minutes at 37°C in 10% CO_2. The cells were then re-suspended in D10 media and centrifuged at 1200 rpm for 5 mins. The cell pellet was resuspended in D10 and plated to the required density.

2.1.2 Human embryonic stem cells and neuroprogenitor stem cells

Human embryonic stem cell (ESCs) lines designated H7 and H1 and their differentiated neuroprogenitor stem cells (NPSCs) were kindly given to me by Dr Wei Cui (Imperial College London),

NPSCs were grown on poly-L lysine (Sigma) and laminin (Invitrogen) coated plates. 6-well plates were coated with 1 ml of 16.7 μg/ml of poly-L lysine in PBS for 1 hr at RT. The poly-L lysine solution was aspirated and 1 ml of 20 μg/ml of laminin in PBS was added to each well and incubated at 4°C overnight. The laminin solution was aspirated before the cells were seeded. NPSCs were grown in N2B27 medium containing 1:1 of DMEM and neurobasal medium

supplemented with 1% B27, 0.5% N2 and 1% L-glutamine (Invitrogen). The medium was replaced every 2 days and 20 ng/ml of EGF and 20 ng/ml of bFGF were added fresh with every passage. To passage the NPSCs, TrypLE express (Invitrogen) was used to separate NPSCs into single cells.

The H1 and H7 designated ES cells were cultured in conditioned mouse embryonic fibroblasts (MEFs) and SR medium (knockout DMEM and knockout serum replacement) with 8 ng/ml bFGF, 50 U/ml penicillin/streptomycin (Invitrogen, Gibco) and 200 μM glutamine (Invitrogen, Gibco) on matrigel-coated plates (Invitrogen, Gibco). The medium was changed daily and cells were maintained in ES cell colonies at confluence.

2.1.3 Freezing cell lines

Adherent cells were washed with PBS and then incubated with 1 ml of 0.25% trypsin. The trypsin was inactivated with 9 mls of D10 media and then the cells were centrifuged at 1000 rpm for 5 minutes. The pellet was re-suspended in D10 media with 10% dimethyl sulfoxide (DMSO) and instantly aliquoted into 1 ml cryotubes. The cells were frozen in a cryo-freezing container containing isopropyl alcohol at -80°C for at least 48 hrs and then stored in the -80°C freezer or in liquid nitrogen.

2.2 GENERAL RNA AND DNA PROTOCOLS

2.2.1 Total RNA extraction

Trizol was used as per manufacturer's instructions (Invitrogen) to lyse cells and extract total RNA. 6 x 35 mm diameter dishes of confluent cells could generate 20 µg - 100 µg of total RNA depending on the cell type. 1 ml of trizol reagent was used per 35mm well, and then 200 µl chloroform was added to isolate total RNA. The isolated total RNA was then precipitated with isopropanol and washed with 80% ethanol before being resuspended in 30 µl of nuclease-free water. 1 µl of the sample was removed for quality check of 28S and 18S RNA on a 1% agarose gel. 1 µl of the sample was measured in a UV spectrophotometer. All RNA samples were stored at –80^{o}C. The quality of RNA samples was checked on a 1% TBE agarose gel containing 0.1 µg/ml of Ethidium Bromide (Sigma). The samples were run at 100 V for 1 hr and the quality of the 28S and 18S rRNA bands were checked under UV light.

2.2.2 Small RNA enrichment

Approximately 100 µg of total RNA was enriched for small RNAs using the miRVana kit (Ambion) according to the manufacturer's instructions. Small RNAs were precipitated with an equal volume of 100% isopropanol and 3 µl of glycogen (Ambion glycoblue) and incubated at –80^{o}C for 30 minutes. The RNAs were centrifuged at max speed in a microcentrifuge at 4^{o}C. The pellet was

washed with 80% ethanol and after air drying the RNA, the pellet was dissolved in a total of 10 µl of nuclease-free water (Ambion).

2.2.3 cDNA synthesis

20-100 ng of RNA was used for the reverse transcription reaction into cDNA. The sample of RNA, 0.5 µg of random hexamers and 1 µl of 10 mM dNTPs were combined with water to make a total of 35µl and the solution was denatured at 65°C for 5 mins. 1 µl of 100 mM DTT, 2 µl of 5 x first strand buffer, 1 µl of RNAseOUT (40 U/µL), and finally 1 µl of SuperScript™ III RT enzyme (200U/µL) (Invitrogen) were added, making the solution a total of 40µl. The sample was incubated at 42°C for 30 mins, then 50°C for 30 mins and stored at -20°C.

2.2.4 Polymerase Chain Reaction (PCR)

1 µl of the RT sample (see chapter 2.2.3) was added to 1 µl of each of the gene specific forward and reverse primers (10 uM), 1 µl of 10 mM dNTPs, 2 µl of 10 x buffer and 1 µl of Taq polymerase (New England Biolabs) and made up to 20 µl with water. The PCR reaction was set at 94°C for 2 mins, followed by 35 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, and terminated at 72°C for 10 mins (extension time is based on a rate of extension of 1 kb/min).

PCR products were run on TBE agarose gels or 5% PAGE to check the product sizes. 1-2% (w/v) agarose, depending on product size, was dissolved in 1x TBE and cooled before adding 0.1 µg/ml ethidium bromide (Sigma-Aldritch). To make the 5% PAGE gels 2.5 ml 10 x TBE, and 6.75 ml of 40% (w/v) 19:1 acrylamide:bis-crylamide was combined and adjusted to 50 ml with water. Then

350 µl of 10% (w/v) ammonium persulphate (APS) and 17.5 µl of TEMED were added and the gel solution was mixed and poured into cassettes (Invitrogen). Samples were mixed with 6 x loading dye containing bromophenol blue and then loaded into the wells of the gel along with a 1kb ladder (New England Biolabs), 100bp ladder (New England Biolabs) or a 50 bp ladder (New England Biolabs). Gels were run at 100V for ~1 hour and then visualised for bands on a UV light box.

2.3 DNA EXTRACTION AND SEQUENCING

2.3.1 DNA gel elution

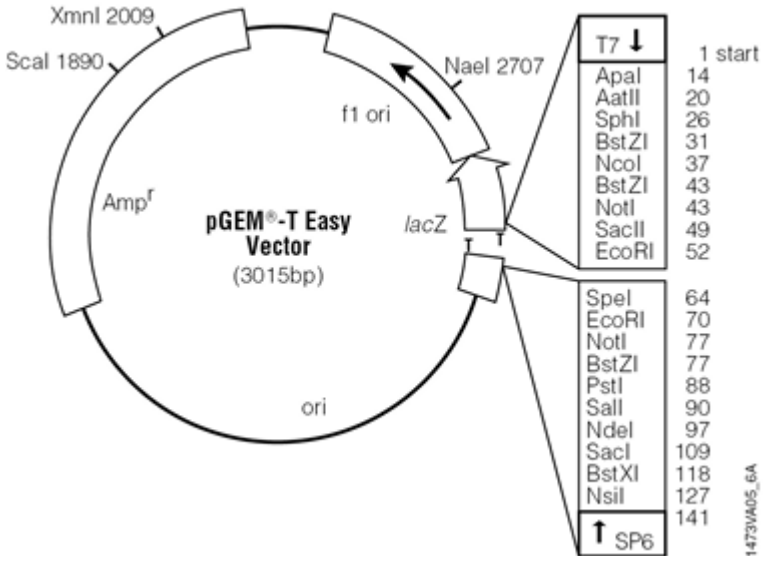
The DNA band of expected size was excised from the gel and eluted with 3 x volume of DNA elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% w/v SDS), at 16^{o}C under agitation overnight or 37 ^{o}C for >2hr. The solution was filtered using spin-x centrifuge filter tubes with a 0.22 µm pore and the remaining gel removed. The solution was precipitated with 3 x volume of -20^{o}C ethanol, plus 0.1 x volume of 3 M sodium acetate and 1 µl of glycogen, overnight at -20^{o}C. The DNA pellet was resuspended in 10 µl of water.

2.3.2 Vector cloning

The DNA eluted from the gel was cloned into pGEM-T Easy vector (Promega – see vector map below) by mixing 3.5 µl of the DNA with 0.5 µl of pGEM-T Easy vector, 5 µl of 2 x rapid ligation buffer and 1µl of T4 DNA ligase. The ligation reaction was incubated at room temperature for 2 hrs, and added to 50 µl of

Dh5α competent cells and incubated on ice for 1 hour. The cells were subjected to heat shock for 30 sec at 42°C, and then placed immediately on ice. 200 µl of YT medium was added to the cells and incubated at 37°C, agitating at 250 rpm for 30 minutes. LB plates supplemented with ampicillin were pre-warmed to 37°C, and 100 – 200 µl of the cells were spread evenly onto the plate and incubated at 37°C overnight.

Vector used for DNA sequencing: pGEM-T easy vector were used in the cloning of RT-PCR products (taken from Promega manual).



2.3.3 Colony PCR

The PCR mix per sample consisted of 2 µl of 10 x Taq polymerase buffer, 0.4 µl of 10 mM dNTPs, 0.2 µl of both the M13 forward and reverse primers (100 µM), 0.73 µl of 1% w/v W-1 detergent (Sigma Brig-35P) and 0.2 µl of Taq polymerase enzyme was mixed together and made up to 20 µl with water. A p10 pipette tip was used to swipe single colonies on the plate and placed inside a PCR tube containing the 20 µl PCR mix. The solution was pipetted repeatedly to ensure sufficient mixing of the solution in each PCR tube and the tube was spun briefly.

The PCR reaction was set at 94°C for 2 mins, followed by 35 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, and terminated at 72°C for 10 mins. A sample of 10 µl of each of the PCR products was then checked on a 1.5% TBE agarose gel for amplification.

Shrimp alkaline phosphatase (SAP) and exonuclease I (Exo I) reaction mix was set up by mixing 2 µl of SAP (1 U/µl) with 0.2 µl of Exo I (20 U/µl) and 12.8 µl water. This was then added to which was added to the 10 µl PCR mix. The samples were then incubated at 37°C for 30 mins, and inactivated at 80°C for 10 mins.

2.3.4 Sequencing

The pre-sequencing PCR was prepared by mixing 4 µl of each sample with 1 µl of M13 forward primer (3.2 pmol/µl) and 2 µl of Big Dye Terminator (BDT) 3 µl water. The PCR thermocycler was set at 94°C for 2 mins, followed by 35 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec, and terminated at 72°C for 10 mins.

8 mls of Big Dye precipitations solution was made with 6.25 mls of 100% ethanol, 0.3 mls of 3 M sodium acetate and 1.45 mls of water. 40 µl of the precipitation solution was added to each of the pre-sequencing PCR products and left at room temperature for 20 mins. The samples were centrifuged at 4000 rpm for 40 mins, and the supernatant was discarded. The invisible pellet was washed with 50 µl of 70% ethanol and centrifuged at 4000 rpm for 5 mins. The supernatant was discarded and the samples were left to dry at 37°C for 30 mins. Samples were

submitted for in-house sequencing with ABI Prism Big Dye Terminator cycle sequencing reaction kit.

2.4 RACE

2.4.1 5'rapid amplification of cDNA ends (5'RACE)

The 5'RACE assay was carried out using the Ambion FirstChoice RLM-RACE kit (Invitrogen AM17000) as per manufacturer's instructions and using reagents provided in the kit unless otherwise specified. Primers used were either provided in the kit or uniquely designed to be gene-specific, they can be found listed in section 2.10.5 under 5'RACE primers.

10 µg of total RNA was treated with Calf Intestinal Phosphatase (CIP) for 1hr at 37^{o}C, and the reaction was stopped by adding ammonium acetate. The solution was then subject to acid phenol-chloroform extraction and the aqueous phase containing RNA was precipitated with equal volume of isopropanol for 10 mins at RT. This solution was then centrifuged at max speed for 20 mins at 4^{o}C. The precipitated pellet was washed in 80% ethanol; air dried and resuspended in 10 µl RNAse-free water. The sample was then treated with Tobacco Acid Pyrophosphatase (TAP) for 1 hr at 37^{o}C. The RNA was then ligated to the 5'RACE adapter with T4 RNA ligase for 1 hr at 37 ^{o}C and half of the sample was reverse transcribed with M-MLV Reverse Transcriptase and random decamers for 1 hr at 42^{o}C. This sample was then ready to undergo nested PCR, first with the 5'RACE gene-specific outer primer followed by the 5'RACE gene-specific inner primer. Both PCR cycles followed with an initial denaturing step at 94^{o}C for 3 mins, then

35 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec, and a final extension of 72°C for 7 mins. 5'RACE PCR products were visualised on a 5% PAGE gel stained with ethidium bromide and all bands were excised for sequencing in accordance with the sequencing protocol (Ch 2.3.4). See Figure 2.1 for an overview of the 5'RACE procedure and the position of primers used in the 5'RACE PCR.

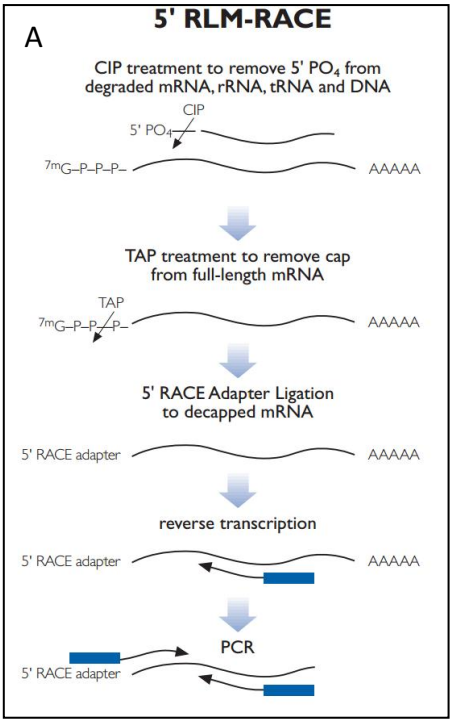
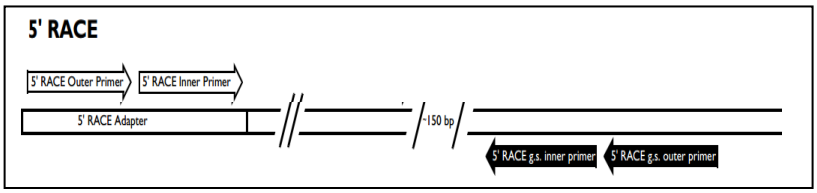


Figure 2.1 Overview of the 5'RACE protocol and 5'RACE amplification of transcripts. A) Diagram of the overview of 5'RACE procedure described in ch 2.4 B) Diagram of primer positions for 5'RACE, white arrows denote the position of primers supplied by the kit and the black arrows represent the position of gene-specific transcripts (Taken from the Ambion FirstChoice RLM-RACE protocol manual)

B



2.4.2 3'rapid amplification of cDNA ends (3'RACE)

The 3'RACE assay was carried out using the Ambion FirstChoice RLM-RACE kit (Invitrogen AM17000) as per manufacturer's instructions and using reagents provided in the kit unless otherwise specified. Primers used were either provided in the kit or uniquely designed to be gene-specific, they can be found listed in section 2.10.4 under 3'RACE primers.

1 µg of total RNA was used in a reverse transcription reaction which involved the binding of an anchored oligo(dT) 3'RACE primer to poly(A) containing RNA transcripts. Anchored oligo(dT) primers require the annealing of the oligo(dT) to the Poly(A) tail and the anchored nucleotide ("NV" see fig 2.2) to the 5'end of the 3'poly(A) tail. MLV-Reverse Transcriptase was used for this reaction which took place at 42°C for 1hr. The sample was then ready to undergo nested PCR, first with the 3'RACE gene-specific outer primer followed by the 3'RACE gene-specific inner primer. Both PCR cycles followed with an initial denaturing step at 94oC for 3 mins, then 35 repeats of 94oC for 30 sec, 60oC for 30 sec, 72oC for 30 sec, and a final extension of 72oC for 7 mins. 3'RACE PCR products were visualised on a 5% PAGE gel stained with ethidium bromide and all bands were excised for sequencing in accordance with the sequencing protocol (Ch 2.3.4). See Figure 2.2 for an overview of the 3'RACE procedure and the position of primers used in the 3'RACE PCR.

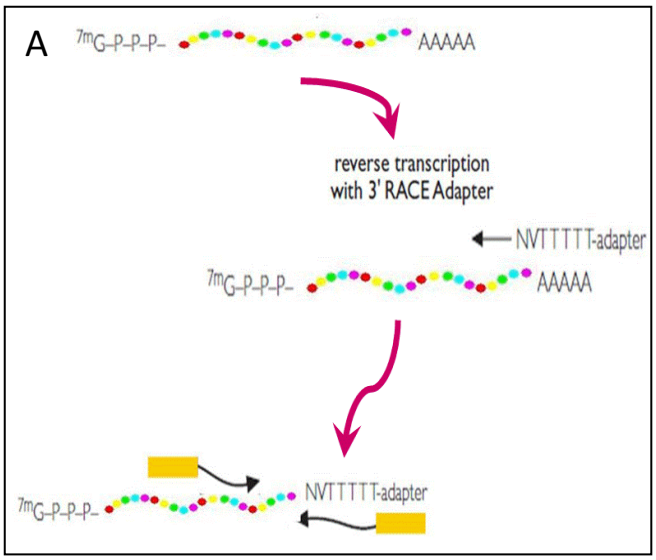
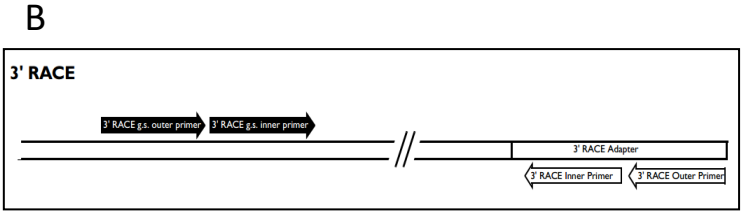


Figure 2.2 Overview of the 3'RACE protocol and 3'RACE amplification of transcripts. A) Diagram of the overview of 3'RACE procedure described in ch 2.4 B) Diagram of primer positions for 3'RACE, white arrows denote the position of primers supplied by the kit and the black arrows represent the position of gene-specific transcripts (Adapted from the Ambion FirstChoice RLM-RACE protocol manual)



2.5 NORTHERN HYBRIDISATION

2.5.1 Labelling of DNA probes^{32}P γATP

DNA oligonucleotide probes were designed to be complementary to the target miRNAs. 50 µM of oligonucleotide probe was labelled by mixing with 1 µl of polynucleotide kinase (PNK) buffer, 1 µl of PNK enzyme, 1 µl of ^{32}P γATP and 6.5 µl of water. The sample was incubated at 37^{o}C for 1 hr. To remove excess^{32}P γATP the samples were precipitated with ethanol by adding 70 µl of Tris-EDTA pH 8.0 (TE) buffer, 20 µl of ammonium acetate (10 M) and 1 µl of glycogen to the sample. The sample was mixed and then 250 µl of -20^{o}C ethanol was added and the sample was left on ice for 1 hr. The sample was centrifuged at max speed at 4^{o}C for 20 mins. The supernatant was removed and the sample was left to air dry briefly before being resuspended in 50 µl TE. The radioactivity of the sample was checked and compared with the supernatant with a Geiger counter, and stored at -80^{o}C.

2.5.2 Northern blot gel

10 – 30 µg of total RNA was denatured at 95^{o}C for 5 mins and separated on a 15% polyacrylamide denaturing gel with 7M Urea in 0.5 x TBE buffer at 100-250V. The gel was washed in 0.5 x TBE to remove excess urea. 6 pieces of Whatman filter paper, and 1 piece of Hybond N+ nylon membrane was cut to the same size as the gel and soaked in 0.5 x TBE. 3 of the pieces of filter paper were then placed on the semi-dry blot apparatus and the membrane was layered on

top. The gel was positioned on top of the membrane before the remaining 3 pieces of filter paper were layered on the gel. Bubbles within the "gel sandwich" were removed by rolling a long pipette over the top, and any excess liquid surrounding the sandwich was wiped away. The semi-dry apparatus was run at 3.3 mA/cm^2 of the gel size for 35 mins (~5 V). The membrane was washed in 0.5 x TBE for 5 mins, and then placed on top of a piece of filter paper and UV cross-linked at 1200 µJoules, twice.

2.5.3 Hybridisation

The membrane was washed with 2 x SSC + 0.1% SDS for 5 mins at room temperature. Hybridisation buffer consisting of 3 ml of 20 x SSC, 1 ml of 50 x Denhardt's solution, 0.25 ml of 10% SDS, and 5.75 ml water, was pre-warmed to 42^{o}C. The membrane was placed inside a 15 ml falcon tube with the transferred RNA side facing upwards, and pre-hybridised with 3 mls of the hybridisation buffer, rotating at 42^{o}C for 30 mins. The solution was removed and fresh hybridisation buffer was added to the tube along with 100 µl of the ^{32}P γATP labelled oligonucleotide probe. The probe was hybridised to the membrane for at least 12 hrs at 42^{o}C, and then washed twice in 2 x SSC + 0.1% SDS at room temperature. The membrane was exposed at -80^{o}C with an intensifying screen for a minimum of 3 days.

2.5.4 Northern blots with double-DIG labelled LNA probes

Pre-designed miRCURY LNA microRNA Detection Probes made by Exiqon with 3' and 5' Digoxigenin (DIG) labels were used instead of the radioactive DNA probes.

The denaturing PAGE and semi-dry blot follows the same procedure however the hybridisation, washing and detection steps differ.

2.5.5 Hybridisation with LNA probes

Hybridisation buffer consisting of 9 ml of 20 x SSC, 3 ml of 50 x Denhardt's solution, 0.75 ml of 10% SDS, and made up to 30 ml with water, was pre-warmed to 42^{o}C. The membrane was placed inside a 250 ml cylindrical container with the transferred RNA side facing upwards, and pre-hybridised with 15 mls of the hybridisation buffer, rotating at 42^{o}C for 30 mins. The solution was removed and fresh hybridisation buffer was added to the tube along with 2 nM of denatured mercury double-DIG labelled LNA microRNA Detection Probes. The probe was hybridised to the membrane for at least 12 hrs at 42^{o}C.

2.5.6 Washing and detection

The membrane was washed twice with a low stringent buffer (2 x SSC and 0.1% SDS in water) at 50^{o}C and then twice with a high stringent buffer (0.1 x SSC and 0.1% SDS in water) 50^{o}C. It was then briefly rinsed twice in 1 x washing buffer (Roche) at 50^{o}C. The membrane was incubated for 3hrs at room temperature in blocking buffer containing maleic acid buffer and 1% BSA (Roche). The blocking buffer is replaced for 30 minutes with DIG antibody solution which is a dilution of the DIG antibody 1 in 10000 in blocking buffer (Roche). The membrane is rinsed again in 1x washing buffer (Roche) four times after which, the membrane is incubated for 5 minutes in 1x detection buffer (Roche). The substrate solution containing a 1 in 100 dilution of CSPD (Roche) in detection buffer is then added

to cover the membrane. This is left to incubate at room temp for 5 minutes before the membrane is sealed and incubated in the dark at 37^{o}C for 15 minutes. The membrane is then exposed at -80^{o}C for 5 mins. The exposure time can be adjusted according the strength of the signal and the substrate last for up to 48 hrs.

2.5.7 Northern blotting of primary miRNAs by agarose-formaldehyde gel

Prior to making the gel the 10 x running buffer was prepared with the following materials; 0.2 M MOPS pH 7, 20 mM Sodium acetate, 10 mM EDTA (pH 8) and RNAse free DEPC-treated water. Following this the 1% agarose formaldehyde gel was cast by melting 1 g agarose in 72 ml RNAse free water. In the fume hood, 10 mls of 10 x running buffer and 18 ml of 37% formaldehyde was added and mixed. Subsequently the horizontal gel cast and comb was assembled and the molten 1% agarose-formadehyde gel was poured and left to set.

Sample preparation

0.1 µg/ml of ethidium bromide was added to 30 µg of RNA and 1:1 denaturing dye (10 ml deionized formamide, 200 µl 0.5 M EDTA pH 8.0, 1 mg xylene cyanol FF, 1 mg bromophenol blue) which was then denatured at 65 ^{o}C for 10 minutes. The sample was then loaded into the 1% agarose-formaldehyde gel in 1 x running buffer for 2-3 hours at 100 Vs. The gel is then visualised by UV to check the integrity of the RNA.

Capillary transfer arrangement

Before transferring, the gel is washed DEPC-water 2 x 5mins and 10 x SSC for 2 x 5 mins. During this time 4 x Whatman filter paper, and 1 piece of Hybond N+ nylon membrane is soaked in 10 x SSC and arranged as described in figure 2.3. The transfer takes place overnight (or two nights) and is topped up with 10x SSC transfer buffer as necessary. After transfer the membrane is washed in DEPC-water and UV crosslinked to fix the RNA.

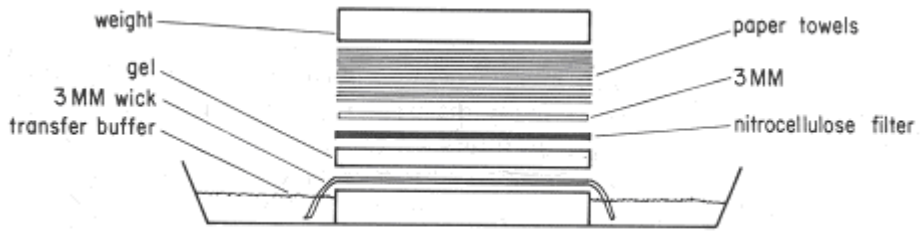


Figure 2.3 Capillary transfer arrangement. From bottom to top: Within a deep dish filled with 10x SSC transfer buffer rests within a platform with 2x whatman filter paper soaked in 10x SSC. On top of this lies the gel with the edges covered in saran wrap to prevent any contact with the buffer disrupting the current. Then Hybond N+ nylon membrane is placed directly onto the gel followed by 2 x whatman filter paper (soaked in 10 x SSC), a stack of paper towels and a 500g weight(picture taken from Molecular Cloning, Maniatis).

Probe labeling with radioactive dCTP

The probe was generated by RT-PCR of genomic DNA using gene specific primers to generate a complementary probe of ~200bs. Probe labelling is performed with radioactively labelled dCTP (Perkin Elemer BLU013H100UC) and the Fermentas DecaLabel kit (K0622) following manufacturer's instructions. Briefly, 25 ng of PCR product is combined with 10 µl buffer and 40 µl water which is then heated in boiling water for 10 mins and snap cooled on ice to denature the DNA. Following radioactive guidelines the sample is then combined with 3 µl MixC (dNTPs without dCTP), 5 µl radioactively labelled dCTP and 1 µl Klenow enzyme.

After incubation for 5 mins at 37°C the labelled probe is precipitated, washed with 80% ethanol and resuspended in 50 µl water. The radioactivity of the sample was checked with a Geiger counter, and the probe was used straight away or stored at -80°C.

Hybridisation

Hybridisation buffer consisting of 15 mls of 20 x SSC, 5 mls of 50 x Denhardt's solution, 2.5 mls of 10% SDS, and made up to 50 ml with water, was pre-warmed to 42°C and then 10 mls was used for pre-hybridisation of the membrane rotating at 42°C for 30 mins. The pre-hybridisation buffer is discarded and replaced with the hybridisation buffer containing 50 µl of probe and left to hybridise at 42°C overnight. The membrane was then washed 4 x SSC + 0.1% SDS at 42°C. The membranes were incubated with x-ray film and exposed at -80°C with an intensifying screen for a minimum of 5 days.

2.6 ISOLATING NASCENT RNA TRANSCRIPTS

2.6.1 Thiouridine labelling

HepG2 cells and mouse NIH 3T3 fibroblasts were cultured in 10cm cell culture dishes (Corning) until 80% confluent. Final concentrations of 200 µM Thiouridine (4sU) (Sigma T4509) was added to cells for 2-3 hours to achieve labelling of nascent RNAs.

2.6.2 Total RNA extraction

To end 4sU labelling, the cell culture medium was removed from cells and 5 mls of Trizol was added immediately to cells. This was incubated at room temperature for 5 minutes to allow nucleoprotein complexes to dissolve and then the lysate was transferred into 15 ml falcons. 1 ml of chloroform was added to 5 mls of trizol and shaken vigorously for 30 secs. After incubation at RT for 5 minutes the falcons were then centrifuged at 13,000 g for 15 mins at 4^{o}C. The aqueous upper phase was transferred to a new 15ml falcon tube with half the reaction volume of isopropanol. After mixing and incubating at RT for 10 mins the falcon tubes were centrifuged at 13,000 g for 10 mins at 4^{o}C. The supernatant was removed immediately and an equal volume of 80% ethanol was added and shaken to detach the pellet. Following this the falcon was centrifuged at 13,000 g for 10 mins at 4^{o}C. Again the supernatant was discarded and to ensure all remaining ethanol was removed the falcon was centrifuged again and remaining ethanol was removed by pipetting using a 200 µl and a 20 µl pipette. The RNA pellet is then resuspended in 100 µl of 1x TE (10 mM Tris and 1 mM EDTA) and dissolved on ice. The RNA was measured on the nanodrop and run on a 1% agarose gel to visualise the integrity of the RNA.

2.6.3 Biotinylation of 4sU labelled RNA transcripts

In 2 ml eppendorfs, 100 ug of 4sU labeled total RNA (from chapter 2.6.4) was added to 200 µg of Biotin-HPDP dissolved in Dimethylformamide (DMF) (Pierce, EZ-Link Biotin-HPDP) (final concentration 100 ng/ul); 100 µl of 10x

Biotinylation buffer (100 mM Tris pH 7.4 and 10 mM EDTA) and 700 µl of DEPC-treated water. This was incubated at RT for 1.5 h with rotation.

To isolate the biotinylated RNA an equal volume of Chloroform was added, mixed vigorously and incubated for 5 minutes at RT until the phases begin to separate. This was then centrifuged at full speed for 5 minutes. The upper phase was carefully transferred to phase lock gel heavy tubes (Eppendorf) (these tubes help avoid loss of material) and the chloroform step was repeated to ensure removal of unincorporated biotin-HPDP. The RNA was then precipitated in 1/10 of volume of 5 M NaCl and an equal volume of isopropanol. This was centrifuged at max speed for 20 mins at 4^{o}C, the supernatant was then removed and the pellet was washed with 80% ethanol and precipitated again at max speed for 10 mins at 4^{o}C. After removal of ethanol as described in chapter 2.6.2 the biotinylated RNA was resuspended in 100 µl (1 x TE).

Analysis of biotinylated RNA was performed by dot blot, this involved placing 1 µl drops of biotinylated RNA dilutions (5 10-fold dilutions were used 1 µg – 1 ng) and a positive (biotinylated oligo) and negative (unlabelled RNA) control on Hybond N+ nylon membrane. The RNA was then UV cross-linked to the membrane. The membrane was incubated in 40 ml blocking solution (20 ml 20% SDS, 20 ml 1 x PBS and 1 mM EDTA) with shaking. The blocking solution was replaced with streptavidin-horseradish peroxidase solution (10 ml 1 x PBS, 10 ml 20% SDS and 20 µl Streptavidin-HRP (Millipore Cat 18-152) and incubated with shaking for 15 mins. The solution was then removed and the membrane was washed 2 x 10 mins with 40 mls 1:1 PBS and 20% SDS followed by 2 x 10

mins with 40 mls 1% SDS in PBS and finally 2 x 10 mins with 40 mls 0.1% SDS in PBS. The membrane was then dried lightly and covered with 2-5 mls ECL reagent for 2 minutes. The membrane was encased in a clear plastic wallet and excess ECL was removed. The membrane was exposed to an x-ray film for 1 – 5 mins.

2.6.4 Biotin affinity purification of 4sU labelled RNA (nascent RNA)

100 µg of biotinylated RNA was captured using 100 µl of µMACS streptavidin beads and columns (MIltenyi). The RNA-bead mixture was incubated at RT with rotation and then the beads were transferred and magnetically fixed to columns. The column was washed three times with 900 µl 65^{o}C washing buffer (100 nM Tri-HCL, pH 7.4, 10 mM EDTA, 1 M NaCl and 0.1% Tween20) followed by three washes with RT washing buffer. The use of the columns eliminated the risk of purifying unlabelled RNA. Labelled RNA was eluted by adding 100 µl of freshly prepared 100mM dithiothreitol (DTT) followed by another elution with DTT 5 minutes later. Recovery of nascent RNAs was achieved using the RNAeasy MinElute Cleanup Protocol following the manufacturer's instructions (Qiagen).

2.7 IMMUNOCYTOCHEMISTRY OF NASCENT RNAS

Equal densities of HepG2 cells (and ESCs) were plated onto 13 mm^2 glass coverslips (VWR) placed in a 24-well plate and left to adhere overnight. The following day cells were left untreated or pre-treated with 1 µM actinomycin D, 200 ng/ml SSA or 2 ng/ml TG003 for 1 hour at 37 ^{o}C in the dark. Subsequently, cells (where indicated) were treated with 1mM 5-fluorouridine (5-F) for 1 hour at 37 ^{o}C in the dark. After treatment, cell staining was performed in low light

levels and at RT. Initially the cells were fixed with 10% formalin (w/v) for 15 minutes and then washed in PBS supplemented with 0.1% BSA and 0.2% Tween20 (PBS-BT). Then the cells were permeabilised with 0.1 Triton-X/PBS for 5 minutes. Cells were washed again with PBS-BT and then 10% FCS in PBS was used for blocking for 30 minutes. The block was then removed and replaced with 1:500 anti-BrdU antibody (in 10% FCS) for 30 minutes. After washing cells three times with PBS-BT, cells were incubated with 1:100 anti-mouse alexa fluor (in 10% FCS) conjugated antibody for 1 hour. Lastly cells were washed twice in PBS-BT and once in distilled water and mounted on glass slides using Prolong antifade with DAPI (Invitrogen). Photographs were taken on an SP5 confocal microscope and LAS AF Lite, Adobe Photoshop and Image J software was used to analyse the data.

2.8 ISOLATION OF NUCLEAR AND CYTOPLASMIC RNAS

Cells were grown in a 6 well plate, at 80% confluency cells were trypsinised in 1 ml, washed with PBS and pelleted. The cell pellet was resuspended in 120 µl LB buffer (10 mM NaCl, 2 mM MgCl, 10 mM Tris-HCL pH 7.8, 5 mM DTT, 0.5% Nonidet-P40) for cell lysis, placed on ice for 5 minutes and centrifuged for 5 mins, 8000 rpm at 4^{o}C . The supernatant containing the cytoplasmic RNA was transferred to a new eppendorf with 120 µl Proteinase K buffer (0.2 M Tris pH 7.5, 25 mM EDTA, 0.3 M NaCl, 2% SDS). The remaining nuclear RNA pellet was resuspended in 120 µl LB buffer and 120 µl Proteinase K buffer was added. The following treatments were performed on both the cytoplasmic and nuclear RNA fractions. The samples were then incubated at 37^{o}C for 20 mins. The RNA was

then extracted with 700 µl Trizol and 140 µl Chloroform followed by vigorous shaking and centrifugation for 5 mins at 13,000 g. The supernatant was transferred to a new eppendorf with 15 µg glycogen and 1 ml 100% ethanol. After thorough mixing, the sample was centrifuged for 20 mins at 13,000 g at RT. The ethanol was removed thoroughly and the pellet was dried before resuspending in 400 µl DEPC-water. The RNA sample was precipitated again with 1/10 vol sodium acetate and 1 ml 100% ethanol. The pellet was washed with 80% ethanol, dried thoroughly and resuspended in 20 µl DEPC-water.

2.9 LENTIVIRUS PRODUCTION AND INFECTION

One million HepG2 cells were seeded into a 10 cm plate the day before transfection. The next day, a plasmid DNA mix of 15 µg lentivector (figure 2.4), 10 µg pCMVΔ8.91 packaging construct and 5 µg VSV-g envelope construct, in the ratio of 3:2:1 (Lentivector: Packaging: Envelope), sterile water and 50µl calcium phosphate were combined to a total volume of 500 µl. This mix was added drop-wise to 500 µl of 2x HBS while mixing rigorously then the solution was incubated at room temperature for 30 minutes. Fresh media was supplied to the HepG2 cells with 25 µM chloroquine (Sigma) and then the plasmid / calcium phosphate mix was added drop-wise. The cells were incubated overnight at 37^{o}C with 5% CO_2. 18 hours post-transfection, the cells were washed twice with D10 and replaced with 8 mls fresh media. The media containing lentivirus was harvested continuously for 2 days, concentrated and stored at -80^{o}C.

50,000 cells were seeded in a 24-well plate one day prior to infection. The next day, 4 µg/ml polybrene (Sigma) was added to the media followed by 10-20 µl of lentiviral particle. The cells were incubated overnight and replaced with fresh media the following day. Total RNA extraction using Trizol was performed after 48 or 72 hours.

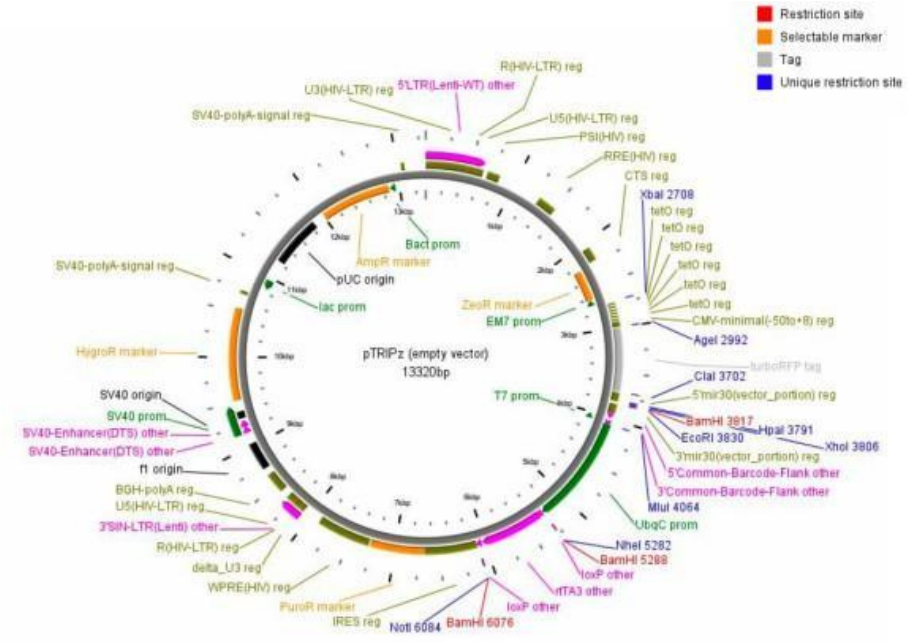


Figure 2.4 pTRIPz inducible lentiviral vector used in the cloning of the three miR-let-7a/miR-let-7f transcripts (see figure 4.2.4) expression vector. pTRIPz inducible lentiviral vector was used in the cloning for construction of let 7a and let 7f expression vector (vector map reproduced from Thermo Scientific Open Biosystems technical manual).

2.10 REAGENTS AND PRIMERS

2.10.1 PCR, Northern blotting and Cloning reagents

15% denaturing PAGE: 21 g urea, 2.5 ml 10 x TBE, 18.75 ml of 40% (w/v) 19:1 acrylamide:bis-acrylamide, Adjust volume to 50 ml with water. Add 350 µl of 10% (w/v) ammonium persulphate (APS) and 17.5 µl of TEMED

5% non-denaturing PAGE: 2.5 ml 10 x TBE, 6.75 ml of 40% (w/v) 19:1 acrylamide:bis-acrylamide, Adjust volume to 50 ml with water. Add 350 µl of 10% (w/v) ammonium persulphate (APS) and 17.5 µl of TEMED

1% Agarose gel: 1 g Agarose, 10 ml 10 x TBE, 90 ml water. Add 0.1 µg/ml of ethidium bromide.

Denaturing loading dye: 10 ml deionized formamide, 200 µl 0.5 M EDTA pH 8.0, 1 mg xylene cyanol FF, 1 mg bromophenol blue,

Non-denaturing loading dye: 0.02% w/v 1 M EDTA pH 8.0, 0.25% w/v xylene cyanol FF, 0.25% w/v bromophenol blue, 15% Ficoll in water.

2.10.2 Other commonly used reagents

10 x PBS: 0.5 M NaH2P04 (30 g/500 ml), 0.5 M Na2HP04 (35.5 g/500 ml)

pH 7.2 for stock solution. 200 ml of stock, add 8.76 g NaCl and make up to 1 litre with distilled water.

10 x TBE: 432 g Tris, 220 g Boric acid, 37.2 g EDTA, Water to 4 litres

20 x SSC 175.3 g NaCl, 88.2 g Sodium citrate, water to 1 litre pH 7

TE Buffer: 10 mM Tris-HCL pH 7.4, 1 mM EDTA

50x DEN buffer: 2ml 50% Dextran (10 g Dextran and 15.4 ml distilled water), 5 ml formamide, 1 ml 10% SDS, 2 ml 5M NaCl

DNA elution buffer: 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% w/v SDS

YT medium: 16 g Tryptone, 10 g Bacto-yeast extract 5 g NaCl, Water to 1 litre

2.10.3 Conditions for commonly used chemicals and drugs

Abbreviations: Dimethylsulfoxide (DMSO), Dimethylformamide (DMF), Phosphate buffer saline (PBS)

Chemical	Diluents	[Stock]	Conditions of storage
Cycloheximide (Cyclo)	100% Ethanol	500 µg/ml	minus 20°C
Actinomycin D (Act-D)	DMSO	500 µM	minus 20°C, in dark
5-flurouridine (5-FU)	Media	N/A	Made fresh
Spliceostatin A (SSA)	Methanol	0.1mg/ml	minus 20°C, in dark
TG003 (Cdc2-like kinase inhibitor)	DMSO	1mg/ml	minus 20°C, in dark
Thriouridine (4sU)	PBS	50mM	minus 80°C, in dark
Biotin-HPDP	DMF	1mg/ml	minus 20°C, in dark

2.10.4 MicroRNA PCR primers

Primer	FWD	REV
miR-302a	GCTCCCCACCACTTAAACG	TCACCAAAACATGGAAGCACTTACTT
miR-302b	GTGGGCTCCCTTCAACTTTA	CACTCCTACTAAAACATGGAAGCA
miR-367	GTTCTTGGCTACAGGCCATT	TTGGCTTAACAATCCATCACC
miR-7	TTGGCCTAGTTCTGTGTGGA	GCTGCATTTTACAGCACCAA
miR-100	GTTGCCACAAACCCGTAGAT	AGGTCCGTGAGATTGCCTAA
miR-125b-1	ATTGTTGCGCTCCTCTCAGT	AGAGCCTAACCCGTGGATTT
beta actin	AGAAAATCTGGCACCACACC	AGAGGCGTACAGGGATAGCA
miR-182	GCCTGGACCATCCTAACTGT	GAGAACAGCAGGTCCAGCAT
miR-183	CGCAGAGTGTGACTCCTGTT	TTCCGACGTCTAGGAGCACT
miR-96	TCCTAGACGTCGGAAACAGG	ATTGCTCAGAGCGGAGAGAC
miR-Let-7f-2	TGTGGGATGAGGTAGTAGATTG	CGTGGGAAAGACAGTAGACTG
miR-9-1	CGGGGTTGGTTGTTATCTTTG	GGGGTTATTTTTACTTTCGGTTATC
miR-9-2	TGTGGGAAGCGAGTTGTTATC	CGGCGATCTTGAAGGAGTT
miR-146	CCAGGCTGAAAGAACTTTGG	GCACCAGAACTGAGTCCACA
LARP7	TGGAAACTGAAAGTGGAAATCA	GAAGATTTGCATCCCCAAAC
miR-302c	TCCCCTTTGCTTTAACATGG	TCCACTGAAACATGGAAGCA
miR-302d	TAACATGGAGGCACTTGCTG	CCACCACACTCAAACATGGA
miR-Let-7a-1	AATACTATTTTTCTTATCACTCAC	CATAAACAAATGCTGCACTAC
miR-Let-7f-1	GTCTTATTTCTCTGTGTCAGAA	ACTGATAAACTGTAAATACAAAAT
miR-450b	GAGATGCCACGTTTTCTTGTT	GAAGATGAGTGGCTCCCAGA
miR-542	CAGATGAAAAGATGGAGGGAAT	CAAAACTATGGATGCAAAATGA
miR-29a	CGACCTTCTGTGACCCCTTA	CCCAACGGTCACCAATACAT
miR-29c	ATCAGACTTGCCACCTGGAC	CTGCTTTTCCCCCTACATCA
miR-34a	CCTCCCCACATTTCCTTCTT	CAACGTGCAGCACTTCTAGG
miR-29c	ATCAGACTTGCCACCTGGAC	CTGCTTTTCCCCCTACATCA
miR-200a	GGTTCTTCCCTGGGCTTC	GAGTAGGAGCTCCGGATGTG
miR-101-1	GGGGAGCCTTCAGAGAGAGT	AGCCACCTGTTTCACATTCC

2.10.5 3'RACE primers

Primer	Outer	Inner
Adapter Primer	GCGAGCACAGAATTAATACGACT	CGCGGATCCGAATTAATACGACTCACTATAGG
miR-302a	GTGGGCTCCCTTCAACTTTA	GCTCCCCACCACTTAAACG
miR-302b	TGCCATTTTGTTTTCTTTCTCC	GTGGGCTCCCTTCAACTTTA
miR-302c	GTGGGCTCCCTTCAACTTTA	TCCCCTTTGCTTTAACATGG
miR-302d	GTGGGCTCCCTTCAACTTTA	TAACATGGAGGCACTTGCTG
miR-367	GCTCCCCACCACTTAAACG	GTTCTTGGCTACAGGCCATT
miR-99a	TGAAGGCCTTTAATGGAGAATATG	TGGGTCTGTGTCAGTGTG
miR-Let-7f-2	CAATTGGTTGCTTCTTTATCCTG	GAGTGAGGT AGTAGATTGT
miR-100	ATGTCACAGCCCCAAAAGAG	GTTGCCACAAACCCGTAGAT
miR-96	TCCTAGACGTCGGAAACAGG	CGATTTTGGCACTAGCACATT
miR-9-1	CTCCTACTCCCGCTGACG	CGGGGTTGGTTGTTATCTTTG
miR-9-2	AATCTTAAGCGCGGCAAGTA	TGTGGGAAGCGAGTTGTTATC
miR-23a	TCTCTTTCTCCCCTCCAGGT	ATGGGATTTGCTTCCTGTCA
miR-27a	GAGGTGTCCCCAAATCTCATTAC	GTCGTGTTCACAGTGgc
miR-182	GATATGAGGGGAAGGGAGGAG	CGGTGGTTCTAGACTTGC
miR-183	GCCTTCAGGTGGAGATAGGAG	TACCGAAGGGCCATAAACT
miR-125b-1	CGAACAGAAATTGCCTGTCA	ATTGTTGCGCTCCTCTCAGT
miR-Let-7a-1	CAGGAAATGAAACCACAGCA	AATACTATTTTTCTTATCACTCAC
miR-Let-7f-1	GGGGAAACCTTTTGCTTCTT	GTCTTATTTCTCTGTGTCAGAA
miR-21	TCATTTTGTTTTGCTTGGGAGGAA	ATCCTGCCTGACTGTCTG
miR-21 (2)	ATCCTGCCTGACTGTCTG	GATGACCATTTATGCACTTTCA
miR-7-1	AAAACTGCTGCCAAAACCAC	TTGGCCTAGTTCTGTGTGGA
U6	ACTCTGTGAC AGTCCCATGG	CGCTTCGGCAGCACATATAC
miR-Let-7a-2	CCTAGCTTTC CTTGGGTCTT GC	GAGGCTTTGA GGGATTCAGT ATATC
miR-Let-7g	AAGTGGTTTTCTAAGGAAGCAGTG	CCTCTGCTGTGAGGATG
miR-214	CAAAAGTCTAACTTGTTTTTAGTTCCA	CTGGCTGGACAGAGTTGTCA
miR-199a-2	CAGCCCTTGAGTGTGTCTCA	AGGACCGGGAGAGGATTTTA
miR-181a-1	GTCTCCCATCCCCTTCAGAT	GACAGGAAACCAAAGGTGGA
miR-205	GGTCCTTGACATCTCCCAAA	CATTCCACCGGAGTCTGTCT
miR-2355	GCATTTGCTGCCTCAGTACA	GGCCCTTTCAGCTACATTCA
miR-153-1	CCTCTCCCCTGTTTTTCTCC	CACAGCTGCCAGTGTCATTT
miR-217	GCACAATAGTTTGCCATGAAA	CTAATGCATTGCCTTCAGCA
miR-4772	CATTGTTCCTTGCATTGGTG	TCAAGTGGAATTGCTTGCTG
miR-3127	GAGTTCGAGACCAGCCTGAG	TGTGGAATGGGAAGGAGAAG
miR-548f-2	AAACACGCAGGAACCTTGAG	TGGTGCGAACATAATTGCAG
miR-1301	TCCTGGTTTTGACCCAAGAG	AAACAGCCTTTCCACTGACG
miR-216a	AGGGGTACATCAGGGCTTCT	GGTCTAACACGGATGGCTGT
miR-216b	TGGCTTGGGAAACTTCACTT	CTGCAGGCAAATGTGATGTC
miR-4784	AGTCTGGGTGACAGGGTGAG	GGAGATGCTGGGACTGAGAG

2.10.6 5'RACE primers

Adapter Outer: GCTGATGGCGATGAATGAACACTG

Adapter Inner: CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG

miR-302 Cluster Outer: TTTGTTATTTTAATGTATCAGCTC

miR-302 Cluster Inner: TACAGATCTTCCTGAACTAGTTCCCAAA

2.10.7 Northern blotting probes

Primer for N.blot probe L7af fwd: CCAGGCGGTCTGATAGAAAG

Primer for N.blot probe L7af rev: GGGTAAAATCACTACCCCACAA

miR-let 7a LNA probe: AACTATACAACCTACTACCTCA

2.10.8 Primers used for checking lentivirus transfection miR-let 7a and miR-let 7f

Primer	FWD	REV
miR-Let7aV	CATAAACAAATGCTGCACTACAT	AATACTATTTTTCTTATCACT
miR-Let7fV	GTCTTATTTCTCTGTGTCAGA	ACTGATAAACTGTAAATACAAAATG
miR-Let7afV	AATACTATTTTTCTTATCACTC	TACCTCACTCTGATAGAGCAATG
let7a fwd xho - rev mlu	TACTCGAG- TCACACAGGAAACCAGGATTA	TCACGCGT - AAATTAACT GATAAACTGTAAATAC
let7a PTripZ	GCTACCTGCACTGCTCCTTC	CCAGGCGGTCTGATAGAAAG
let7f PTripZ	AACTGGGGCACAGATAATCG	TCCTCAGGGAAGGCAATAGA
let7af PTripZ	AACACCGAGATGCTGTACCC	AGTTGAGCACCACAGGAAGG

2.10.9 Primers used in Chapter 5

Primer	FWD	REV
RING4	AGCCGATACCTTCACTCGAA	TGGAAAAACTCCGTCTCCTG
TRIM21	CAATCGACAGCTAGCCAACA	TAATGCCACCTGGAGCTTCT
P48	AGGTCCAGCTGTCTGGAAGA	ATGGCATCCTCTTCCTCCTT
MXA	CCGGAAGACGAAAATGAAAA	CATTTGTGGAACTCGTGTCG
ISGI51	GAGAGGCAGCGAACTCATCT	CACCAGGATGCTCAGAGGTT
6.16	CTCGGAGAGCTCGGACAGC	CATTCAGGATCGCAGACCAG
OAS1	CTGGCTGAATTACCCATGCT	TGCAGGTCCAGTCCTCTTCT
9.27	GATCAACATCCACAGCGAGA	ATGAGGATGCCCAGAATCAG
MXA In7	GCATCGACCTCATTGACT	CTGACCTTGCCTCTCCACTT
OAS1 In1	TCCAAGCTCAGTCAGCAGAA	AGCCTGGACCTCAAACTTCA
6.16 In1	AACCGTTTACTCGCTGCTGT	CTTACCCGCATTCTCACCTG
18S Rrna	AAACGGCTACCACATCCAAG	CCTCCAATGGATCCTCGTTA
5.8S Rrna	GACTCTTAGCGGTGGATCACTC	GACGCTCAGACAGGCGTAG
EEF1A1	GCATCCTACCACCAACTCGT	TCCATTGGTGGGTCATTTTT
RPL37	AACGTCATCGTTTGGAAAGC	TTAGCCTTGGCACTCCAGTT
SERPINA1	GGGAAACTACAGCACCTGGA	CCCCATTGCTGAAGACCTTA
HMBS	AACCCTCATGATGCTGTTGTC	CAGATGCGGGAACTTTCTCT
RPS3A	TGAAAGCACCTGCTATGTTCA	GTCACGGGTAAGATCCATGC
G6PD	GGAGAATGAGAGGTGGGATG	CGGCTTCTTGGTCATCATCT
HPRT1	ACCCCACGAAGTGTTGGATA	ATTCAAATCCCTGAAGTATTCATT
Ki67 (mmu)	TTGTAGTAGGTAGAGCGCAGATTGA	CTTCTCCTTCACTGGAGTCTTGAAC
Cytokeratin 8 (mmu)	CAACAAGCGTACAGAGATGGA	CTGCAACTCACGGATCTCCT
N-Cadherin (mmu)	CCAGGAAAAGTGGCAGGTAG	CACTGTGCTTGGCAAGTTGT
Caspase 3 (mmu)	GGAGGCTGACTTCCTGTATGC	GATCTGTTTCTTTGCGTGGA
B-Catenin (mmu)	GTGCAATTCCTGAGCTGACA	ACAATGGCAGACACCATCTG
Vimentin (mmu)	ATGCTTCTCTGGCACGTCTT	AGCCACGCTTTCATACTGCT
RPL19 (mmu)	TGACCTGGATGAGAAGGATG	TTCAGCTTGTGGATGTGCTC
ATP5B (mmu)	Pre-ordered from primer design	

2.11 BIOINFORMATICS PROGRAMS

miRBase: http://microrna.sanger.ac.uk/sequences/

UCSC Genome Browser: http://genome.cse.ucsc.edu/cgi-bin/hgBlat

Ensembl Genome Browser: http://www.ensembl.org/index.html

NCBI: http://www.ncbi.nlm.nih.gov/

Pubmed: http://www.ncbi.nlm.nih.gov/pubmed/

Primer3: http://frodo.wi.mit.edu/

NEB cutter V2.0: http://tools.neb.com/NEBcutter2/

OligoAnalyzer 3.1:

http://eu.idtdna.com/analyzer/applications/oligoanalyzer/default.aspx

Alternative splicing database: http://t.caspur.it/ASPicDB/simple_search.php

Sequence editor: http://www.fr33.net/seqedit.php

Database of new exon boundaries: http://www.dbass.soton.ac.uk/dbass5/

Multiple sequence alignment: http://www.ebi.ac.uk/Tools/msa/clustalw2/

Poly(A) mining data tool: https://github.com/coggy/signals_in_sequence

Finnzymes multiple primer analyser:

https://www.finnzymes.fi/java_applets/multiple_primer_analyzer.html

miRGator v3.0: http://mirgator.kobic.re.kr/

CHAPTER 3 - EXTENSIVE POST-TRANSCRIPTIONAL REGULATION OF MICRORNAS IN ESCS

3.1 INTRODUCTION

3.1.1 MicroRNA signature in human stem cells

In collaboration with Dr Wei Cui, our laboratory made miRNA libraries from ESCs before and after differentiation into NPSCs. Solexa sequencing of these libraries identified miRNAs that were exclusively or preferentially expressed in ESCs or NPSCs (table 3.1.1). Figure 3.1.1 compares the expression of miRNAs in ESCs and their neural derivatives, NPSCs, by sequencing and northern blotting. For example, mir-302a was sequenced 2370 times in ESCs but not once in NPSCs and this result is supported by the northern blot results (figure 3.1.1). The three visible bands for miR-182 in ESCs represent isomers that differ in size by one or two nucleotides (figure 3.1.1). Our sequencing data and others show a distinct miRNA expression profile in stem cells signifying specific miRNA influence in maintaining pluripotency and influencing differentiation. Dicer-null and DGCR8 knock out ESCs do not fully downregulate pluripotency markers and show defects in differentiation demonstrating the significance of miRNA expression in ESCs (see Introduction).

MicroRNA	ESC	MicroRNA	NPSC
302a	2475	let-7f	4356
182	905	let-7a	3701
183	350	9	2169
302b	115	92b	885
31	83	29a	559
302d	75	let-7c	481
302c	68	26a	393
1323	58	let-7b	371
367	45	let-7e	326
96	9	125b	297
		24	297
		100	197
		7	150
		146b	88
		23a	53
		99a	44
		27a	43

Table 3.1.1 MicroRNA expression data taken from in ESC and NPSC sequencing libraries (100ug total RNA). Table shows sequencing reads of miRNAs that are highly and preferentially expressed in ESCs or NPSCs and not detected in the other cell type (n=0) (Dr Elcie Chan)

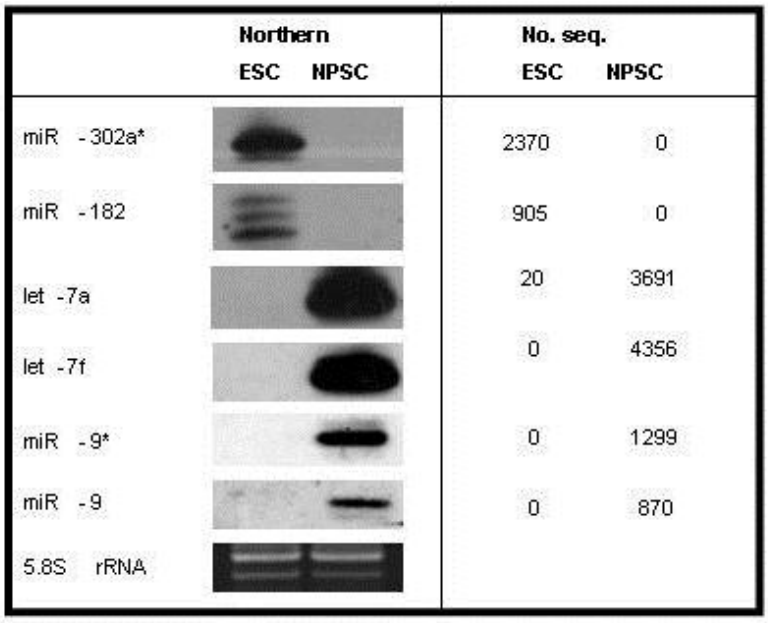


Figure 3.1.1 Northern blot and Sequencing reads for miRNAs specific to ESCs and NPSCs (Chan et al 2009 unpublished)

3.1.2 Post-transcriptional regulation of microRNAs

During early mouse development the expression of many miRNA transcripts are inhibited post-transcriptionally at the Drosha stage (Thomson et al., 2006). Given the conserved nature of miRNA transcripts across a number of species it is possible that a similar mode of miRNA regulation is present in human ESCs. Extensive regulation of miRNAs in human ESCs is to be expected given the significant impact of miRNAs in pluripotency (Thomson et al., 2006, Garzon et al., 2009).

Differential expression of miRNAs can be seen across different cell and tissue types as well as different developmental stages. Whether regulation occurs by RNA binding proteins, RNA editing, 3' modification or other miRNAs themselves,

post-transcriptional regulation is known to occur at any stage of biogenesis (figure 3.1.2) and there is reasonable evidence to suggest an abundance of post-transcriptional regulation (Siomi and Siomi, 2010, Ryazansky et al., 2011, Davis and Hata, 2009, Newman and Hammond, 2010, Thomson et al., 2006, Obernosterer et al., 2006). Differential expression of miRNAs within polycistronic clusters, discrepancy in expression of the miRNA strands and different miRNA isomer expression can be post-transcriptionally dictated by developmental timing and external conditions (Zhang et al., 2009).

Despite recognition of post-transcriptional mechanisms in miRNA regulation there have only been reports of a few miRNAs regulated post-transcriptionally in human ESCs and ESC differentiation, the most well-known example being Lin28 regulation of the let-7 miRNA family. My aim is to classify the method of regulation of miRNAs highly expressed in ESCs in an attempt to understand the mechanisms involved in regulating ECS differentiation via miRNAs.

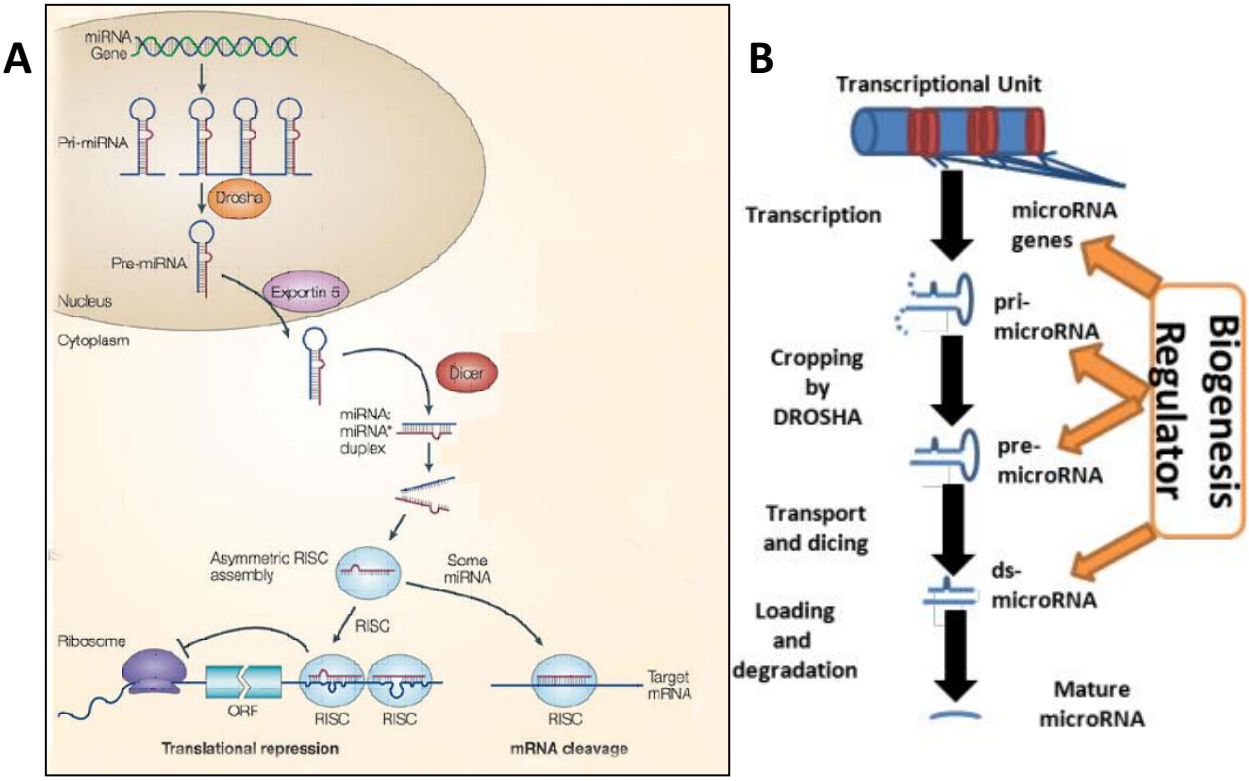


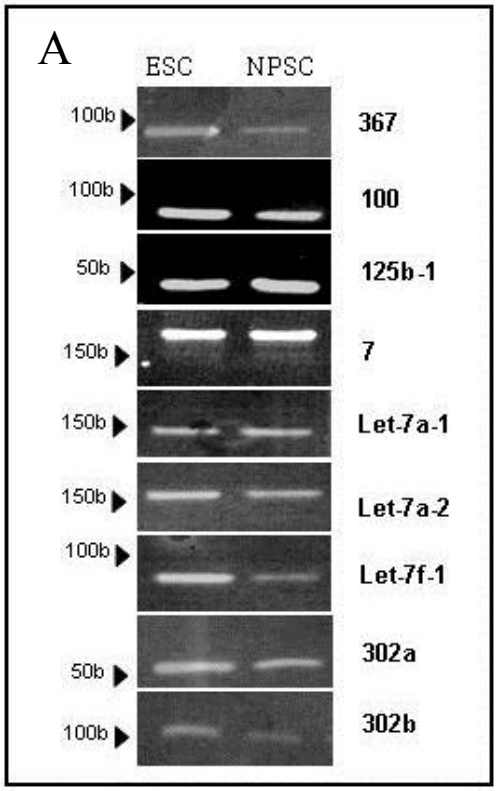
Figure 3.1.2 Many stages of miRNA biogenesis are open to post-transcriptional regulation A) Schematic diagram of the miRNA biogenesis (He and Hannon, 2004) B) Phases of miRNA biogenesis that could be subject to post-transcriptional regulation (taken from Sumazin, P http://www.genome.gov/Multimedia/Slides/TCGA1/TCGA1_Sumazin.pdf)

.

3.2 RESULTS

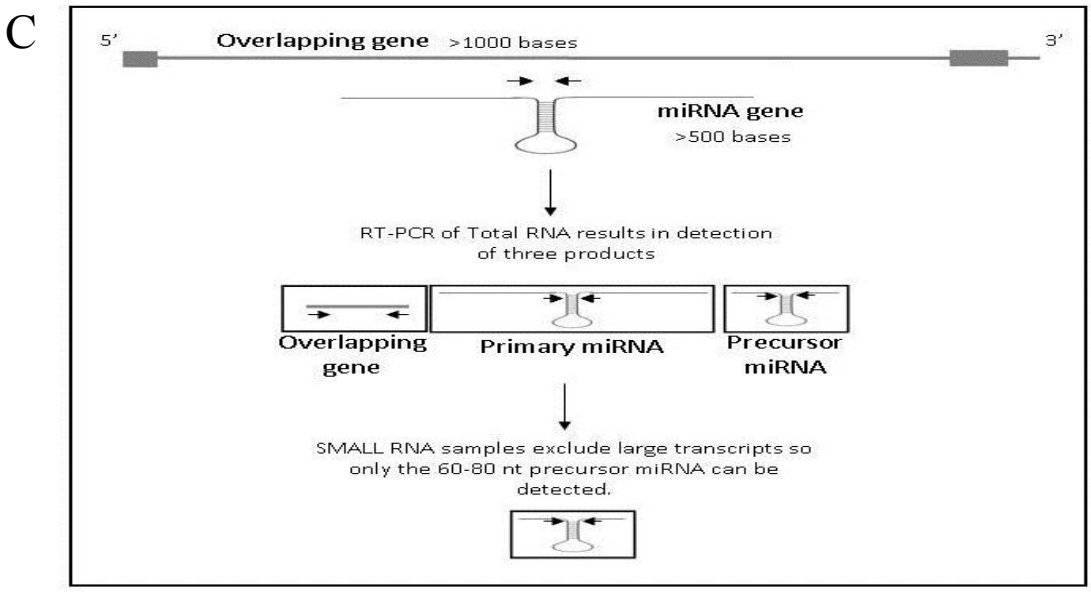
3.2.1 RT-PCR methods cannot control for detection of confounding transcripts that overlap miRNAs of interest

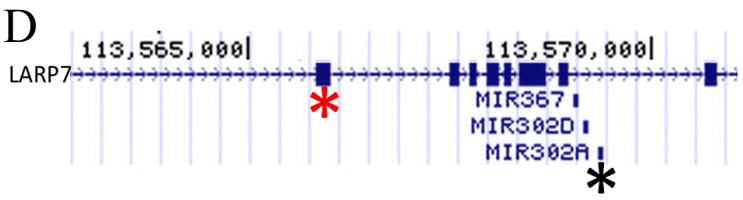
I performed RT-PCR using primers that I initially thought were specific to miRNAs of interest; these experiments resulted in the consistent detection of a large number of transcripts in both ESCs and NPSCs despite the differential expression of mature miRNAs and precursor miRNAs (Fig 3.2.1A-B, Fig 3.2.2). The implied abundance of post-transcriptional regulation of miRNAs in ESCs and NPSCs that we detected seemed unusual and did not correlate with current reports. Because the majority of human miRNAs are located within intronic regions of transcribed genes, it seemed possible that our RT-PCR approach was detecting overlapping transcripts of the miRNA genes. (figure 3.2.1C). This was confirmed in figure 3.2.1E where the overlapping gene, LARP7, can be detected in both cell lines indicating that the primer pairs for 302a also amplify the LARP7 transcript (figure 3.2.1D).



B

MicroRNA	ESC	NPSC
Let-7f	0	4356
Let-7a	20	3701
125b	0	297
100	0	197
7	0	150
367	45	0
302a	2475	0
302b	115	0





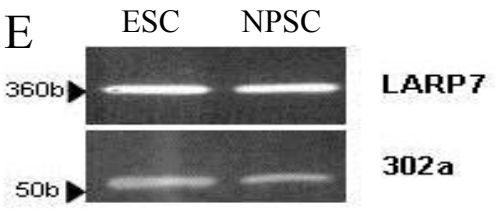


Figure 3.2.1 RT-PCR of miRNA is confounded by overlapping transcripts A) RT-PCR of total ESC and NPSC RNA show that miRNAs expression can be detected in both cell types despite one of the cell types not producing the mature miRNA. B) The table shows sequencing data illustrating the mature miRNA expression in each cell type. C) The primers, denoted by the arrows, can amplify the primary miRNA, precursor miRNA and the overlapping genes in total RNA samples D) Organisation of LARP7 overlapping miR-302, the genomic position of the primers used in RT-PCR reaction shown in fig 3.2.1E is highlighted by the red star (LARP7) and black star (302a) (See materials and methods for primer sequences) E) Shows the detection by RT-PCR of miR-302a and its overlapping gene LARP7 in both cell lines, using specific primers for these transcripts the position of which is highlighted by the red and black stars in fig3.2.1D, each star represents a pair of RT-PCR primers.

3.2.2 Problems with RT-PCR analysis of miRNA expression are confirmed when investigating precursor miRNAs

As previously mentioned RT-PCR analysis of miRNAs can lead to inadvertent detection of other gene transcripts that are not the miRNA of interest. To avoid this problem I analysed precursor miRNAs only. This was achieved by isolating small RNAs of less than 200 bases by using the mirVANA protocol. This resulted in the exclusion of overlapping genes and also primary miRNA transcripts because of their large size, so that only precursor miRNAs could be analysed (figure 3.2.2). RT-PCR of small RNA showed that pre-miRNAs could only be detected in the cells which expressed the mature miRNA (figure 3.2.2B) these bands were sequenced to confirm their miRNA identity.

Although the analysis of small RNA removed the problem of overlapping transcripts, there was a limitation with this technique in that we were also unable to analyse large primary miRNA transcripts. This is important because it is known that let-7 precursor miRNAs are expressed in ESCs, even though mature miRNA is not made (Bagga et al., 2005, Newman et al., 2008, Heo et al., 2008). However, our analysis of small RNA from ESCs gave no indication of miR-let-7 expression (Fig 3.2.2B).

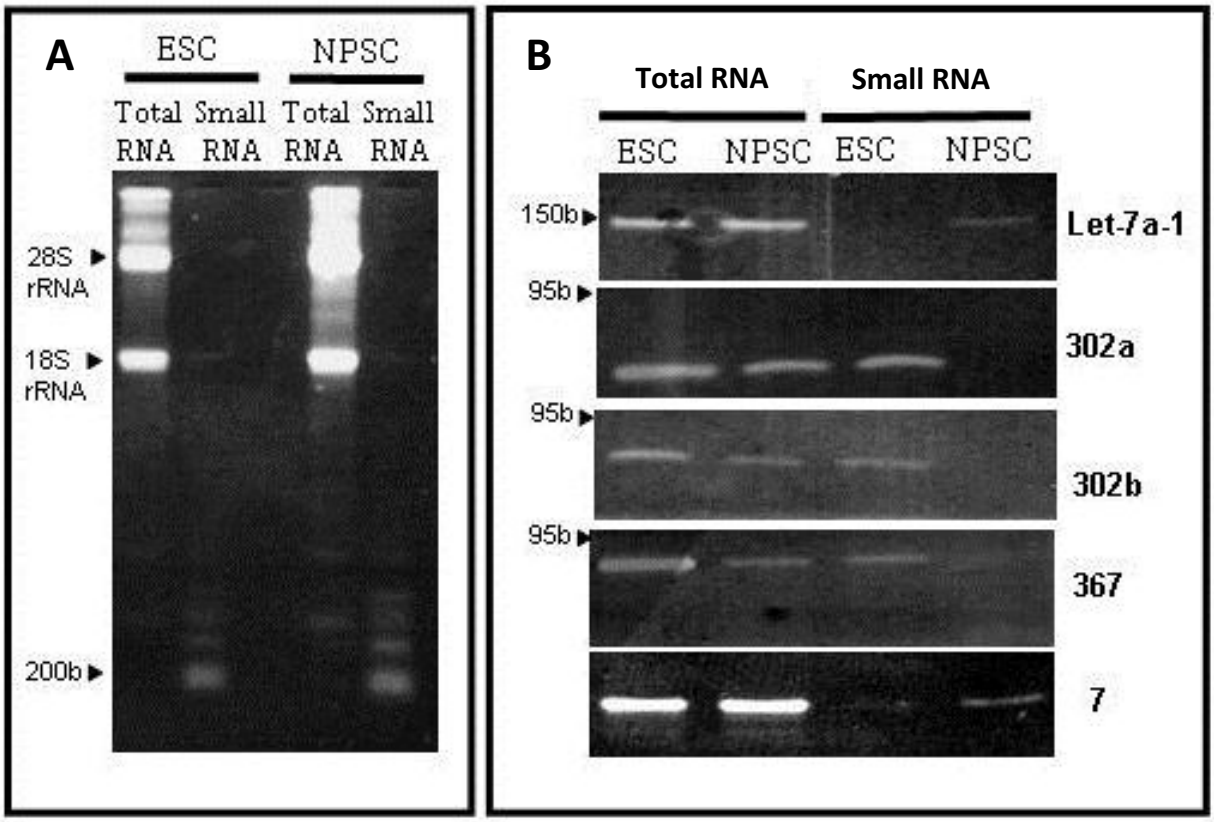


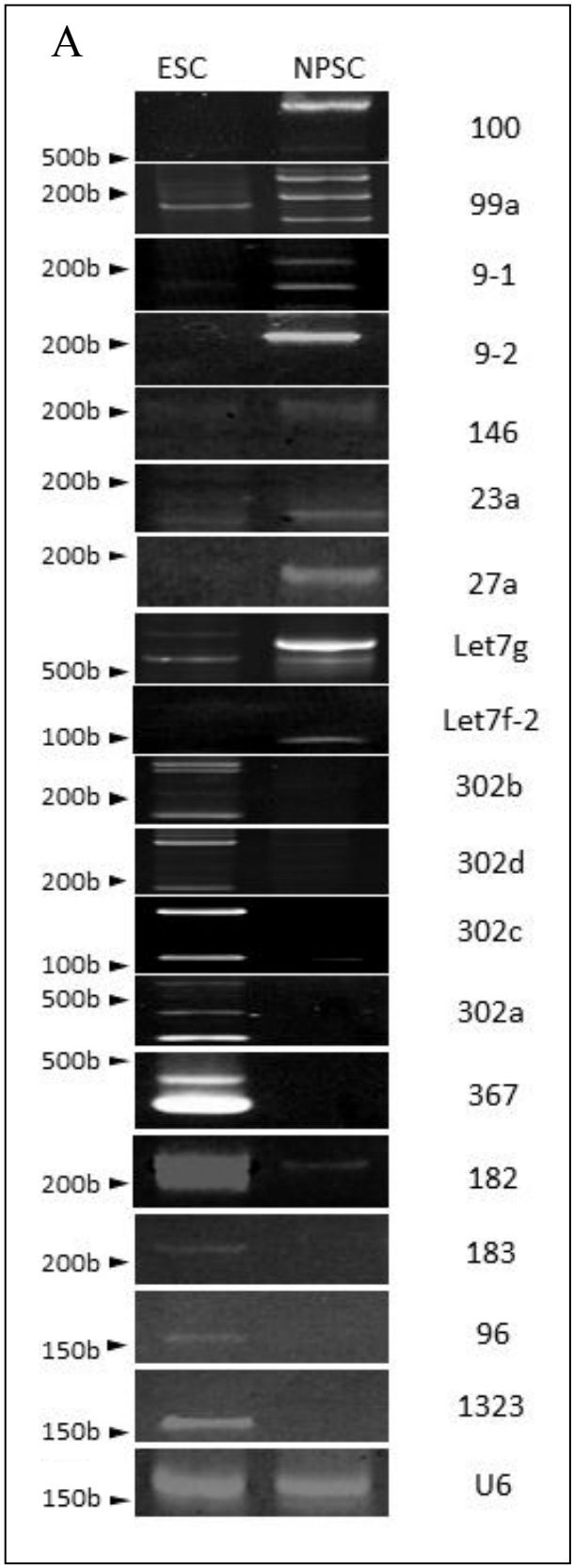
Figure 3.2.2 RT-PCR analysis of precursor microRNAs A) Small RNA extraction of ESC and NPSC RNA. Small RNAs less than 200 nucleotides in length were isolated from total ESC and NPSC RNA using the Ambion mirVana kit. Evidence of isolation can be seen by the appearance of low molecular weight RNA and the disappearance of the rRNAs on a 1% agarose gel. B) RT-PCR of cDNA generated from total RNA and small RNA from ESC and NPSCs. Total RNA and small extracted RNA below 200 bases were used for the reaction. Primers used for "Total RNA" and "Small RNA" were the same for each miRNA (see fig 3.2.1c).

3.2.3 3'RACE successfully overcomes detection of confounding overlapping transcripts allowing the identification of primary miRNAs regulated by transcription and post-transcriptionally

In order to avoid the detection of transcripts that overlap miRNAs, yet still be able to analyse primary miRNA expression, we decided to use 3' Rapid amplification of cDNA ends (3'RACE) for our analysis of miRNA expression. 3'RACE involves the binding of an anchored oligo(dT) primer sequence to the poly-A tail of RNA transcripts, subsequent cDNA synthesis and then PCR with a 3' primer specific to the oligo(dT) adapter and a 5' primer specifically binding the desired miRNA transcript. The theory behind using this method is that the overlapping transcript and the miRNA have different 3'ends and paired with the 5' primer specific to the miRNA, the two transcripts that could be potentially generated by RT-PCR would be clearly distinguishable by size. In many cases overlapping transcripts are transcribed antisense to the miRNA therefore there is no possibility of detecting the overlapping transcript using the 5' miRNA specific primer. Consequently this technique allowed successful distinction of the primary miRNAs from the overlapping transcript. In addition the technique only amplified transcripts with poly (A) tails therefore distinguishing pri-miRNAs from pre-miRNAs.

Using this technique we were able to show that 18 out of 29 primary miRNA transcripts are regulated by transcription (fig 3.2.3). This is evident by the preferential expression of the primary transcript in the cell type that expresses the mature miRNA (fig 3.2.3). This indicates that transcription is one of the main

mechanisms of controlling cell-specific miRNA expression. In figure 3.2.3 where more than one band can be seen, the sequencing results indicate that the different bands represent alternate forms of primary miRNA transcripts; this is discussed at greater length in chapter 4. In miR-99a, miR-let-7g and miR-182 the bands in the cell lines that do not express mature miRNAs were sequenced and found to be non-specific.



B

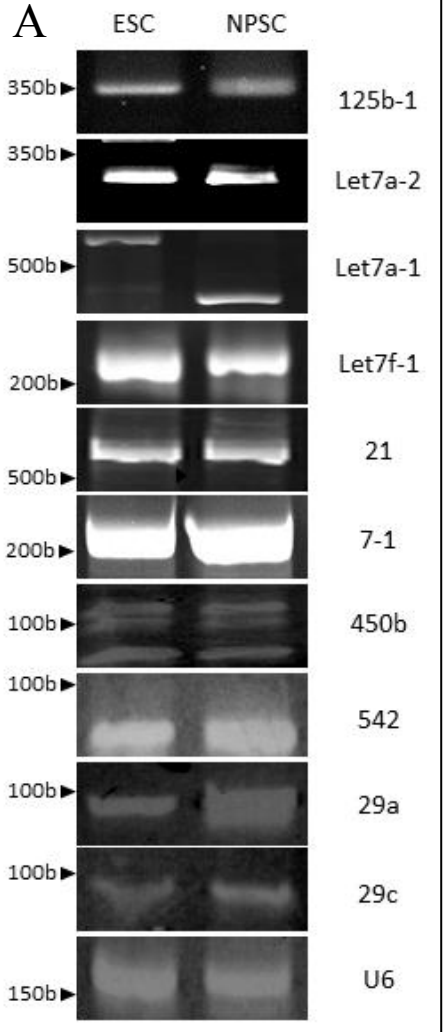
MicroRNA	ESC	MicroRNA	NPSC
302a	2475	let-7f	4356
182	905	9	2169
183	350	100	197
302b	115	146b	88
302d	75	23a	53
302c	68	99a	44
1323	58	27a	43
367	45		
96	9		

Figure 3.2.3 Transcriptional regulation of miRNAs specific to ES and NPS cells. A) 3'RACE PCR and RT-PCR detects miRNA primary transcripts that are differentially expressed in either ESCs or NPSCS where the mature miRNA is expressed. This is indicative of transcriptional regulation. The U6 gene is used as a control. In miR-99a, miR-let 7g and miR-182 the bands in the non-expressing cell line were sequenced, confirming their non-specificity. B) All miRNAs that were sequenced and preferentially expressed in either ESCs or NPSCs and not detected (n=0) in the other cell type are listed with their sequencing reads.

3.2.4 Many miRNAs in ESCs are regulated post-transcriptionally

Post-transcriptional regulation of miRNAs was inferred for 11 out of 29 miRNAs investigated (Fig 3.2.4). A pattern emerged showing a prevalence of post-transcriptionally regulated miRNA in ESCs. MiR-let-7a and miR-let-7f are known to be regulated post-transcriptionally by ESC-specific RNA binding protein Lin28. MiR-7 and miR-21 are known to be post-transcriptionally regulated in other cell types however it is not yet known how miR-125b-1, miR-542, miR-450b, miR-29a and miR-29c are regulated.

A



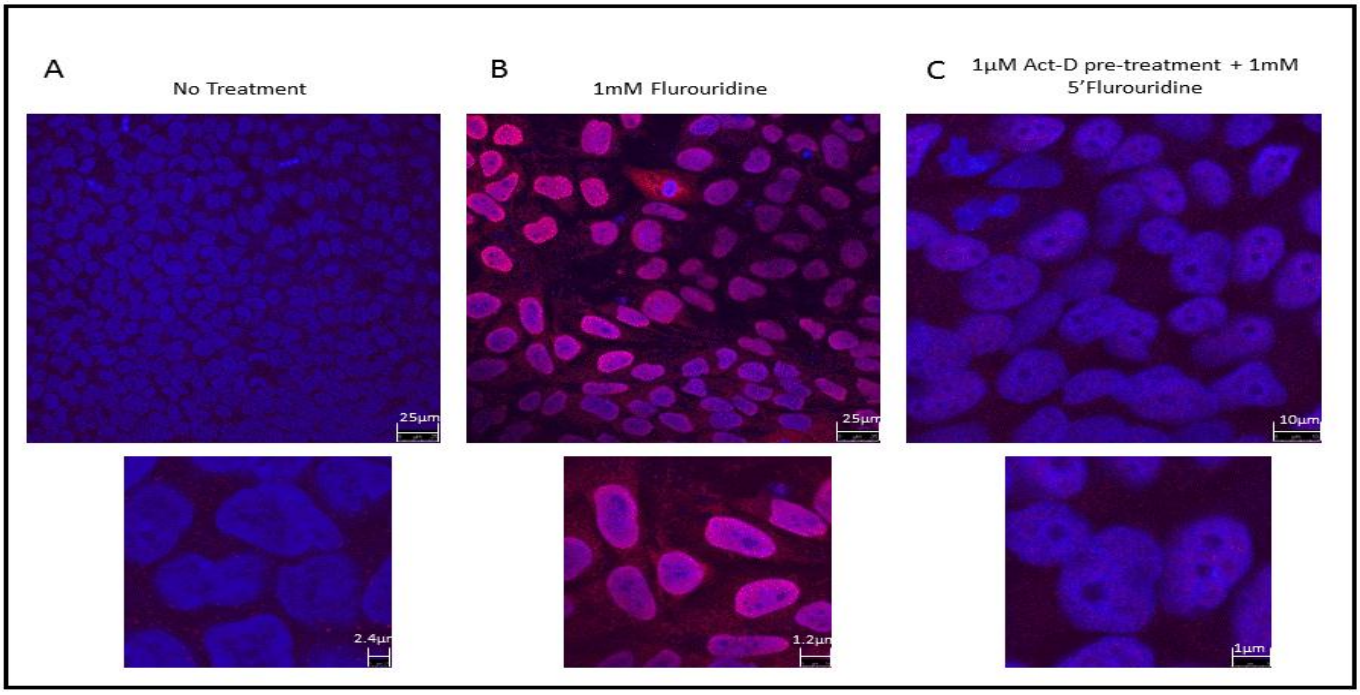
B

MicroRNA	ESC	NPSC
Let-7f	0	4356
Let-7a	20	3701
21	445	3162
29a	30	559
125b	0	305
7	0	150
542	0	16
29c	0	15
450b	0	12

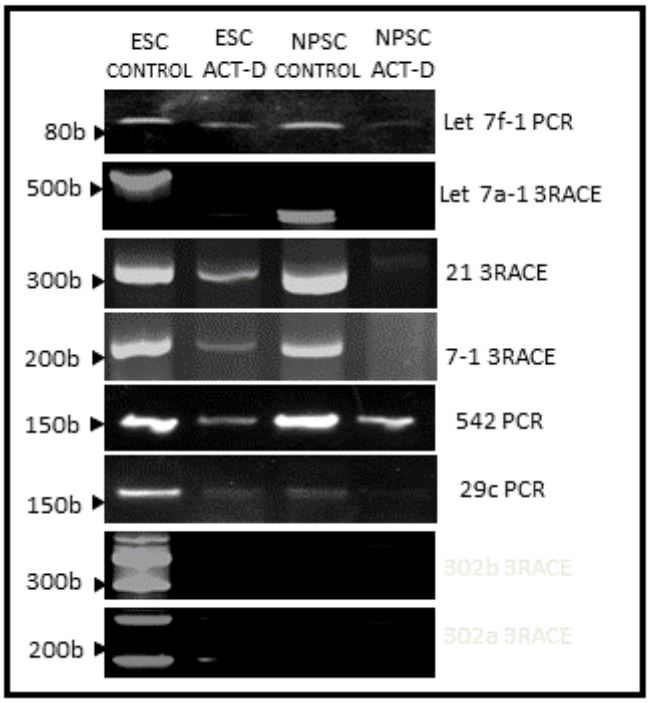
Figure 3.2.4 Post-transcriptional regulation of miRNAs in ESCs. A) 3'RACE PCR of miRNA primary transcripts that are expressed in ESCs and NPSCs despite low expression of the mature miRNAs. This is indicative of post-transcriptional regulation. B) The sequencing data showing the proportion of mature miRNAs in ESCs and NPSCs.

3.2.5 Post-transcriptionally regulated miRNA are actively transcribed and turned over

The validity of post-transcriptionally regulated primary miRNA was uncertain as it was not yet known if the primary miRNAs were being actively expressed and turned over as they would be in cells making the mature miRNA. Consequently the active transcription of some of these primary miRNAs was tested by inhibition of transcription with actinomycin D (Act-D). The efficacy of actinomycin D at inhibiting RNA synthesis was demonstrated by its blocking of 5-fluorouridine (5-FU) incorporation into nascent RNA (fig 3.2.5 A-C). The inhibition of primary miRNA expression post treatment with Act-D indicated that the transcripts were actively transcribed and turned over in ESCs despite the lack of mature miRNA expression (fig 3.2.5 D).



D

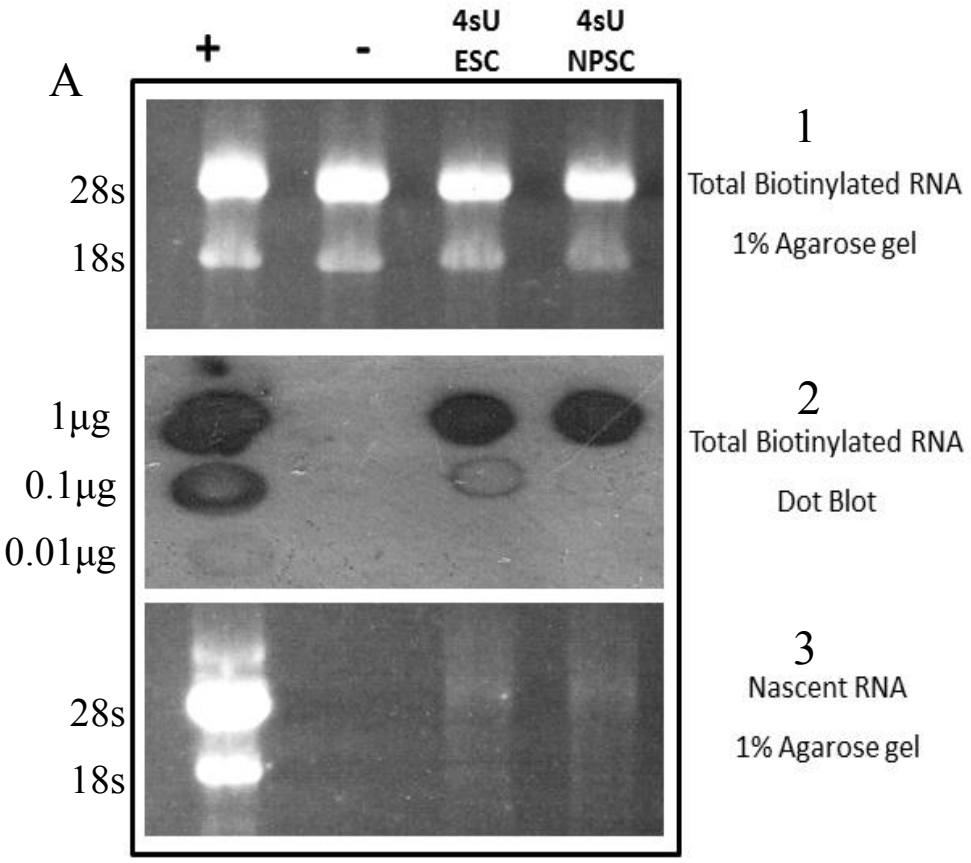


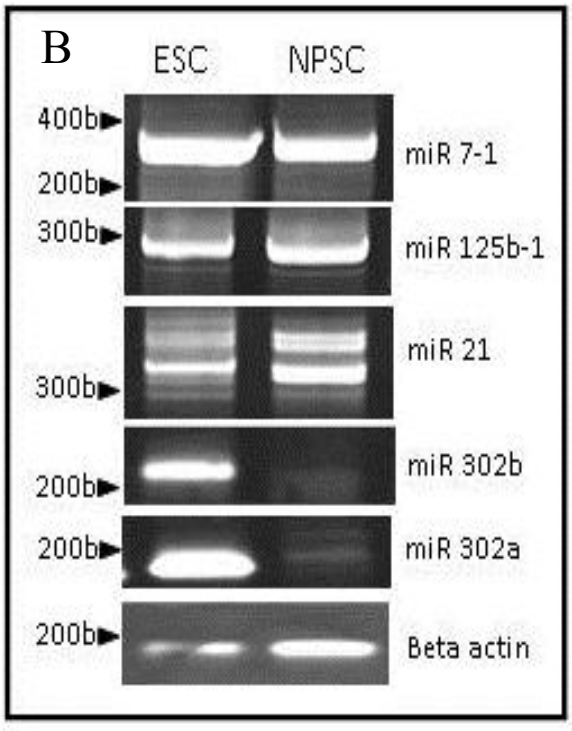
E

MicroRNA	ESC	NPSC
Let-7f	0	4356
Let-7a	20	3701
21	445	3162
7	0	150
302b	115	0
302a	2475	0
542	0	16
29c	0	15

Fig 3.2.5 RNA turnover of primary miRNA indicates that post-transcriptionally regulated miRNA are consistently being processed. A-C) Actinomycin D inhibits transcription in ESCs A) ESCs were not subjected to treatment B) ESCs treated with 1mM 5-Fluorouridine (5-F) incorporation into nascent RNA C) or pre-treated with 1 µM Actinomycin D for 30 minutes and then treated with 1mM 5-Fluorouridine. Cells were then stained by immunocytochemistry using either anti-BrdU antibody or DAPI. Photographs were taken on an SP5 confocal.D) 3'RACE PCR and RT-PCR of total ESC RNA treated with DMSO (Control) and 1µM Actinomycin D (Act-D) for 4 hours, indicates that transcripts in both ESCs and NPSCs are turned-over after 4 hours of inhibiting transcription with Actinomycin D. MiR-302b and miR-302a are transcriptionally regulated so transcription and turnover is only detected in ESCs. E) The table shows sequencing data illustrating the mature miRNA expression in each cell type.

To further verify the expression of primary transcripts, nascent ESC RNAs were biotin affinity purified to omit stable RNAs. This was achieved by incorporating a uridine base analogue, thiouridine (4sU), into growing cells. Subsequent biotinylation of newly synthesised RNAs allowed the biotin affinity purification of the nascent RNAs with streptavidin beads. Figure 3.2.6a1 shows that ESC and NPSCs samples were run on a 1% agarose gel after biotinylation to ensure RNA integrity, following this biotinylated samples underwent a dot blot assay to ensure that only samples with 4sU were labelled (fig 3.2.6a2) and finally, after biotin affinity purification, samples were run on a 1% agarose gel to show that in samples without 4sU (-), RNA was absent and samples that had incorporated 4sU were purified and had not degraded (obvious by the presence of the 28S and 18S rRNAs) (fig 3.2.6a3) (refer to chapter 5 for further details about this technique). 3'RACE of these samples show that post-transcriptionally regulated primary miRNAs are expressed in newly transcribed RNA confirming that miRNA genes are actively transcribed in ESCs although they are not processed into mature miRNAs (fig 3.2.6b).





C

MicroRNA	ESC	NPSC
21	445	3162
125b	0	297
7	0	150
302b	115	0
302a	2475	0

Fig 3.2.6 Detection of miRNAs in nascent RNA samples from ESCs and NPSCs indicates that post-transcriptionally regulated miRNAs are continuously being transcribed. A) Detection and integrity analysis of 4sU incorporated RNA. 1) 1% agarose gel of total RNA after biotinylation of ESC RNA with no 4sU ("-") and "4sU ESC" and "4sU NPSC" representing the RNA of ESCs and NPSCs treated with 200 µM 4sU 2) Dot blot assay showing the thiol-specific biotinylation of RNA. 10 fold dilutions of RNA was blotted down the membrane. 3) 1% agarose of RNA after biotin affinity purification, the "+" control was total ESC RNA. The "-" control was ESCs RNA with no 4sU showing that only 4sU incorporated RNA could be purified. +: Total ESC RNA; -: ESC RNA with no 4sU incorporation; 4sU ESC: ESCs treated with 200 µM 4sU; 4sU NPSC; NPSCs 200 µM 4sU. In C) +: 5'-biotinylated DNA oligo. B) 3'RACE PCR and RT-PCR of nascent ESC and NPSC RNA show that post-transcriptionally regulated miRNA are continuously being made. C) The table shows sequencing data illustrating the mature miRNA expression in each cell type.

3.2.6 Over-active expression of primary miRNAs in ESCs suggest a prevalence of post-transcriptional regulation of miRNAs

In order to estimate the number of miRNAs that are regulated at a post-transcriptional level in human ESCS, I analysed all of the miRNA genes on chromosome 1 and 2 (miRGator v3.0: http://mirgator.kobic.re.kr/) that are poorly expressed in ESCs but strongly expressed in differentiated cell types.

Initially I investigated 37 miRNAs on chromosome 1 and 2 that are highly expressed in mature human cells but not ESCs. Out of these, 18 primary miRNAs were expressed in ESCs (Figure 3.2.7). Figure 3.2.7 shows that four miRNA genes (miR-34a, miR-101-1, miR29c and miR-200a) were analysed by RT-PCR and the remaining miRNAs were analysed by 3'RACE as their transcripts overlap with other genes. Treatment with actinomycin D revealed that 16/17 of these miRNAs were actively transcribed and turned over in ESCs despite the lack of mature miRNA expression (Fig 3.2.7). MicroRNA 200a is expressed in ESCs however it is not turned over; this could be an example of a dead-end product that is not actively transcribed. MicroRNA 101-1 was used as a positive control as this miRNA is expressed in ESCs and turned over. The remaining twenty of the 37 microRNAs did not show evidence of primary miRNA expression in ESCs. We have confirmed that many of these miRNAs are regulated by transcription in ESCs as primary miRNA expression is evident in expressing cell types using the same primers.

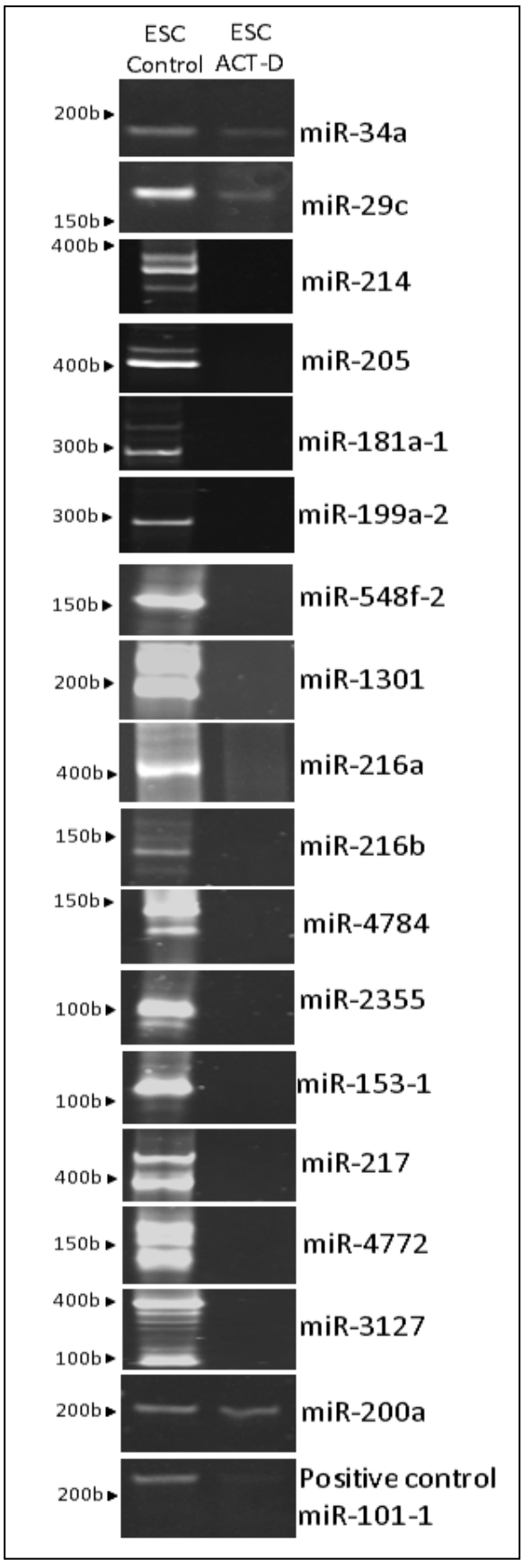


Figure 3.2.7 Over-expression of primary miRNAs in ESCs implies that the processing of many miRNAs is blocked. 3'RACE PCR and RT-PCR of total ESC RNA treated with DMSO (Control) and 1μM Actinomycin D (Act-D) for 4 hours, indicates that the majority of primary transcripts expressed in ESCs are turned-over after 4 hours of inhibiting transcription with Actinomycin D. There are no sequencing results available for these miRNAs in ESCs indicating that these mature miRNAs are not expressed.

3.3 DISCUSSION

3.3.1 Transcriptional regulation of miRNAs

A great of deal of evidence suggests that like most other protein coding and non-coding RNAs, transcription is a major regulator of miRNA biogenesis (Wang et al., 2011, Marson et al., 2008). 18 of the 29 miRNAs investigated were regulated by transcriptional mechanisms (figure 3.2.3). Evidence of transcriptional regulation in ESCs can be seen clearly in the relationship between miR-302 and pluripotency factors such as Oct4 and Sox2, which are known to bind the promoter of the miR-302 gene cluster in order to regulate its transcription (Marson et al., 2008). In parallel, the miR-302 cluster is known to regulate Oct4 and Sox2 resulting in a feedback mechanism that is able to fine tune gene expression to achieve cell determination (Liao et al., 2011). Epigenetic regulation of miRNAs is a key factor in transcriptional control (Suzuki et al., 2011). In mouse ESCs, miR-9 is under epigenetic regulation resulting in silencing in ESCs and expression in NPSCs. The miR-9 promoter contains bivalent domains, so its expression is dependent on two histone marks, when H3K4me3 is lost and H3K27me3 is retained the genes are transcribed, however when the opposite occurs expression is silenced (Suzuki et al., 2012, Marson et al., 2008). Regulation of miRNAs by transcription factors and epigenetic markers operate in a feedback loop mechanism whereby upregulation or downregulation of miRNAs influence signalling cascades which can cause reversible fine tuning of gene expression within a cell (Babiarz and Blelloch, 2008). Regulation of miR-9 and

miR-302 demonstrates the importance of reversible transcriptional regulation indicating that feedback mechanisms are in place to promote cell fate.

3.3.2 Post-transcriptional regulation of microRNAs is prevalent in ESCs

There is growing evidence for regulation of miRNA biogenesis which could occur at any of the steps in the processing pathway including Drosha cleavage, nuclear-cytoplasmic export, Dicer cleavage and RISC loading and silencing (Heo et al., 2008, Davis and Hata, 2009, Kedde et al., 2007, Michlewski et al., 2008) (fig 3.1.3).

Deep sequencing of miRNAs demonstrates a huge difference in miRNA expression between ESCs and NPSCs (supplementary table S1), suggesting important functions for those miRNAs in maintaining pluripotency and promoting differentiation. Analysis of differentially expressed miRNA genes indicates that the majority of miRNAs were regulated transcriptionally. The 11 out of 29 miRNAs that were regulated by post-transcriptional mechanisms were regulated in this manner in ESCs only. Prevalence of post-transcriptional miRNA regulation in ESCs is a phenomenon also observed in mouse embryonic development (Thomson et al., 2006). Thomson et al., 2006 describes extensive post-transcriptional regulation at the Drosha processing stage resulting in downregulation of several miRNAs in mouse ESCs and primary tumours. Our data presents a lack of correlation between primary, precursor and mature miRNA expression levels in ESCs while differentiated NPSCs appeared to show more consistent expression between the different miRNA stages; these observations were previously made in mouse ESCs (Thomson et al., 2006). This

infers that Drosha regulation is necessary to regulate miRNA expression in early developmental stages which could otherwise be detrimental (Thomson et al., 2006).

As expected, the post-transcriptionally regulated miRNAs I identified included miR-let 7a-1, miR-let 7a-2 and miR-let 7f-1 for which the mechanism of post-transcriptional activity is well known (Heo et al., 2008). Interestingly miR-let 7f-2 was identified to be regulated transcriptionally which implies that the two independent sources of mature miR-let-7f are controlled by a complex transcriptional/post-transcriptional system. Evidence suggests that miR-542, miR-450b, miR-29a, miR-29c, miR-125b-1, miR-21 and miR-7 are also regulated post-transcriptionally in ESCs. Expression of these miRNAs have known implications in stem cell development and several cancers (Bray et al., 2011, Gebeshuber et al., 2009, Lui et al., 2007, Sengupta et al., 2008, Shalom-Feuerstein et al., 2012) and miR-125b and miR-21 have even been proposed to be suitable biomarkers for specific types of cancers (Cho, 2010, Chen et al., 2008).

In many types of cells, miR-21 is post transcriptionally regulated at the Drosha step by Smad proteins (Davis and Hata 2009). The Smads are signalling proteins of the TGF-β family signalling cascade and can be co-precipitated in a complex with Drosha on the pri-miR-21 hairpin upon miRNA stimulation with TGF-β (Davis and Hata 2009). Furthermore, factors upregulating or downregulating mature and precursor miR-21 did not alter the expression of primary miR-21 again indicating post-transcriptional regulation at the Drosha processing step (Davis and Hata, 2009). Davis and Hata, 2009 detail many miRNAs that are post-

transcriptionally regulated by Smads implying a crucial role of rapid alteration of miRNA expression in cytokine signalling.

It is possible that this mechanism of post-transcriptional regulation also occurs in ESCs however it has not been possible to perform loss-of-function assays in ESCs. ESCs are notoriously difficult to transfect because of their characteristic clonal growth pattern consequently it has not yet been possible to identify if Smad proteins are responsible for post-transcriptional regulation of miR-21.

MicroRNA 7 is described as a tumour suppressor in glioblastoma cells (Kefas et al., 2008). In ESCs miR-7 was revealed to be post-transcriptionally regulated, this is corroborated by data from glioblastoma tumour cells indicating that primary miR-7 is constantly expressed despite radical reduction of precursor and mature miR-7 expression (fig 3.2.4-6). Post-transcriptional regulation of miR-7 in glioblastoma tumours seemed specific to the miR-7 however the mechanism remains unknown (Kefas et al., 2008). In this case, miRNA expression and regulation in ESCs appear to match miRNA expression data in different cancer cell lines.

The literature remains elusive with regard to regulation of miR-125b-1, miR-450b, miR-542, miR-29a and miR-29c despite their significant biological relevance in controlling pluripotency, immune responses and roles in several cancers (Boldrup et al., 2012, Bray et al., 2011, Ma et al., 2011, Sengupta et al., 2008, Sun et al., 2013, Wang and Gu, 2012, Cui et al., 2012a, Cui et al., 2012b). Our data is the first to indicate that these miRNAs are regulated post

transcriptionally in ESCs. These miRNAs may also be post-transcriptionally regulated in cancer cells because of the similar properties shared with ESCs and the extreme differences in expression of these miRNAs in cancer (Bray et al., 2011, Sengupta et al., 2008, Sun et al., 2013, Wang and Gu, 2012). The mechanism behind post-transcriptional regulation of these miRNAs is yet to be confirmed. RNA editing and 3' modifications cannot be detected in the sequencing data of these miRNAs therefore it is likely that RNA binding proteins play a role in their regulation.

In order to further investigate the potential of ESCs to over-actively transcribe primary miRNAs, we systematically analysed microRNAs located on chromosome 1 and 2. The miRNAs investigated are highly expressed in differentiated cell types and poorly expressed in ESCs (figure 3.2.7). These results show that 16/37 primary microRNAs were actively expressed in ESCs despite a lack of mature miRNA expression in ESCs. Twenty out of 37 primary microRNAs showed no evidence of expression in ESCs thus indicating transcriptional regulation of these miRNAs. Nevertheless, preliminary analysis has shown that 43% of primary microRNAs from chromosome 1 and 2 are actively expressed in ESCs. This supports our evidence and others (Thompson et al 2006) that post-transcriptional regulation of primary miRNAs is widespread in ESCs.

3.3.3 The role of post-transcriptional regulation of miRNAs

Post-transcriptional mechanisms controlling miRNA expression in ESCs allows for rapid, tissue-specific and stringent fine tuning of miRNA expression. During differentiation of ESCs it is necessary to tightly control the expression of genes which promote cell determination. Consequently it is likely that miRNAs which regulate mRNAs involved in ESC differentiation are consistently transcribed and ready to be processed upon physiological stimulus. In differentiated lineages like NPSCs, they are defined in cell type and do not have the ability to differentiate therefore constant transcription of miRNA genes with no immediate function may not be necessary. As a result, ESCs ability to differentiate remains the rationale behind the extensive post-transcriptional regulation of miRNAs. The extensive post-transcriptional regulation of miRNA biogenesis that we observe in ESCs has been reported in different cell types and linked to human diseases, including cancer (Thomson et al., 2006, Siomi and Siomi, 2010).

CHAPTER 4 - EXTENSIVE ALTERNATIVE POLYADENYLATION OF MICRORNAS

4.1. INTRODUCTION

Solexa deep sequencing of ESCs and NPSCs gave rise to an abundance of anomalous miRNA expression data that did not match the predicted transcriptional models. For example the expression levels of mature miRNAs within clusters (e.g. the miR-302 cluster) differed greatly despite being regulated by the same promoter (Table 4.1.1). When using 3'RACE to investigate miRNA regulation we observed extensive use of potential alternative polyadenylation sites in miRNA genes expressed in ESCs and NPSCs. Alternative polyadenylation may provide an explanation for differential expression of miRNAs within clusters and between cell types.

MicroRNA	ESC	NPSC
Let-7f-1	0	4356
Let-7a-1	20	3701
Let-7d	0	183
99a	0	44
Let-7d	0	475
125b-2	0	273
Let-7a-2	0	3701
125b-1	0	273
100	0	197
302b	115	0
302a	2475	0
302c	68	0
302d	75	0
367	45	0

Table 4.1.1 Sequencing reads of miRNAs within clusters. MicroRNAs are grouped into clusters according to their genomic location and the relative sequencing numbers represent their expression in ESCs and NPSCs (Solexa sequencing by Dr Elcie Chan)

4.1.1 Alternative polyadenylation

Polyadenylation and 3' processing play a crucial part in eukaryotic mRNA synthesis which consists of co-transcriptional endonucleolytic cleavage followed by the addition of repeated adenosine bases by a poly(A) polymerase (PAP) enzyme. The structure and location of the 3'UTR and poly(A) tail is thought to influence all areas of mRNA processing including mRNA stability, translational efficiency and nuclear-cytoplasmic export (Colgan and Manley, 1997). 3' end positioning is dictated by a number of factors including the poly(A) signal (PAS), usually defined by multiple *cis*-elements e.g. an AAUAAA hexamer, a U- or GU-rich downstream element (DSE) and other sequences (Shi, 2012). PAS recognition is dependent on the cleavage and polyadenylation specificity factor (CPSF) complex which recognises the AAUAAA sequence (or alternative A-rich hexamer), while the cleavage stimulation factor (CstF) and the cleavage factor Im (CFIm) complexes bind auxiliary motifs such as the U- or GU- rich DSE (Fig 4.1.1) (Colgan and Manley, 1997).

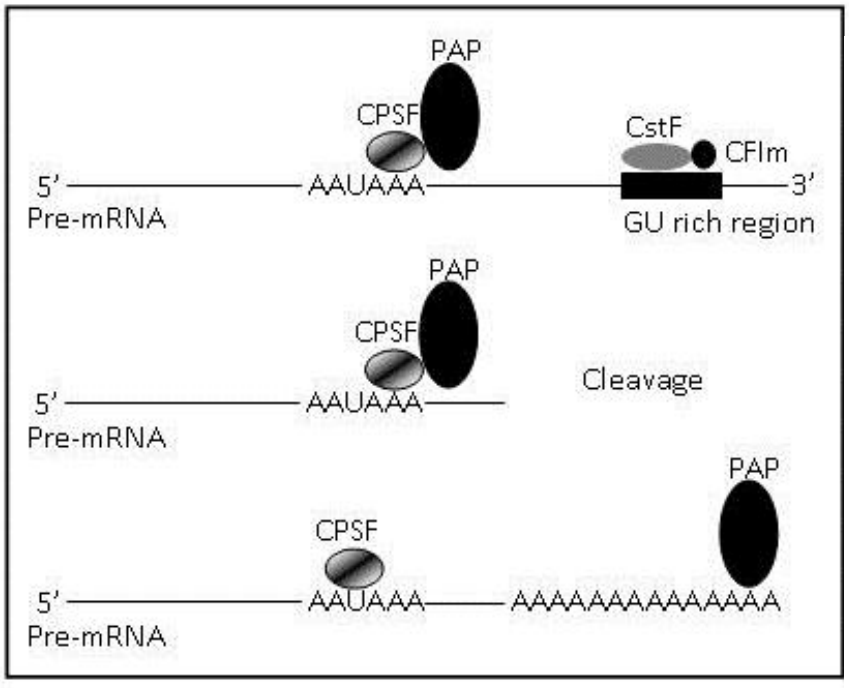


Figure 4.1.1 Polyadenylation of pre-mRNA. Diagram of the steps in eukaryotic polyadenylation and the factors involved. Initially proteins bind to the polyadenylation signal (PAS) AAUAAA and the downstream GU rich region. Shortly after, the 3' GU rich region is cleaved and the Poly(A) polymerase (PAP) catalyzes the addition of a string of adenosine bases to the 3'end of the mRNA. Abbreviations: CPSF: Cleavage and polyadenylation specificity factor, CstF: Cleavage stimulation factor, CFIm: Cleavage factor Im.

Alternative polyadenylation (APA) has been documented for some time however the mechanism has only recently been recognised as a major regulator of RNA expression (Boutet et al., 2012, Lutz, 2008, Mueller et al., 2013).

Upon completion of various global sequencing projects including the human genome project it was realised that over half of all mammalian genes are alternatively polyadenylated with many evolutionarily conserved alternative polyadenylation signals (Lutz, 2008, Mueller et al., 2013). The discovery of such extensive and widespread APA came as a surprise and furthermore over 70% of genes were found to be alternatively polyadenylated in human and yeast which implies that APA is far more ancient a mechanism than first imagined (Lutz, 2008, Mueller et al., 2013).

Lutz (2008) describes two types of alternative polyadenylation (fig 4.1.2); type I polyadenylation refers to genes that have two or more PASs in the 3'UTR (normally located with the canonical PAS) and type II polyadenylated genes have a PAS signal located anywhere on the gene producing a truncated protein coding transcript (fig 4.1.2). Type I APA can influence regulation of the transcript whereas Type II APA can result in the production of aberrant truncated protein isoforms.

Changes in 3'UTR length of mRNA (type I APA) can have a huge effect on abundance through a variety of mechanisms. Shortening of 3'UTRs can result in a decrease in miRNA targeting of transcripts and removal of NMD motifs affecting

turnover of transcripts and affecting transcript localization and translational efficiency (Mueller et al., 2013).

Many factors can determine APA activity including activity of core polyadenylation factors e.g. CstF and CFIm, tissue specific RNA binding proteins, and mutations of canonical PASs (Gruber et al., 2012).

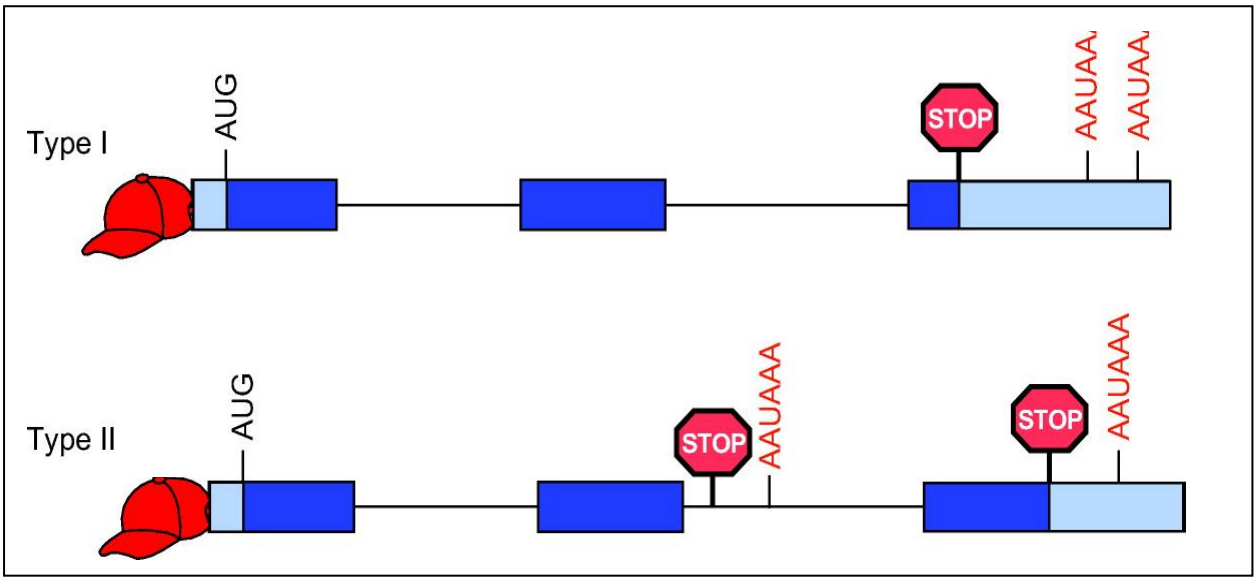


Figure 4.1.2 Types of alternative polyadenylation. For type I alternative polyadenylation there is more than one polyadenylation signal in the 3'UTR. Type II alternative polyadenylation involves alternative splicing coupled with alternative polyadenylation (Lutz, 2008).

4.1.2 Alternative polyadenylation and stem cells

Tissue specific APA has been investigated bioinformatically revealing a pattern of APA bias, where a particular PAS is in significantly higher use than another in certain tissue and cell types (Shi, 2012, Winter et al., 2007, Licatalosi and Darnell, 2010, Ji and Tian, 2009). This is likely to be a consequence of the cellular choice of a promoter which may dictate the PAS that defines the 3'end of the mRNA, however no mechanism for this is known (Mueller et al., 2013, Ji and Tian, 2009).

An example of tissue specific APA is seen in the expression of Opitz syndrome gene MID1 where the cellular choice of an alternative transcriptional start site (TSS) and promoter dictates the use of a particular PAS (Winter et al., 2007).

Recent microarray and RNA sequencing analyses have indicated that APA is important in regulating stem cell pluripotency and differentiation (Mueller et al., 2013). Global mRNA analyses in a variety of species have led to the discovery of the dynamic nature of APA in proliferating cells such as ESCs (Mueller et al., 2013). Proximal PASs generates mRNAs with shorter 3'ends in ESCs whereas in differentiated cells distal PASs generate mRNAs with long 3'UTRs. Furthermore, microarray data for cells at different stages of reprogramming of iPSCs displayed shorter 3'UTR lengths compared to the somatic cells from which they were derived (Ji and Tian, 2009). Ji et al 2009 show 3'UTR lengthening during mouse embryonic development again showing that proliferating cells like ESCs have mRNA with shorter 3'UTRs than differentiated cell types (Mukhopadhyay et al., 2009). Mueller et al., 2013 theorise that global changes in 3'UTR length seen in ESCs could contribute to general miRNA availability and activity.

APA is functionally significant to the regulation of the mRNA by miRNA and is particularly prevalent in cancer cells which have high expression of short 3'UTR mRNA isoforms that evade miRNA regulation (Clancy et al., 2011, Mayr and Bartel, 2009). APA also has the potential to serve as a biomarker in cancer as there is a precise APA molecular signature that is distinguishable between similar tumour subtypes (Di Giammartino et al., 2011).

A bioinformatic study revealed that based on EST information, alternative mRNAs lacking target sequences for specific miRNAs were preferentially expressed in cell lines where the targeting miRNAs were also expressed (Legendre et al., 2006). This implies that the alternative transcripts would be able escape regulation by miRNAs but the canonical transcript with the long 3'UTR could not (Legendre et al., 2006). This has been experimentally verified with Pax3, a myogenic regulator expressed in two cell subtypes, satellite cells (SCs) and quiescent satellite cells (QSCs). Both cell types express Pax3 and miR-206 which is a negative regulator of Pax3; however QSCs produce an alternatively polyadenylated Pax3 transcript with a shorter 3'UTR escaping miR-206 regulation resulting in higher protein expression in these cells (Boutet et al., 2012).

An abundance of mRNA contributes to reducing the effects of miRNAs and siRNAs by behaving like miRNA sponges. The availability of miRNAs which act on multiple transcripts is directly affected by the abundance of any one of the targeted mRNAs. Therefore mRNAs with long 3'UTRs can act as sponges, so a global reduction of 3'UTR length as seen in ESCs and cancer cell types can result in an overall increase in free miRNAs (Mueller et al., 2013).

4.1.3 Regulation of microRNA biogenesis by alternative polyadenylation

The number of novel miRNAs has been rapidly growing due to advancements of sequencing technologies, with a current estimate of over 2578 miRNAs in the human genome. Despite these advances the primary miRNAs which encode them

remain largely uncharacterized. Consequently there are few published examples of alternative polyadenylation regulating primary miRNA expression (Ribas et al., 2012).

A single primary miRNA, miR-21 is reported to have two sources of origin which I confirm in ESCs and NPSCs (supplementary figure S1). One transcript has an independent promoter within the VMP1 gene intron just a couple of hundred bases away from the precursor making a primary transcript separate from the VMP1 gene. The other uses the promoter of VMP1 and alternative polyadenylation to make a hybrid VMP1-miR-21 transcript that is processed by the microprocessor machinery and made into mature miR-21 (Ribas et al., 2012). This shows that alternatively polyadenylated miRNA transcripts can be functional and processed in the same way as canonical primary miRNAs. In addition it could allude to the importance of promoter choice in miRNA regulation.

More and more evidence points towards cross-talk between transcription and primary miRNA processing. It has been reported that exogenous primary miRNAs without poly(A) tails cannot be processed into mature forms and are retained at the transcription site (Pawlicki and Steitz, 2010). Furthermore the choice of promoter is likely to play a part in primary miRNA production. It has been shown that expression of a primary miRNA from a miRNA promoter *in vitro* results in higher expression of Drosha protein and increased processing of the primary miRNA compared to different promoters e.g. cytomegalovirus (CMV) driving the same primary miRNA (Ballarino et al., 2009). It can be concluded that

many intricate features could play a role in primary miRNA processing and alternative 5' and 3' processing may play a large part despite remaining relatively unknown as a post-transcriptional method of regulating miRNAs.

The alternative polyadenylation of clustered miRNAs may promote distinct miRNA expression. From our deep sequencing data and others published it is obvious that miRNAs within clusters, sharing the same promoter, can have very different expression profiles and the method of post-transcriptional regulation for miRNAs within a cluster has not yet been identified (Thomson et al., 2006, Tang and Maxwell, 2008).

4.2 RESULTS

4.2.1 Annotation of the miR-302 cluster with evidence of extensive alternative polyadenylation

3' Rapid amplification of cDNA ends (3'RACE) allows amplification of a specific gene with an unknown 3'end. The method involves binding of an anchored oligo(dT) primer adapter to the poly(A) tail of the RNA and then amplifying the transcript with primers against the oligo(dT) adapter and a gene-specific sequence.

This method identified several putative 3' ends of each miRNA within the miR-302 cluster; this is evident by the multiple bands seen in the ESC lanes in figure 4.2.1a. These bands were sequenced and the genomic positioning of the primary miRNAs were mapped and can be visualised in figure 4.2.1b, note that the 5'ends

are artificially determined by the gene specific primer. Figure 4.2.1b displays the numerous putative primary transcripts produced by alternative polyadenylation matched against the canonical annotation of primary miR-302 (Barroso-delJesus et al., 2008).

Detection of alternative polyadenylation sites using anchored oligo(dT)s increase validity of the alternative primary transcripts generated by 3'RACE. This is because anchored oligo(dT)s have been proved to bind to genuine 3' poly(A) tails even in competition with internal poly(A) sites that are up to 14 adenine nucleotides long (Nam et al., 2002). Anchored oligo(dT) primers consist of a string of 20 deoxythymidylic acid (dT) residues followed by two additional nucleotides that act as an anchor allowing the primer to anneal only at the 5'end of the poly(A) tail of RNA, providing more efficient cDNA synthesis for 3'RACE (see chapter 2). In agreement with this, we found that the majority of the 3' ends of our 3'RACE transcripts did not coincide with internal polyA sites.

To help verify the authenticity of the APA sites, potential polyadenylation signals were mined using a web-based tool (created by Mr T. Coggrave) which annotated the 12 known poly(A) motif variants in a given sequence (Kalkatawi et al., 2012) (supplementary table S2) (Polyadenylation mining web tool: https://github.com/coggy/signals_in_sequence). The data mining tool found 17 potential polyadenylation signals between the genomic positions chr4:113,568,706-113,569,780. Polyadenylation signals are reported to be found ~10-30nts upstream from the poly(A) tail, in this case putative polyadenylation signals could be predicted upstream from a few of the novel 3'ends of the miR-

302 cluster denoted by the "*" next to the miRNA transcripts in figure 4.2.1b (supplementary figure S2 shows the sequence in full with annotated PASs and 3'ends aligned with the mouse sequence to show conservation of the PASs). As there are many additional factors involved in APA recognition we cannot dismiss the other transcripts discovered but further evidence is required to prove that the transcripts are genuine.

Different alternative 3'ends could promote or reduce expression of individual miRNAs from the miR-302 cluster. In addition, there also seem to be splicing mechanisms in place to isolate miR-302b and miR-302b* from the remaining miRNAs of the miR-302 cluster. Figure 4.2.1c shows that the splicing would disrupt miR-302b and remove the remaining miRNAs. It is possible that miRNAs 302c, 302a, 302d and 367 could still be made from the spliced lariat but not miR-302b as it would be unable to form a stem region (fig 4.2.1c).

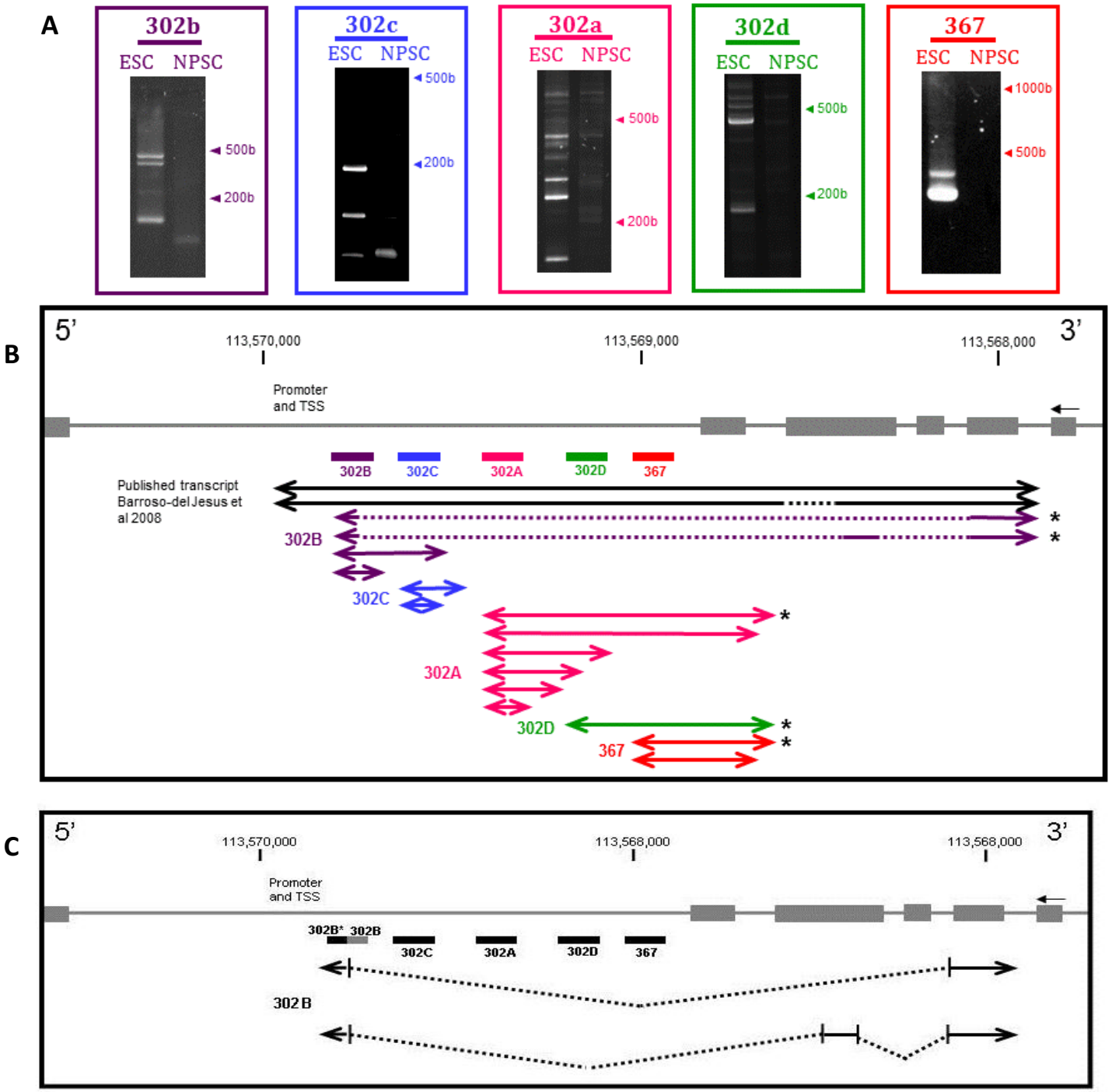


Figure 4.2.1 The 302 microRNA cluster has many 3' ends potentially allowing differential expression of miRNAs within the cluster A) 5% PAGE of 3' RACE PCR products of primary miR-302 in ESCs and NPSCs B) A Schematic diagram of the sequencing results from the 3' RACE PCR products. The sequences are aligned to the LARP7 gene intron where the miRNA cluster resides. The black arrows represent sequences identified as the primary miRNA for the cluster by Barroso-del Jesus et al., 2008. The arrows are colour coordinated to represent the different 3'RACE bands of the different miRNA. The arrows which are starred (*) represent transcripts whose 3'end is verified by a nearby PAS. The dotted line signifies alternative splicing of the transcript. C) Illustration of 3'RACE showing two different splicing variations of the miR-302 cluster. The primary transcript produced only includes the miR-302b*. With only half a precursor miRNA it would not be possible to form the stem loop that is recognised by the microprocesser machinery therefore degradation of this transcript is likely.

4.2.2 Annotation of the transcription start site of the microRNA 302 cluster

5'RACE of the miR-302 cluster corroborates existing data annotating the TSS 153 nts upstream of precursor miR-302b (Barroso-delJesus et al., 2008) (Fig 4.2.2). The outcome of individual miRNA expression within the miR-302 cluster could also be dependent on the use of alternative 5' TSS. However unlike miR-21 whose alternative 3' polyadenylation is accompanied by alternative 5' TSS, multiple 5' RACE experiments of transcripts downstream of the annotated 5' TSS indicated there were no other TSSs in the miR-302 cluster (data not shown).

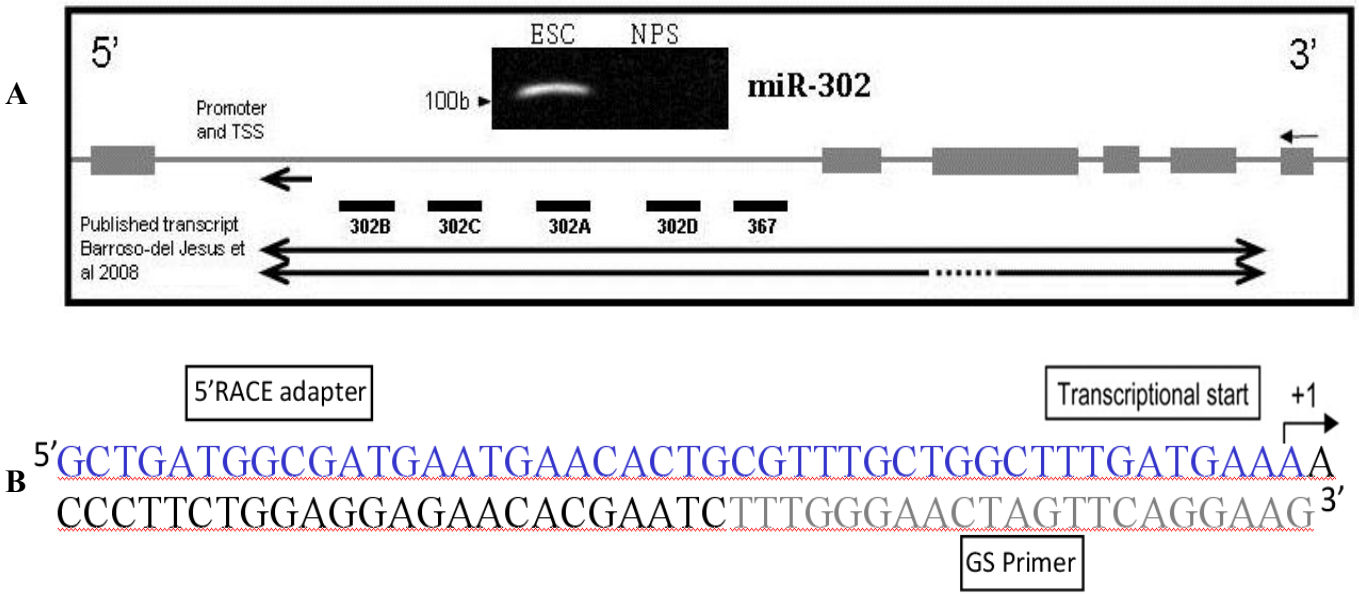


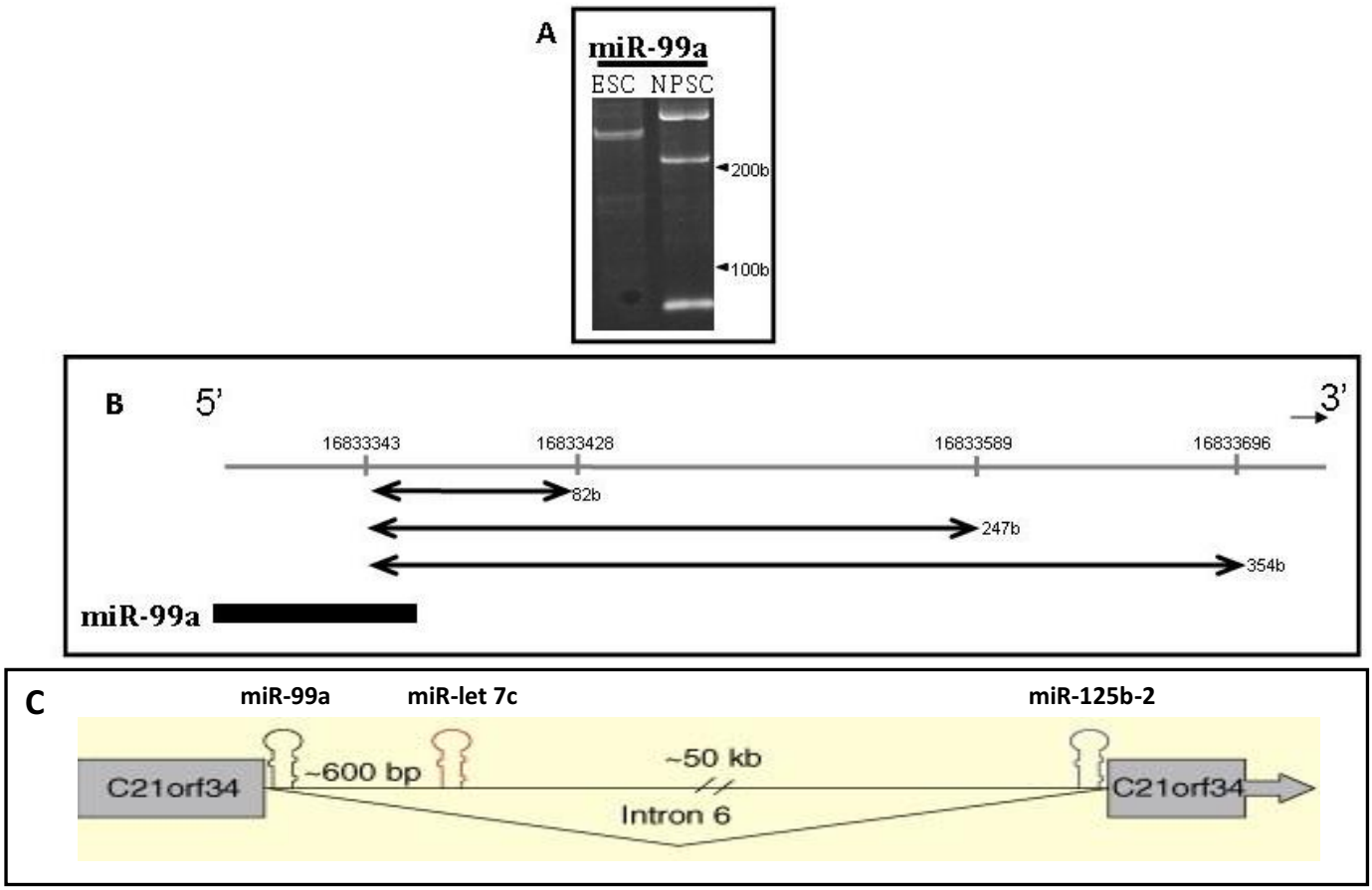
Figure 4.2.2 Identification of miR-302 transcription start site corroborating existing data from Barroso-del Jesus et al 2008. A) The band in the ESC line represents the transcript identified as the TSS displayed in the schematic diagram. The TSS was identified using 5'RACE which involves ligating an adapter to the 5'end of the RNA and using primers specific to the gene and adapter to perform nested PCRS. As expected a TSS could not be detected in NPSCs. B) Sequencing of the 5'RACE product, 5'RACE adapter is highlighted in blue lettering, the 3' gene specific primer (GS primer) is highlighted in grey. The transcription start site is annotated with the (+1) arrow, indicating the position that 5'adapter annealed to.

4.2.3 Alternative polyadenylation of a singular miR-99a

Another example of alternative polyadenylation occurs in miR-99a, a conserved intronic miRNA located on chromosome 21 next to miR-let-7c and miR-125b-2 however it is unknown if the three miRNAs are polycistronic. Interestingly the sequencing data indicates that the expressions of the three mature miRNAs are different in NPSCs (fig 4.2.3c, Table 4.1.1).

Figure 4.2.3a shows that there are three alternative 3'UTRs in NPSCs and the single band in the ESC line proved to be non-specific after sequencing. Figure 4..2.3d shows the conserved sequence alignment of human and mouse miR-99a. The highlighted yellow sequence shows putative PASs (using PAS mining tool: https://github.com/coggy/signals_in_sequence) and the red sequence indicates the position of the three novel 3'ends established by DNA sequencing of the bands from figure 4.2.3a.

As illustrated the multiple 3'ends of pri-miR-99a are supported by potential conserved PASs within 30 nt upstream of the poly(A)tails. MicroRNA 99a expression is heavily downregulated in several types of cancers and studies show that RNA in cancer cells are highly alternatively polyadenylated, with a tendency towards shorter 3'UTRs which can be subject to less post-transcriptional regulation. If miR-99a is transcribed together with nearby miRNAs, miR-let-7c and miR-125b-2, the function of multiple 3'ends could be to differentiate expression of the miRNAs within the cluster as well as between cell types and allow fine tuning of expression under different conditions.



D

```
HUMAN       CCCATTGGCATAAACCCGTAGATCCGATCTTGTGGTGAAGTGGACCGCACAAGCTCGCTT
MOUSE       CCCATTGACATAAACCCGTAGATCCGATCTTGTGGTGAAGTGGACCGCGCAAGCTCGTTT
            *******.****************************************.******** **

HUMAN       CTATGGGTCTGTGTCAGTGTGGTAATCTGACAAAATGCTATACACAGTGCCGTTCAACAA
MOUSE       CTATGGGTCTGTGGCAGTGTGGTGATCTGACAAAGCACTATGTAAGTGTTGCTCACTCAC
            ************* *********.**********. .****. *..      * ..:**.

HUMAN       TAGTTCAGTAAATCCTGTTAAACTCCAGTTGATTATATACTTTTGAAGTCATTATATTTT
MOUSE       ATCTCAATACACTCCTGCTCACACTCCAGTCCTCTACACAACCTCGGAACAATAGCATTT
            :: * .* :.*.***** *.*..  *.. * .* ::.:..:  * ...:**:** .:***

HUMAN       CCTTTGTTTTTAATGTTTATTTTAATCATTGTCTGTCTTACAAGGCAGGCTTCAATTCTC
MOUSE       CCATTGTTCTTCATGGATCTTTTAATCCTCTCATGTCTTACAAGGCAAGTTTTAATTGTT
            **:***** **.*** :*.********.*   .**************.* ** **** *

HUMAN       AACAACTTGGAAGCGTTTATATCACACCCATTCAAGTTCGATTCCATGTACAGTAAATTG
MOUSE       GACTTTCTGGAAGGATTTGTAGCAGAGCCATTCAAGGCTAAATACAAGTGCAATAAACTG
            .**::  ****** .***.** ** * *********   .*:*.**:**.**.**** **

HUMAN       CATAAGAAAGTTGAACCTTTATAGCAGG---GTTTGGACCAGGATCTGAATAGATTCTGT
MOUSE       CTTTCAAAAGGTTGGTGTTTTTAAATAGCTGGCTTTGAGCAAACTCTGCATAGCTGTTGT
            *:*:..**** * ..  ***:**..:.*   * ** ** **...****.****.*  ***

HUMAN       TCAGAAACTTCAGTGCATACTGTAGTGC---TTAGTGTACCTCTTCAACTTCAGTAAGAA
MOUSE       CATGAAACTTTGGCACACTTTTTCTTGATGGTTAATGTGTCTTATCTATTTCAACAACA-
             .:******* .* .** : * *. **.   ***.***. ** :**:* ****. ** *

HUMAN       AATTAAAGACTACCATTGGATTTATATGTTTTTATTGAGCCTATTTCTCCTGAAAAAATG
MOUSE       TAGTAAAGTCTATAACTAGATTATTATGTTA----------------------------
            :* *****:*** .* *.****::******:
```

Figure 4.2.3 miR 99a shows extensive alternative polyadenylation A) 5% PAGE of 3'RACE PCR products representing primary miR-99a in ESCs and NPSCs B) A Schematic diagram of the sequencing results from the 3'RACE PCR products. The sequences show 3 different 3'end termini. The different length arrows represent the different sized 3'RACE bands (Fig 5A). C) A schematic of the genomic situations of miR-99a, miR-let 7c and miR-125b-2 (adapted from Roush and Slack 2008). D) Conservation of 99a sequence between human and mouse species. Matching sequences are labelled with a "*" below the nucleotides. Underlined sequence represents the pre-miR-99a, the sequences in yellow represents potential PASs and the red sequence annotates the multiple 3' ends established by 3'RACE and ABI 3730xl DNA sequencing.

4.2.4 MicroRNA let 7a/f cluster forms different primary transcripts in ESCs and NPSCs.

The microRNA let 7 family is well known to be post-transcriptionally regulated in ESCs, and our deep sequencing data confirms the absence of mature miR-let-7 expression in ESCs (table 4.1). 3'RACE of the miR-let-7a/f cluster revealed differential alternative polyadenylation of the primary miRNA between ESCs and NPSCs (fig 4.2.4a).

The gel in figure 4.2.4a representing miR-let-7a shows a larger band in the ESC sample. After sequencing of all the bands we were able to determine that three alternatively polyadenylated primary miR-let-7a/f transcripts were being made. Again, in these samples the 5' end was artificially determined by the gene specific primer.

In NPSCs individual primary transcripts for the two miRNAs were potentially made whereas in ESCs a large primary transcript incorporating both miR-let-7a and miR-let-7f was expressed (fig 4.2.4b). It is possible that NPSCs also make the large primary transcript incorporating both miRNAs; however limitations of 3'RACE means that the technique preferentially amplifies shorter transcripts implying that expression of longer transcripts may not be detected (fig 4.2.4b). As mentioned previously, let-7 miRNAs are specific to NPSCs and known to be post-transcriptionally regulated in ESCs (Heo et al 2008); meaning that the primary or precursor form is present in ESCs but the mature transcript is not processed. With this is mind, the difference in polyadenylation of the transcripts

between the two cell lines could play a role in how the transcript is inhibited in ESCs but processed in NPSCs.

The miR-let-7 family is most characterised for their mechanism of post-transcriptional regulation in ESCs. Van Wynsberghe et al., 2011 report that Lin28 binds to primary miR-let-7 co-transcriptionally, since polyadenylation also occurs co-transcriptionally it is possible that it plays a role in Lin28 silencing of miR-let-7. The novel 3'ends of miR-let-7a/f can be found downstream from a putative conserved PAS however the genomic distance appears to be further than 30 nts (fig 4.2.4c).

A

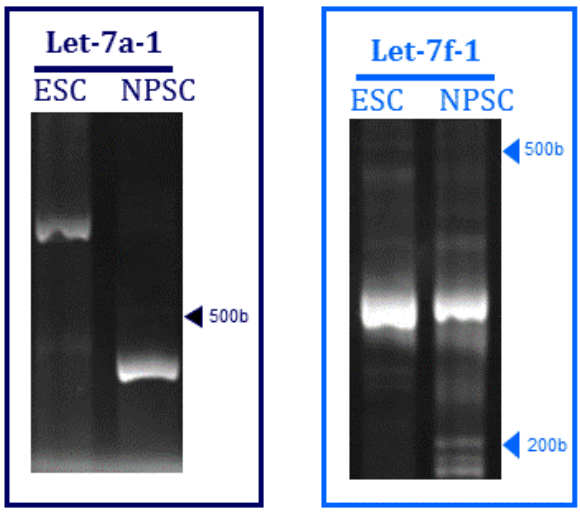
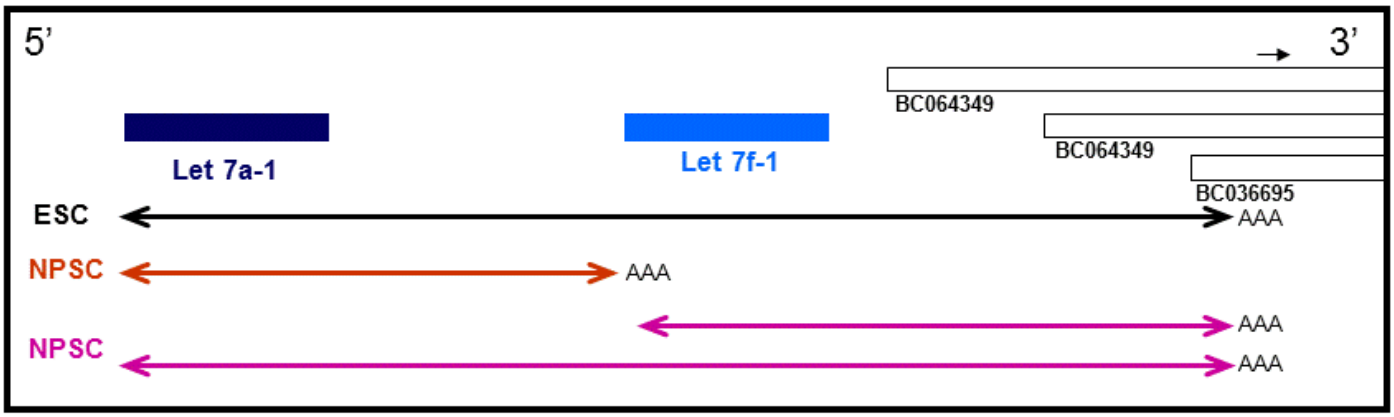


Figure 4.2.4 Primary miR-let-7a and miR-let-7f are differentially polyadenylated in ESCs and NPSCs. A) 5% PAGE of 3' RACE PCR products representing primary miR-let 7 in ESCs and NPSCs B) A schematic diagram of the sequencing results from the 3' RACE PCR products. The sequences are aligned to the genomic position of the precursor miR-let-7a-1 and miR-let-7f-1. The black arrow represents the sequence of the 3' RACE PCR of miR-let-7a-1 in ESCs. The orange arrow represents the 3' RACE PCR of miR-let-7a-1 in NPSCs. The pink arrow represents the 3' RACE PCR of miR-let-7f-1 in NPSCs. C) Conservation of let7a and let7f sequence between human and mouse species. Underlined sequence represents the precursor miRNAs, the sequence in yellow represents potential PASs and the red sequence annotates the multiple 3' ends established by 3'RACE and ABI 3730xl DNA sequencing.

B



C

```
HUMAN      GCGGGAATTCGATTTTTCACCATTCACCCTGGATGTTCTCTTCACTGTGGGATGAGGTAG
MOUSE      tcagctgtgattactttcaacattcaccctggatgttctcTTCACTGTGGGATGAGGTAG
                                                   ********************

HUMAN      TAGGTTGTATAGTTTTAGGGTCACACCCACCACTGGGAGATAACTATACAATCTACTGTC
MOUSE      TAGGTTGTATAGTTTTAGGGTCACACCCACCACTGGGAGATAACTATACAATCTACTGTC
           ************************************************************

HUMAN      TTTCCTAACGTGATAGAAAAGTCTGCATCCAGGCGGTCTGATAGAAAGTCAGTTAACTAA
MOUSE      TTTCCTAAGGTGATGGAAAAGTCTGCATTCATGGGGTCTCATAGGAAACCAAGAACAAAC
           ******** *****.************** ** * ***** ****.**. **. :*..:*.

HUMAN      TTGTACAATATTTAAGATTAACTTGTCTTAAAGAGATGTAGTGCAGCATTTGTTTATGGC
MOUSE      TGCAGT-GTTTTAAAGTATATCTTGCCTTAAAAGCATTTG------------CTTATGCT
           *  :.  .*:**:***::**:**** ******.. ** *.             *****

HUMAN      CTGGAAATAAATTAATTTAGAGATAAAGTCTGTAGCAAGTACACTGGATGGGGGTGGGGA
MOUSE      ATGCATGCTTTTTAATTTAACAAGT-----------------AGATGATGGGAATGGGC-
           .** *:. :::********...* :                 * : ******..****

HUMAN      AACCTTTTGCTTCTTGTCTTATTTCTCTGTGTCAGAATAAATGTATTTTTTTATTTTGAT
MOUSE      AAGTTTTTACTTACT----TTGCATTCTGTGTCAATGTAAATG-AAGATTTAATTTTGAT
           **  ****.***. *    *:  : *********.:.****** *: :***:********

HUMAN      TTATGCTGATAATTTTATGTTGAAATTTTCTTTCGAAAGAGATTGTACTTTCCATTCCAG
MOUSE      TGATGATGATGATTTAATGTTAA-----CCTTTCTTGAAAGAGACTGATGCTTATTCCAG
           * ***.****.****:*****.*      ***** :.*.*** : *..*    *******

HUMAN      AAGAAAACATTGCTCTATCAGAGTGAGGTAGTAGATTGTATAGTTGTGGGGTAGTGATTT
MOUSE      AAGAAAAACTTGCTCTATCAGAGTGAGGTAGTAGATTGTATAGTTGTGGGGTAGTGATTT
           *******..************************************************

HUMAN      TACCCTGTTCAGGAGATAACTATACAATCTATTGCCTTCCCTGAGGAGTAGACTTGCTGC
MOUSE      TACCCTGTTTAGGAGATAACTATACAATCTATTGCCTTCCCTGAGGAGTAGACTTGCTGC
           ********* ************************************************

HUMAN      ATTATTTTCTTTTTATTTAGATGATATTAAAACTCGGAAGAATTAATTTTGACATTTTGT
MOUSE      ATTGTTTTTATCTAGAGGATATTATAAAACATAGGtgaattctgacctt-------ttgt
           ***.**** :* *:.:  * ** ***::*.*:.

HUMAN      ATTTACAGTTTATCAGTTAATTTTCTCTGTTCAAGTAGTACAGTAGGCACAGATTAACAT
MOUSE      tt----tattttacagttcatttcctcagtgcaataaacattgttataacatttttagtg

HUMAN      TTAAATTTTTCGCATATGGTATATTTCAGAAATTTGAAGTGAGCAAAAAAAAAAAACCTA
MOUSE      ttcatgtt----------------taaagaaatttggacaaa------------------
```

miR-let-7a-1

miR-let-7f-1

4.2.5 Each alternatively polyadenylated let 7a/f primary transcript found in ESCs and NPSCs can be processed into precursor miRNAs in HepG2s.

To further investigate the nature of the three primary transcripts, each of the sequences plus ~100 bases extra upstream and downstream of the sequences annotated in figure 4.2.5a-c were cloned into pTRIPz inducible lentiviral vector and packaged with pCMVΔ8.91 packaging construct and VSV-g envelop construct.

Figure 4.2.5d is a northern blot of total RNA from HepG2 cells infected with the three different pri-miRNA sequences and probed with a double-DIG labelled LNA that binds to the mature miR-let-7a and miR-let-7f sequence. Figure 4.2.5d shows that even uninfected HepG2 cells express mature miR-let-7a and miR-let-7f, however the precursor cannot be detected. In the control sample (PTE) infected with an empty vector the same can be seen. However in the samples infected with primary miR-let-7a and primary miR-let-7f (as separate transcripts) precursor miRNA expression is visible. The difference in size of 7 nts between the precursor miRNAs verifies the sizes reported on miRBase. The cells infected with the long precursor (miR-let-7af) incorporating both precursor transcripts (specific to ESCs) is also able to make both the precursor transcripts of miR-let-7a and miR-let-7f. Unfortunately due to the expression of mature miR-let-7a and miR-let-7f in control samples it was not possible to establish whether the infected transcripts could be processed into mature miRNAs. MicroRNAs let-7a and let-7f are ubiquitously expressed in most cell types and their sequence is conserved amongst many species, therefore it proved difficult to find a cell type that did not express miR-let-7a or miR-let-7f. As a result I was not able to

establish if the lentivirally expressed primary transcripts were made into mature miRNAs even though precursor production was obvious.

Expression of specific miR-let-7a, miR-let-7f or miR-let-7af pTRIPz transcripts was confirmed by RT-PCR of the sequence specific to the pTRIPz vector and the miRNA transcript (fig 4.2.5e and supplementary figures S3, S4 and S5). This ensured that the infected cell lines did make the miRNA expressed by the lentiviral vector and that this is the precursor transcript detected by northern blot.

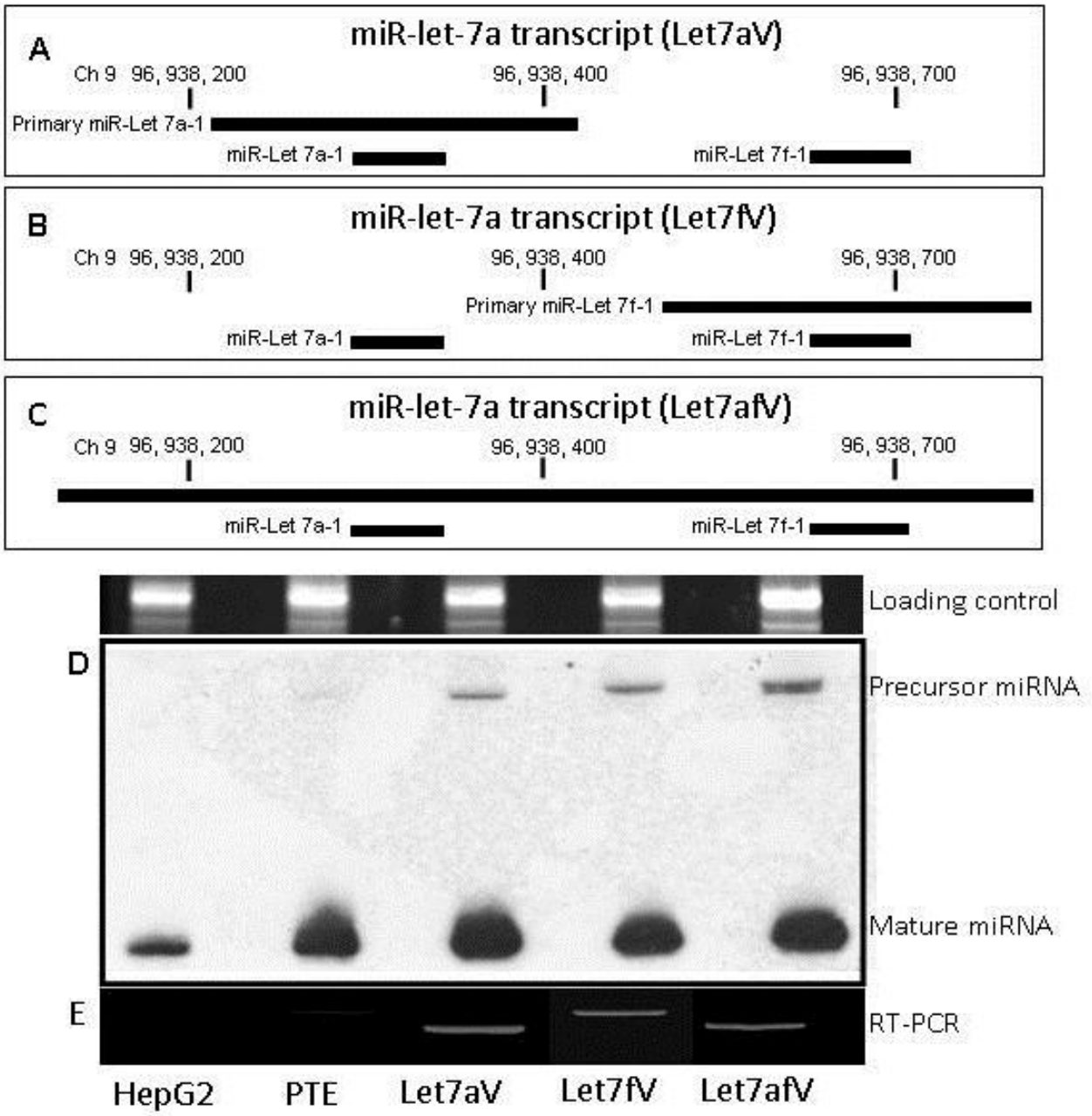


Figure 4.2.5 Lentivirus infection of HepG2 cells with primary miR-let-7a and miR-let-7f of different sequence lengths indicates that alternative polyadenylation alone cannot inhibit precursor miRNA production A-C) Displays UCSC BLAT results of the sequencing of the DNA inserts used to make miR-let-7a, miR-let-7f and miR-let-7af expressing lentiviral vectors. D) Northern blot of HepG2 cell infected with the 3 different lentiviral vectors expressing the various miRNAs in the cells expressing both miR-let-7a and miR-let-7f, two precursor transcripts can be seen indicting that all transcripts are capable of making precursor miRNAs. E) RT-PCR confirms the expression of the DNA insert-vector sequence in the infected cell lines.

4.3 DISCUSSION

4.3.1 MicroRNA gene clusters

Despite the discovery of over 2578 miRNAs in the human genome alone, very few miRNA genes have been annotated; miRNA promoters, transcriptional start sites and 3'ends remain elusive. Lack of knowledge about miRNA gene structure and miRNA biogenesis presents a problem as the increasing importance of miRNAs in biological processes and disease becomes known.

MicroRNA gene clusters are composed of two or more miRNA genes within proximal distance on a chromosome. They can be located in a polycistron and co-expressed to coordinate the regulation of multiple processes such as embryonic development and cell differentiation. MicroRNA gene clusters are predicted to be formed by insertion of tandem repeats and many clusters are evolutionarily conserved across several species (Zhang et al., 2009). Genomic studies have found that miRNA genes across species have a tendency to be found in clusters rather than being scattered across the genome (Lagos-Quintana et al., 2001, Lai, 2002). The discovery of more and more miRNAs in the human genome has revealed that about 40% of human miRNAs are located less than 3 kb from adjacent miRNAs (Kim et al., 2008, Altuvia et al., 2005).

4.3.2 Alternative polyadenylation in miRNA clusters

MicroRNAs investigated in this study were chosen because of their differential expression in ESCs and NPSCs regardless of their genomic location in relation to other miRNAs. Therefore the high occurrence of alternative polyadenylation in clustered miRNAs is likely to be widespread and is possibly a regulatory feature for individual expression of miRNAs.

Table 4.1.1 shows the vast differences in mature miRNA expression between individual miRNAs found in clusters e.g. miR-let-7a expression value is 3631 whereas the expression value for miR-let-7d is only 183. However the miRNA sequencing data refers to mature miRNA expression which includes miRNAs generated from different miRNA genes therefore their expression levels may not represent the exact expression from the primary miRNA cluster of interest. For example miR-let-7a-1 is located in a cluster with miR-let-7f-1 and miR-let-7a-2 is clustered with miR-125b-1 therefore the proportion of mature miR-let-7a made from each genomic location cannot be determined. This demonstrates the need for primary and precursor miRNA sequencing data in ESCs and differentiated cell types. Due to the varying degrees of sensitivity of sequencing techniques and the transient nature of precursor miRNA expression, this sequencing data in ESCs is not widely available and that which exists is inconclusive.

The miRNA-302 cluster is polycistronic, located on chromosome 4 and produces 5 mature miRNAs 302a/b/c/d and miR-367 (and their star counterparts) in a single transcription unit. The cluster is exclusively expressed in pluripotent cells and is evolutionarily conserved over a variety of species. The miRNA-302 cluster

is one of the few whose miRNA gene structure has been identified and the promoter characterised and functionally validated (Barroso-delJesus et al., 2008). Experimental evidence shows that each miRNA within the miR-302 cluster undergoes alternative polyadenylation which, in theory, could allow the expression of several different putative primary transcripts. It is possible that the miR-302 cluster could express separate primary transcripts for each individual miRNA or even a primary transcript incorporating a few but not all of the miRNAs could be preferentially expressed. MicroRNAs in the miR-302 cluster can only be generated from the locus on chromosome 4 and no other methods of post-transcriptional regulation of the miR-302 cluster have been identified. As polyadenylation tends to stabilise RNA transcripts it is likely that these alternative primary transcripts are expressed in order to be processed rather than to be degraded. Consequently it is possible that the differential expression seen in the deep sequencing data between the mature miRNAs could be influenced by alternative polyadenylation (fig 4.3.1).

The use of alternative polyadenylation will result in the generation of alternative primary miRNAs that do not incorporate all miRNAs within a cluster and therefore individual loop regions of the precursor miRNAs may be missing. The loop regions of miRNAs are important because they are known to be targets for RNA binding proteins (Thornton and Gregory, 2012, Viswanathan et al., 2008, Chakraborty et al., 2012) (see Introduction). Consequently alternative polyadenylation might be expected to alter both the miRNA and the RNA binding proteins known to associate with the primary miRNAs. To support this theory, tertiary RNA structure and interaction of stem-loops within a primary miRNA

cluster have been shown to play a huge role in processing (Chaulk et al., 2011, Chakraborty et al., 2012). Alterations of tertiary RNA structure can potentially sequester RNA binding sites to either promote or impede processing of miRNAs. Within the miR-17-92 cluster, the tertiary structure actually limits the accessibility of miR-92a resulting in decreased expression compared with other miRNAs in the cluster. In addition it has been shown in mice that altering the 3' end of pri-miR-17∼19 disrupts the tertiary RNA structure resulting in increased miRNA expression (He et al., 2005, Chaulk et al., 2011). Therefore it is possible that APA could alter tertiary structure of primary miRNA resulting in altered binding of associated proteins thereby influencing miRNA expression levels.

The 3'RACE of miR-302 also suggests that alternative splicing is active within the cluster however it is not yet clear what influence alternative splicing has upon miRNA production.

The miR-99a, miR-let-7c and miR-125b-2 cluster is located on chromosome 21 in the human genome and occurs in almost all species that express miR-let-7, however the order of the miRNAs and the spacing between them on the chromosome differ between species (Roush and Slack, 2008). In Drosophila the miR-99a cluster is polysistronic therefore transcribed as one primary miRNA. However in humans it is not yet known if this cluster is polycistronic or if it is further processed into individual primary miRNAs before processing by Drosha (Roush and Slack, 2008). As miR-99a and flanking sequences are conserved in Drosophila it is likely that human miR-99a is part of a cluster with miR-let-7c and miR-125b-2. 3'RACE of miR-99a identified three alternate 3'ends which may

function to regulate the expression of the individual miRNAs within the miR-99a cluster.

In the human genome the conserved miR-let-7a/f (also found with miR-let-7d) gene cluster found on chromosome 9, has been annotated with identification of four alternate 5'TSS and one 3'end, with all three miRNAs included in each of the alternative transcripts (Wang et al., 2011 and Zhao et al., 2011).

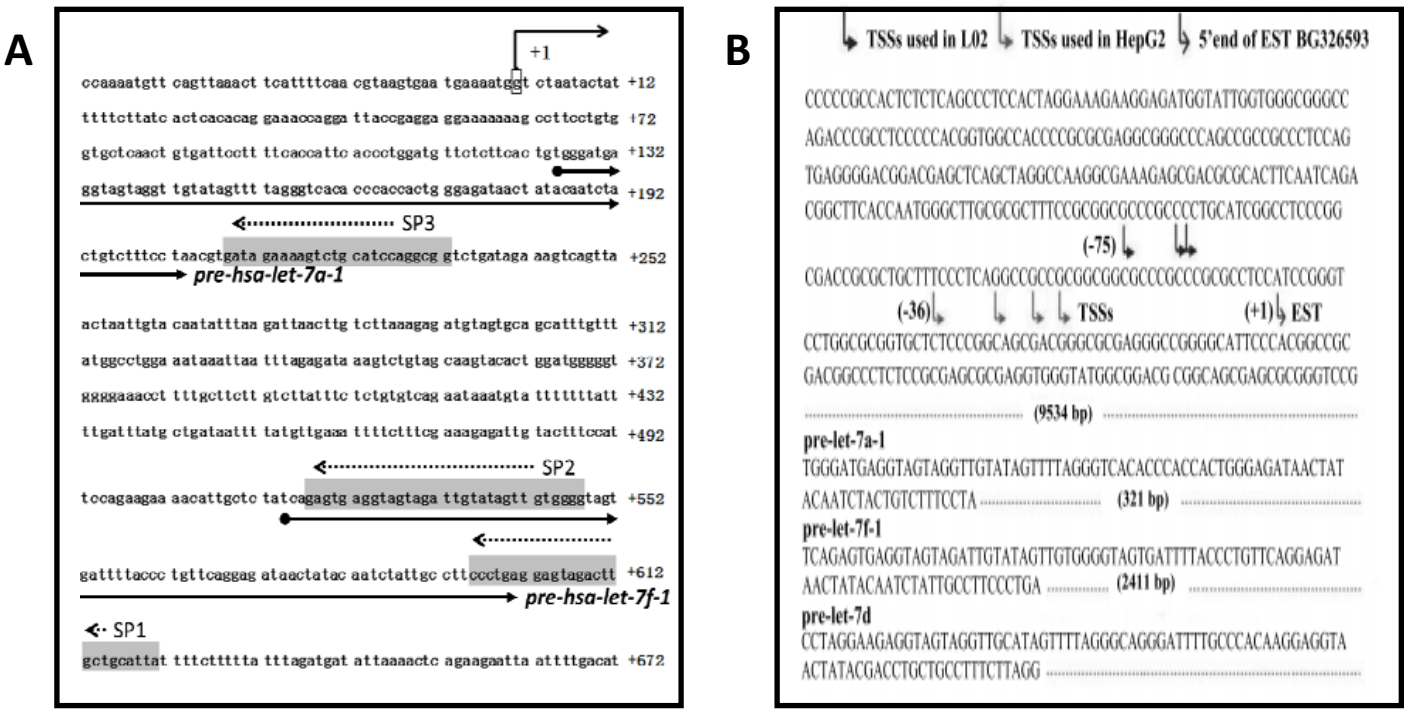


Fig 4.3.1 Maps of the 5' end of primary miR-let-7a-1/let-7f-1 cluster reproduced from Wang et al 2011 (A) and Zhao et al 2011 (B). In both publications the transcription start site detected by 5' RACE is indicated by labelled arrows.

The miR-let-7a/f cluster is regulated by MYC which binds directly to the promoter causing downregulation of the miR-let-7a/f cluster promoting tumorigenesis (Roush and Slack, 2008). The let-7a/f cluster is known to be post-transcriptionally regulated almost entirely by the Lin28 RNA binding protein in ESCs at the primary and precursor miRNA stage. Our results show that the miR-let-7a/f cluster is alternatively polyadenylated in NPSCs, utilizing a PAS between miR-let-7a and miR-let-7f resulting in separate primary transcripts. Alternative polyadenylation could be a novel mechanism for regulation of miR-let-7a/f specific to ESCs and is likely to play a role in the expression of precursor and mature miRNAs. Further experiments show that expression of separate primary transcripts for miR-let-7a-1 and miR-let-7f-1, identical to the alternatively polyadenylated transcripts, can be processed into precursor miRNAs in HepG2 cells which do not express Lin28. It is also possible that in ESCs the block in Drosha processing of the pri-miR-let-7 due to Lin28 binding could result in the inhibition of polyadenylation between pre-miR-let-7a and pre-miR-let-7f. Further investigation into the connection between alternative polyadenylation of miR-let-7 transcripts and Lin28 silencing of let-7 miRNAs could help to further unravel the mechanism behind post-transcriptional regulation of miR-let-7 in human ESCs. Unfortunately this has been difficult to investigate in ESCs as loss-of-function assays have not yet been possible to achieve due to the characteristic clonal growth pattern of ESCs making them resistant to transfection.

We have discovered extensive putative alternative polyadenylation of primary miRNAs, a phenomenon that has only recently been reported in miRNAs and could be an important novel factor in regulating miRNA biogenesis. We show

that alternative polyadenylation occurs in primary miRNA processing within and between cell types; potentially playing a role in regulating the miR-302 cluster and miR-let-7a/f as well as other miRNAs (miR-99a and miR-21 – see supplementary figure S1) confirming existing literature (Ribas et al., 2012). Alternative polyadenylation has the potential to play a regulatory role in miRNA synthesis, regardless of whether the miRNA is under transcriptional or post-transcriptional control.

CHAPTER 5 - DEVELOPING A WORKING ASSAY FOR THE NOVEL ANTI-TUMOUR DRUG SPLICEOSTATIN A

5.1 INTRODUCTION

5.1.1 Spliceostatin A and TG003

Spliceostatin A (SSA) is a specific inhibitor of splicing that possesses potent anti-proliferative and anti-tumour characteristics by causing arrest in the G1 and G2/M phases of the cell cycle. SSA is one of six synthetic inhibitors derived from the fermentation broth of *Pseudomonas* spp however SSA has the most enhanced stability and inhibitory activity (Kaida et al., 2007) (figure 5.1.1).

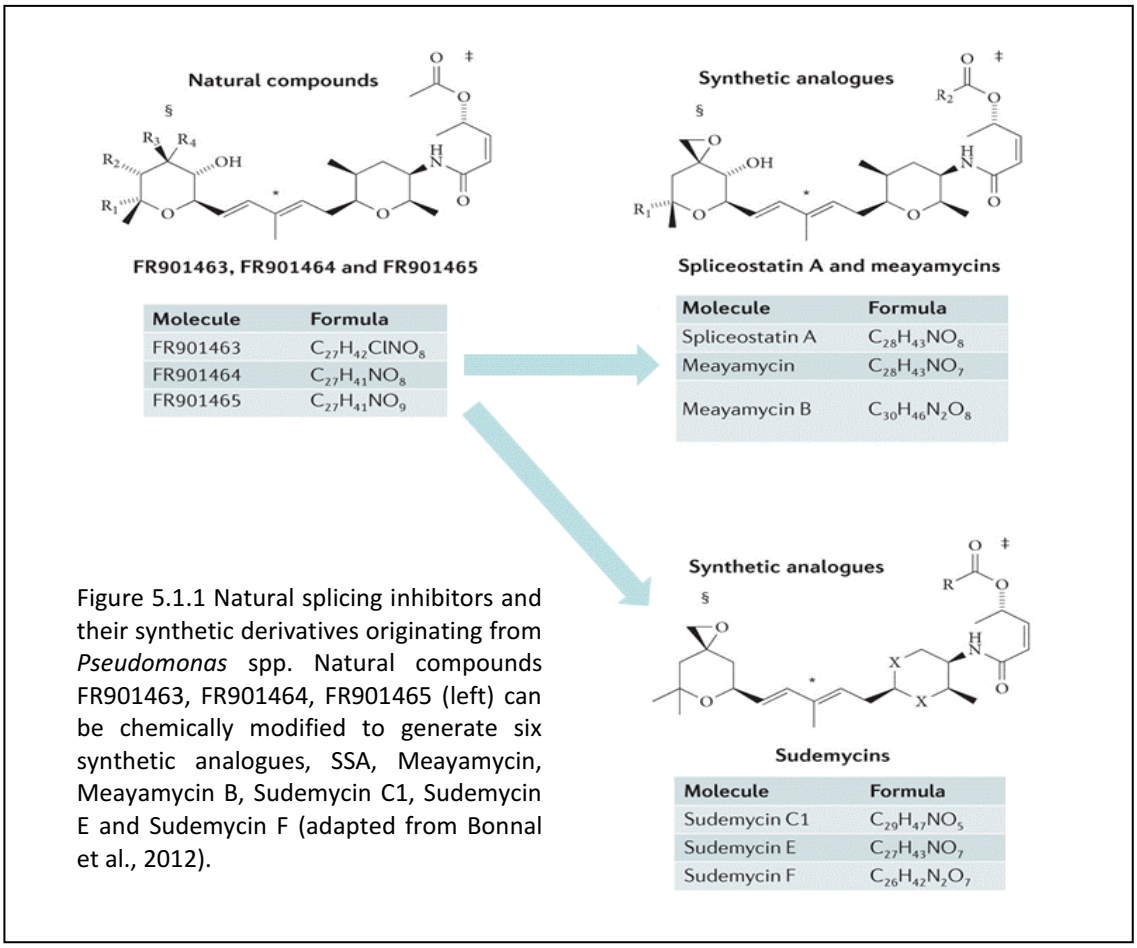


Figure 5.1.1 Natural splicing inhibitors and their synthetic derivatives originating from *Pseudomonas* spp. Natural compounds FR901463, FR901464, FR901465 (left) can be chemically modified to generate six synthetic analogues, SSA, Meayamycin, Meayamycin B, Sudemycin C1, Sudemycin E and Sudemycin F (adapted from Bonnal et al., 2012).

Following U1 snRNP recognition and association of the pre-mRNA 5' splice site, the 3' splice site (and polypyrimidine tract) is initially primed by the branch-point binding protein and U2 snRNP auxillary factors (Corrionero et al., 2011). These interactions result in an RNA secondary structure that activates further pre-mRNA spliceosomal processing (Kent et al., 2003). Subsequently the 17S U2 snRNP complex is enlisted, composed of 2 protein subcomplexes (SF3a and SF3b) and a number of polypeptides. SF3b is the larger of the two subcomplexes and is composed of at least eight subunits. U2 snRNP forms a pre-spliceosomal complex with the pre-mRNA by base-pairing to the branch site causing the branch site adenosine to bulge thus facilitating the initial catalytic splicing reaction. The largest and most conserved subunit SF3b-155 binds in a sequence-independent manner to the 5' and 3' end of the branch sequence holding the U2 snRNP to the location of the branch site (Kent et al., 2003).

Kaida et al 2007 initially showed that SSA can be immunoprecipitated with SF3b. It was then proven that SSA association with SF3b caused inhibition of the binding of SF3b to pre-mRNA branch site (Kaida et al., 2007). This initiates a conformational change that reduces RNA binding efficiency and disables a proofreading mechanism (Bonnal et al., 2012). Roybal and Jurica 2010 also reports that SSA prevents spliceosome assembly prior to U2 snRNP addition destabilising U2 snRNP and their pre-mRNA complexes which prevents subsequent spliceosomal assembly. These results are supported by the finding that knockdown of SF3b1 expression mimics the drug induced effects. (Kaida et al., 2007, Roybal and Jurica, 2010). Further studies show that SSA specifically

prevents the association of a 155-kDa subunit of SF3b (SF3b-155) with the pre-mRNA resulting in a lack of interaction of the U2 snRNP with the branch site (Corrionero et al., 2011). Furthermore SSA inhibition of SF3b can cause nuclear retention of pre-mRNA (Kaida et al., 2007).

Microarray analysis demonstrates that SSA does not globally inhibit splicing (Corrionero et al., 2011). Corrionero et al 2011 show that individual pre-mRNAs are diversely affected by SSA treatment. Some pre-mRNA 3'splice sites show different sensitivity to SSA treatment which can vary according to the base-pairing potential of the branch site with U2 snRNA (Corrionero et al., 2011). This signifies a role of SSA in influencing alternative splicing in addition to inhibiting general splicing (Corrionero et al., 2011). SSA induced alternative splicing results in the decline of cell cycle gene expression including cyclin A2 which could explain the mechanism behind the anti-proliferative effect of SSA (Corrionero et al., 2011).

Despite targeting basic steps in splice site recognition, SSA appears to have a cytostatic (not cytotoxic) effect on cells, especially in cancer cells (Bonnal et al., 2012). This could be due to a number of factors including cellular permeability, splicing demand and drug concentrations (Bonnal et al., 2012). In addition, SSA has been shown to affect only a subset of genes *in vivo* (Kaida et al., 2010, Younis et al., 2010).

The other inhibitor used in this study is TG003, a benzothiazole inhibitor of the SR protein kinases CLK1 and CLK4. Phosphorylated SR proteins CLK1 and CLK4 assist splicing reactions and spliceosome formation. Therefore TG003-induced inhibition of CLK1 and CLK4 can inhibit pre-mRNA splicing (Muraki et al., 2004, Hagiwara, 2005).

The focus of my research is to investigate the effects of the novel anti-tumour splicing inhibitor SSA on RNA transcripts *in* vivo and TG003 is used as a control to verify and compare its effects.

5.1.2 The importance of splicing inhibitors in disease

Currently there are only a few identified splicing inhibitors known to target the splicing pathway (Bonnal et al., 2012). Pre-mRNA splicing inhibitors promise to reveal original insights into the spliceosome and splicing pathways (Bonnal et al., 2012).

Abnormal splicing plays a major role in the production of aberrant proteins. This highlights the importance of pre-mRNA splicing inhibitors in the field of disease therapy (Bonnal et al., 2012). Targeting modulators of alternative splicing of mRNA transcripts is a potential treatment for diseases such as muscular dystrophies and neurodegenerative disorders (Bonnal et al., 2012). For example TG003 is of potential relevance for the treatment of Duchenne muscular dystrophy and SSA displays anti-tumoural properties (Bonnal et al., 2012). Thus SSA may be used not only as a powerful tool for analysing pre-mRNA splicing but

also for developing new therapeutics (Kaida et al., 2007). Since the discovery of SSA much has been revealed about the mechanism of splicing inhibition however its effect on RNAs *in vivo* and the mechanism behind its anti-tumour capabilities remain elusive.

5.1.3 Aims

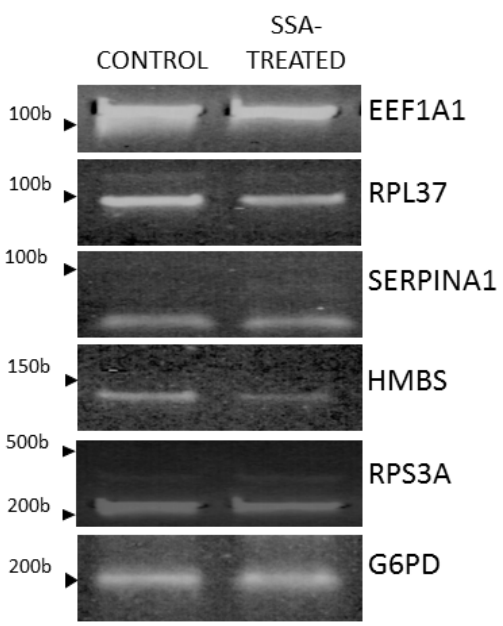
Kaida et al., 2010 and Younis et al., 2010 observed only partial inhibitory action of SSA in RT-PCR assays despite SSAs complete inhibition of splicing *in vitro* and they also show that even at high SSA concentration, splicing inhibition *in vivo* appeared to be restricted (Kaida et al., 2010, Younis et al., 2010). After preliminary RT-PCR experiments, I confirmed these observations showing that <25% of genes investigated appeared to be affected by SSA (figure 5.2.1). Many reports have been published about the effects of SSA on splicing suggesting that it is unlikely that the lack of effect on the mRNAs I assayed is due to the efficiency of SSA. Therefore we wondered if the effect of SSA on RNA splicing is masked by the presence of RNAs with long half-lives, which obscure RT-PCR analysis.

With this in mind, I intend to establish if the effect of Spliceostatin A *in vivo* can be better observed on newly generated RNA. SSA treated cells will be subjected to treatment with Interferon α in order to induce gene expression and then isolated nascent RNAs will be analysed to attempt to observe the total effect of SSA on RNA. This will help establish a working assay for potentially important therapeutic anti-cancer treatments and provide insight into the function of splicing components and their effects on RNA synthesis.

5.2 RESULTS

5.2.1 The effect of Spliceostatin A is obscured by existing stable RNAs

We obtained small amounts of SSA in collaboration with Professor Minoru Yoshida and the concentration of SSA used (200 ng/ml) was based upon a previous titration and current literature (Kaida et al., 2007, Roybal and Jurica, 2010, Corrionero et al., 2011). Microarray analysis shows that SSA did not have an effect on a large subset of genes *in vivo* (Kaida et al., 2010). I verified this observation, noting that <25% of genes investigated appeared to be affected by SSA (figure 5.2.1).



Genes affected by SSA	Genes with no detectible affect by SSA
FANCA ^{k562}	IDH1 ^{k562}
Bactin ^{k562}	Sympk ^{k562}
RPL3 ^{k562}	PPIA ^{HEK293}
P27 ^{HeLa}	IGFBP3 ^{HT1080}
RIOK3 ^{HeLa}	FGFR1 ^{HEK293}
BRD2 ^{HeLa}	PPP2R5D ^{HEK293}
	NNMT ^{HEK293}
	TMSB4X ^{HEK293}
	SMARCA4 ^{HEK293}
	PCBP2 ^{HEK293}
	COLIA1 ^{HEK293}
	ARGHDIA ^{HEK293}
	OAF ^{HEK293}
	FFG ^{HEK293}
	DNAJB1 ^{HeLa}
	EEF1A1 ^{HT1080}
	RPL37 ^{HT1080}
	SERPINA1 ^{HT1080}
	HMBS ^{HT1080}
	RPS3A ^{HT1080}
	G6PD ^{HT1080}

Figure 5.2.1 SSA shows little visible effect on RNA by RT-PCR analysis. A) PCR of expressed genes in HT1080 cells (control) and SSA-treated HT1080 cells. The gel shows that gene expression does not change with the addition of SSA in total RNA samples. B) Table of genes which have been investigated by RT-PCR for effects of treatment with SSA (the cell lines used to investigate the different genes are stated in superscript).

To establish if the effects of SSA were masked by existing RNAs *in vivo,* RT-PCR reactions were performed on newly synthesised RNA transcripts. To induce the synthesis of RNA transcripts, cells were treated with Interferon α (IFNα). IFNα is a cytokine that specifically induces the expression of a large number of genes (Der et al., 1998) (supplementary table S3). The primers for RT-PCR were designed in two adjacent exons under the hypothesis that the addition of SSA would increase the unspliced intron-containing transcripts. All RT-PCR reactions have been repeated and sequenced to verify the correct transcript.

Figure 5.2.2 illustrates a number of RT-PCR reactions performed on cDNA generated from HT1080 cells treated with a combination of IFNα, SSA and Cycloheximide. Cycloheximide inhibits nonsense-mediated mRNA decay (NMD), which is a method of post-transcriptional regulation of irregular mRNA. NMD inhibitors such as Cycloheximide may allow detection of mRNA splicing that has been inhibited but whose transcript could not be detected because of premature degradation due to NMD. Figure 5.2.2 shows that 7/8 of the gene transcripts analysed following treatment with IFNα showed some response to SSA. Figure 5.2.2 shows three examples; RING4, TRIM21 and P48 where IFNα has induced increased expression of the spliced mRNA transcript. Upon the addition of SSA, (figure 5.2.2, lanes C-E) the levels of the spliced transcripts (lower band) are reduced and the levels of unspliced mRNA transcripts are increased (higher band). Only the RT-PCR of RING4 illustrates the value of the addition of Cycloheximide, where the unspliced transcript can be seen clearly in the Cycloheximide treated sample (fig 5.2.2E) indicating that NMD has been inhibited, preventing degradation of the transcript. RT-PCR of TRIM21 shows

that the expression of the spliced transcript is reduced with the addition of SSA, whereas the RT-PCR of P48 shows that the amount of spliced transcript was nearly negligible. Furthermore SSA showed a negative effect on the expression of 4/8 of the genes induced by IFNα (MxA, ISGI5, 6-16 and OAS1) (fig 5.2.2). The 9-27 gene (fig 5.2.2) displayed a reduced response to SSA compared to the other genes tested; however a reduced response to IFNα could also be seen.

Figure 5.2.3 shows that the inhibition of MxA, 6-16 and OAS1 expression is due to splicing inhibition. When observing different exon/intron boundaries of these genes it becomes clear that SSA also inhibited their splicing. Presumably unspliced transcripts are difficult to detect by RT-PCR due to NMD which may not be efficiently inhibited by Cycloheximide (fig 5.2.3).

The effects of SSA and inhibiting splicing on RNA transcripts induced by IFNα could be reproduced with a different splicing inhibitor, TG003 (fig 5.2.4), confirming that the results were due to splicing inhibition (fig 5.2.4).

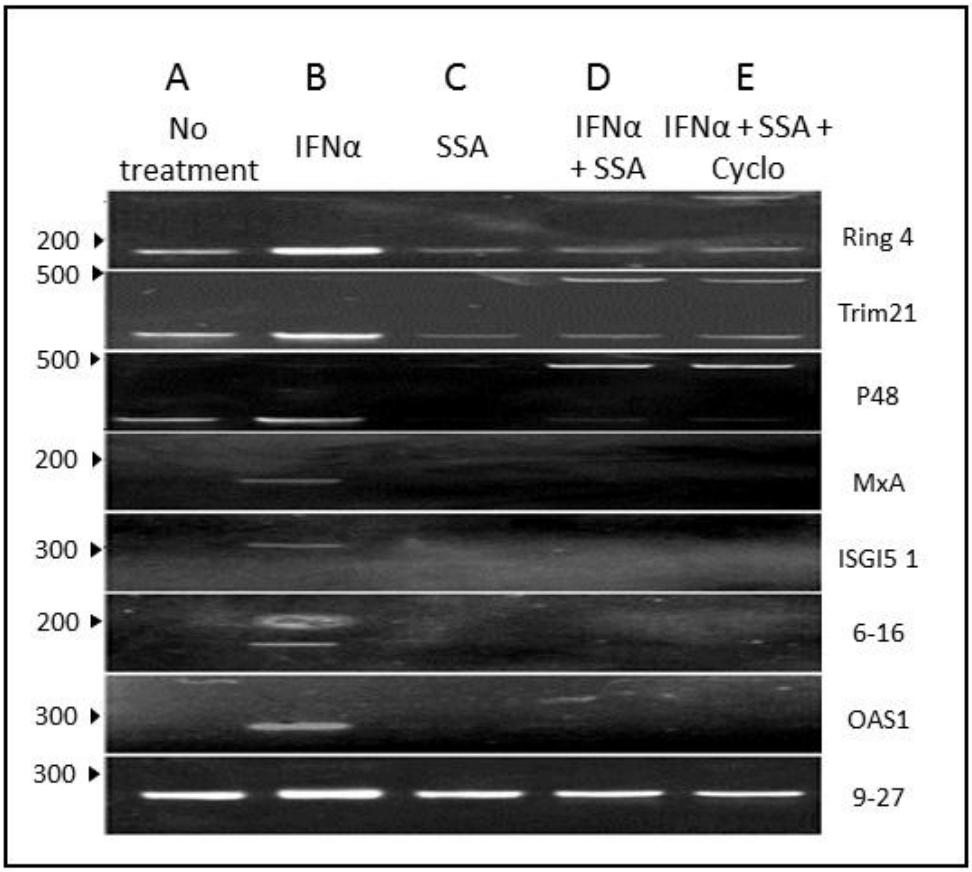


Figure 5.2.2 Spliceostatin A has a visible effect on genes induced by IFNα. cDNA synthesis and RT-PCR of genes known to be induced by IFNα in human HT1080 cells under varied conditions A) no treatment B) treated with IFNα (2 µg/ml) for 2 hours C) treated with SSA (2 ng/ml) for 2.5 hours D) pre-treated with SSA (2 ng/ml) for 30mins and then incubated with IFNα (2 µg/ml) for 2 hours E) pre-treated with SSA (2 ng/ml) for 30mins and then incubated with IFNα (2 µg/ml) and cyclohexamide (1 µg/ml) for 2 hours.

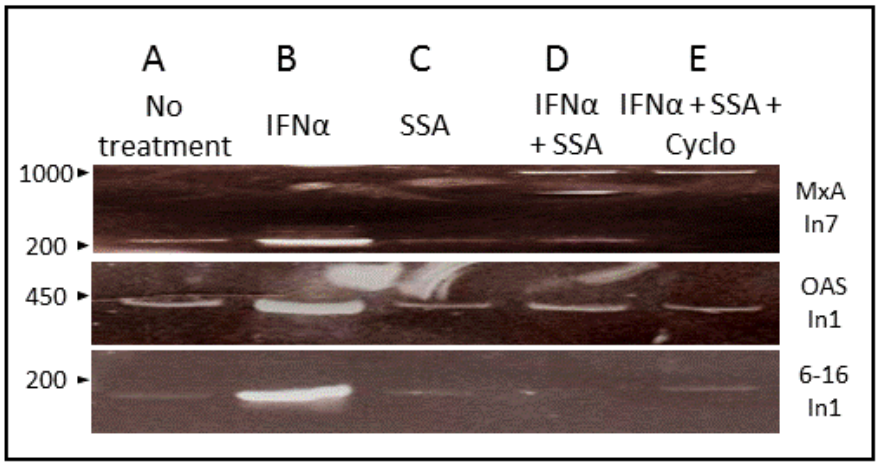


Figure 5.2.3 Detection of splicing inhibition for MxA, OAS1 and 6-16 genes. See Fig 5.2.2 A) no treatment B) treated with IFNα (2 µg/ml) for 2 hours C) treated with SSA (2 ng/ml) for 2.5 hours D) pre-treated with SSA (2 ng/ml) for 30mins and then incubated with IFNα (2 µg/ml) for 2 hours E) pre-treated with SSA (2 ng/ml) for 30mins and then incubated with IFNα (2 µg/ml) and cyclohexamide (1 µg/ml) for 2 hours. MxA In 7 – refers to RT-PCR of exons flanking the 7th intron from the 5' UTR.

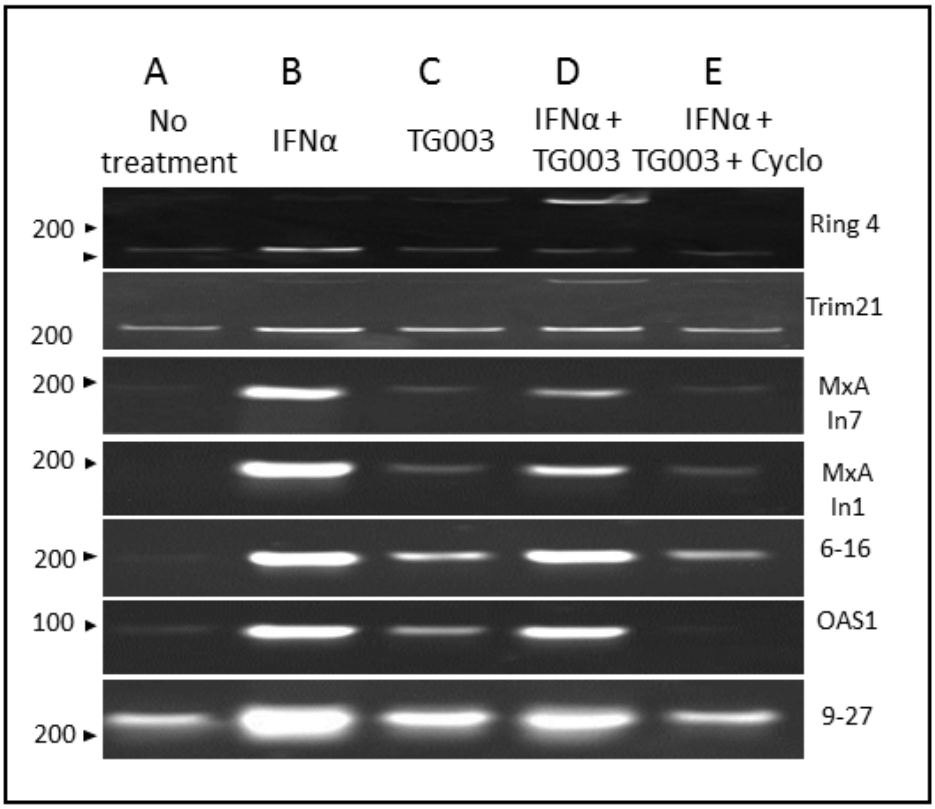


Figure 5.2.4 Alternative splicing inhibitor TG003 has a similar effect on RNA splicing as Spliceostatin A. RT-PCR of genes known to be induced by IFNα in human HT1080 cells under varied conditions A) no treatment B) treated with IFNα (2 μg/ml) for 2 hours C) treated with TG003 (2 ng/ml) for 2.5 hours D) pre-treated with TG003 (2 ng/ml) for 30mins and then incubated with IFNα (2 μg/ml) for 2 hours E) pre-treated with TG003 (2 ng/ml) for 30mins and then incubated with IFNα (2 μg/ml) and cyclohexamide (1 μg/ml) for 2 hours

5.2.2 Effects of Spliceostatin A on nascent RNAs

Although IFNα is able to induce the expression of a variety of genes, this technique did not allow us to have a global view of the effect of SSA on all RNA transcripts. Consequently we decided to establish a method for isolating and analysing newly synthesised (nascent) mRNA.

Nascent RNA was isolated by labelling and selectively enriching nascent mRNA *in vivo* by incorporation of exogenous 4-thiouridine (4sU). 4sU is a naturally occurring uridine derivative which allows isolation of nascent RNA transcripts with minimal interference to cell growth and gene expression (Windhager et al., 2012, Dolken et al., 2008, Kenzelmann et al., 2007). Following 4sU-labelling, total cellular RNA undergoes thiol-specific biotinylation which can then be separated into "tagged" (nascent transcripts) and "untagged" (pre-existing) RNAs using μMACs streptavidin-coated magnetic beads and columns (as described in Dolken et al., 2008). In order to ensure that RNA integrity was maintained throughout the protocol, 1% agarose gels were run after biotinylation and biotin affinity purification (figure 5.2.5a and 5.2.5c). Figure 5.2.6a shows that after biotinylation, samples were run on a gel to check that RNA had not degraded; this can be established by the presence of distinct 28S and 18S rRNA bands. To check that biotinylation of samples was thiol-specific, I performed a dot blot assay (fig 5.2.5b). This assay required blotting 10 fold dilutions of RNA in order to detect biotin incorporation by using streptavidin-horseradish peroxidase solution followed by x-ray exposure with ECL reagent. Figure 5.2.5b shows that biotin can only be detected in the positive control and 4sU incorporated samples.

The small amount of exposure seen in the negative control is background signal and is common when performing this type of dot blot (Lars Dölken, University of Cambridge, personal communication). The gel run after biotin affinity purification verifies that only samples with incorporated 4sU were pulled down and the integrity of the RNA is maintained (obvious by the 28S and 18S rRNA) (fig 5.2.5c). The nascent RNA was then recovered and analysed for effects of SSA on splicing via RT-PCR (Figure 5.2.6).

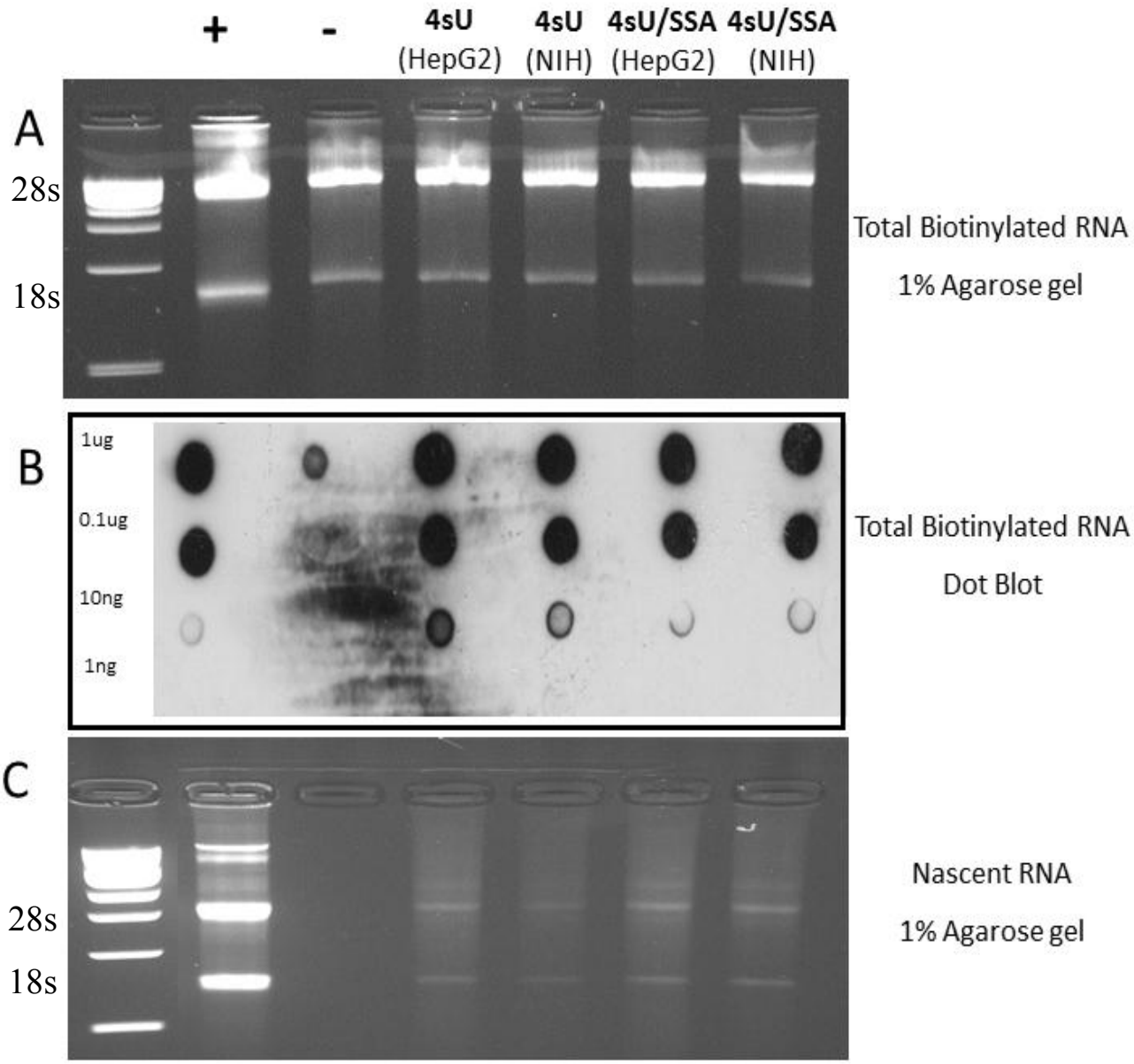
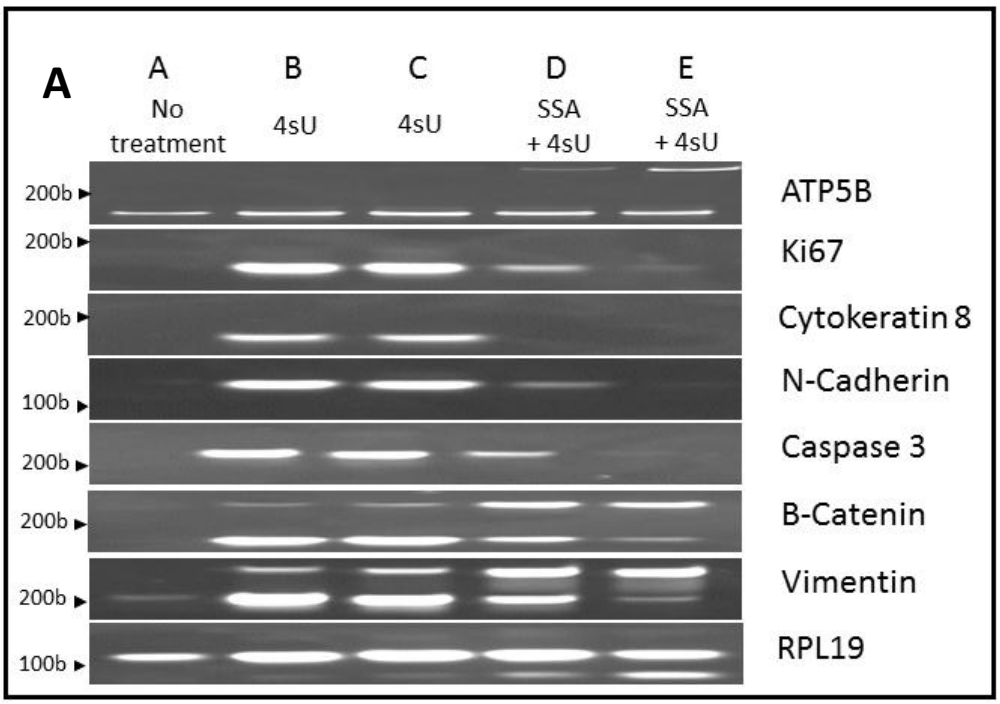


Figure 5.2.5 Detection and integrity analysis of biotinylated and biotin affinity purified RNA. A) 1% agarose of total RNA after biotinylation of RNA with no 4sU ("-") and "4sU" RNA treated with 200 µM 4sU; "4sU SSA"; RNA treated with SSA (2 ng/ml) and 200 µM 4sU. B) Dot blot assay showing the thiol-specific biotinylation of RNA. 10 fold dilutions of RNA were blotted down the membrane. C) 1% agarose of RNA after biotin affinity purification, the "+" control was total RNA. The "-"control was RNA with no 4sU showing that only 4sU incorporated RNA could be purified. +: Total HepG2 RNA; -: HepG2 RNA with no 4sU incorporation; 4sU: RNA treated with 200 µM 4sU; 4sU SSA; RNA treated with SSA (2 ng/ml) and 200 µM 4sU. In C) +: 5'-biotinylated DNA oligo. RNA used in this experiment was derived from HepG2 cells and NIH 3T3 fibroblasts (NIH)

Pre-treatment with SSA resulted in a visible effect on 12/14 genes. 8 out of 14 of these genes clearly showed an appearance of the unspliced transcript in some cases accompanied by the loss of the spliced transcript (fig 5.2.6). 4/14 genes (ki67, cytokeratin 8, N-cadherin and caspase 3) showed a complete loss of the spliced transcript upon the addition of SSA (fig 5.2.6). Again this result is likely to be due to NMD rather than an effect on transcription. Lastly 2/14 genes (RPL19 and G6PD) showed no visible effect of SSA by the RT-PCR method (fig 5.2.6). An effect on G6PD can be seen when splicing is inhibited by TG003 therefore it is possible that G6PD may be less sensitive to SSA treatment. In addition, RPL19 is an abundant transcript in cells therefore some bias may be occurring in favour of isolating non-nascent transcripts which could mask the effects of SSA on RPL19.

Overall these results show that analysing the effect of SSA on nascent RNA provided an enhanced view of the effects of the drug on global RNA and it has revealed differential effects of SSA on distinct RNA transcripts.



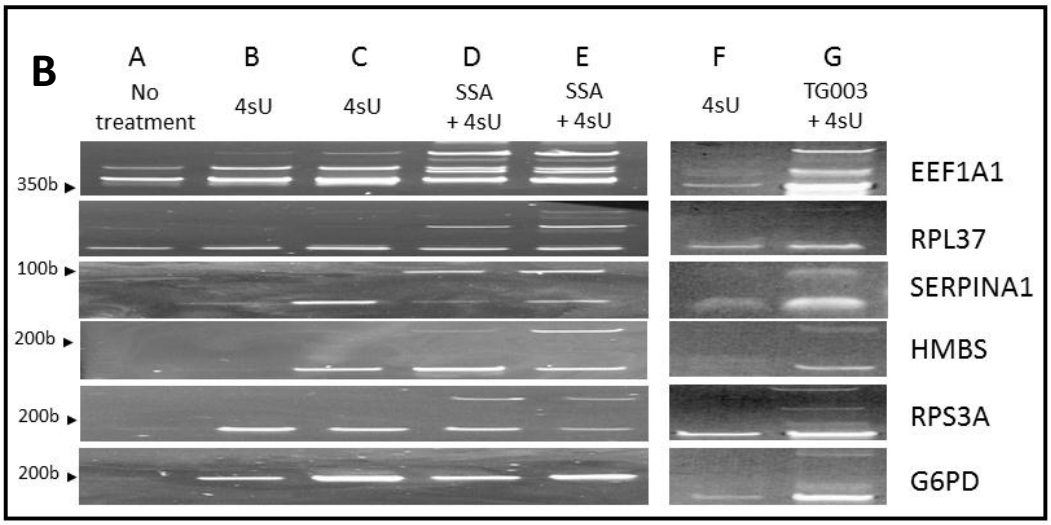


Figure 5.2.6 Spliceostatin A has a visible effect on newly synthesised RNAs in mouse NIH 3T3 fibroblasts (A) and human HepG2 (B). cDNA synthesis and RT-PCR of nascent RNAs in human HepG2 cells under varied conditions A) no treatment B & C) treated with 200 µM thiouridine (4sU) base analogue for 2 hours D & F) pre-treated with SSA (200 ng/ml) for 30mins and then incubated with 200 µM 4sU for 2 hours. E) Treated with 200 µM thiouridine (4sU) base analogue for 2 hours F) pre-treated with TG003 (2 µg/ml) for 30 mins and then incubated with 200 µM 4sU for 2 hours. All samples were then biotinylated and biotin affinity purified with streptavadin beads so that nascent RNA could be isolated (as described in Dolken et al., 2008).

5.2.3 Inhibition of the spliceosome causes retention of RNA in the nucleus

I then investigated the effects of SSA on RNA localisation and export in cells. Immunocytochemistry analysis shows that upon treatment with SSA, newly synthesised RNA labelled with 5'Flurouridine can be seen to localise largely in the nucleus (Figure 5.2.7). Both SSA and TG003 cause nuclear retention of RNAs to different degrees, with SSA being more potent and causing enhanced nuclear retention of RNAs (Figure 5.2.7 and figure 5.2.8). Samples treated with SSA show 56% colocalisation of nascent RNAs with the nucleus; TG003-treated samples show less nuclear retention with 36% and control samples show only 12% nuclear colocalisation (figure 5.2.8). This result can be seen upon titration of the drugs, proving that the difference between the effects of inhibitors is not due to differences in concentration of the drugs (figure 5.2.9). RNA retention in the nucleus could be assumed to be largely ribosomal RNAs (rRNAs) as they form ~80% of all RNA transcripts in a cell. Curiously, a majority of rRNAs are intronless therefore the effect of inhibiting the splicing machinery on these transcripts could be unprecedented. However it is possible that inhibiting the spliceosome may result in knock-on effects on other cell processes thereby having an indirect effect on intronless genes (Maniatis and Reed, 2002, Grzybowska, 2012).

Kaida et al., 2007 suggests that the SF3b subunit of the U2 snRNP, which is a target of SSA, plays a key role in nuclear retention of RNA. SF3b is known to interact with the retention and splicing (RES) complex implicating that without functional SF3b, pre-mRNAs leakage occurs. Figure 5.2.7 c & f show some leakage

of RNA from the nucleus however retention of RNA in the nucleus is much more prevalent.

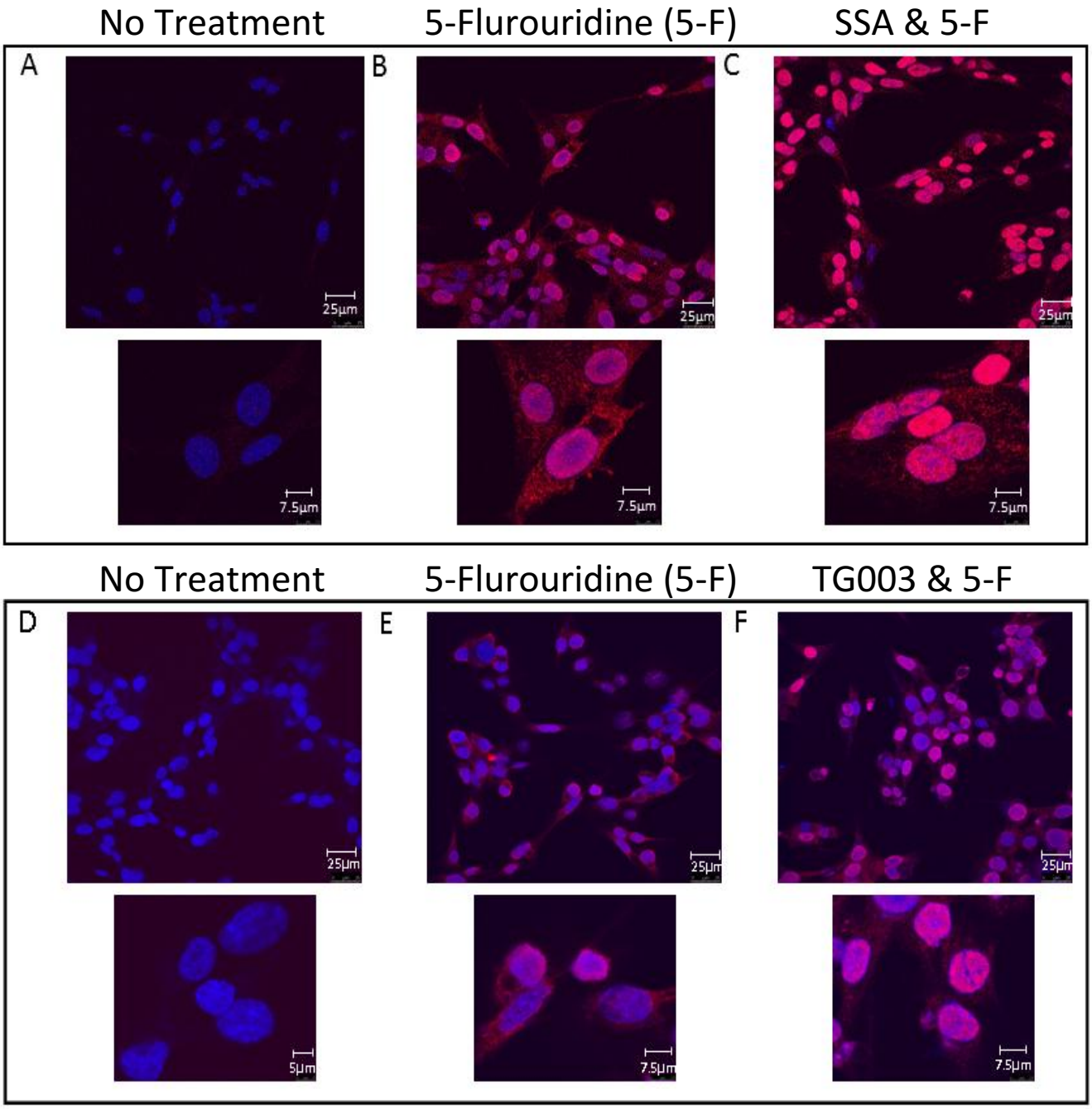


Figure 5.2.7 Spliceostatin A and TG003 splicing inhibitors cause retention of RNAs in the nucleus. A-F) Immunocytochemistry analysis of HepG2 cells A & D) No treatment B & E) Treatment with 1 mM 5-Flurouridine for 2 hours C) Pre-treatment with 200 ng/ml SSA for 30 mins and then 1 mM 5-Flurouridine for 2 hours F) Pre-treatment with 2 µg/mlTG003 for 30 mins and then 1 mM 5-Flurouridine for 2 hours. Cells were then stained by immunocytochemistry using either anti-BrdU and Alexa Fluor 555 antibody or DAPI. Photographs were taken on an SP5 confocal.

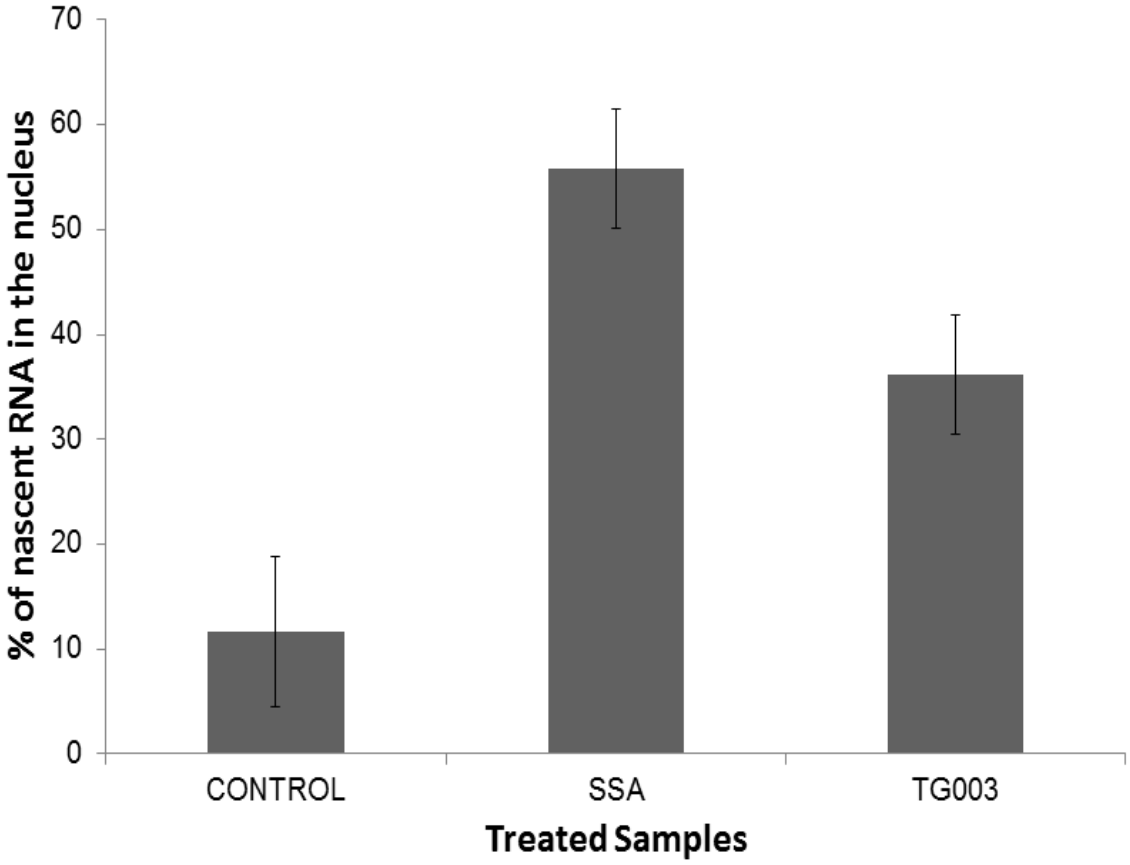


Figure 5.2.8 Quantification of immunocytochemistry analysis of HepG2 cells (figure 5.2.7). Values represent percentage of colocalisation of nuclear stain DAPI with nascent RNA (5'F anti-BrdU staining). "Control" represents samples treated with 1mM 5-Flurouridine for 2 hours. "SSA" represents SSA pre-treated samples and then 1 mM 5-Flurouridine for 2 hours. TG003 represents TG003 pre-treated samples and then 1 mM 5-Flurouridine for 2 hours. Fiji image j and JACoP software was used to obtain colocalisation values from confocal images generated in four independent experiments – error bars represent standard error of the mean.

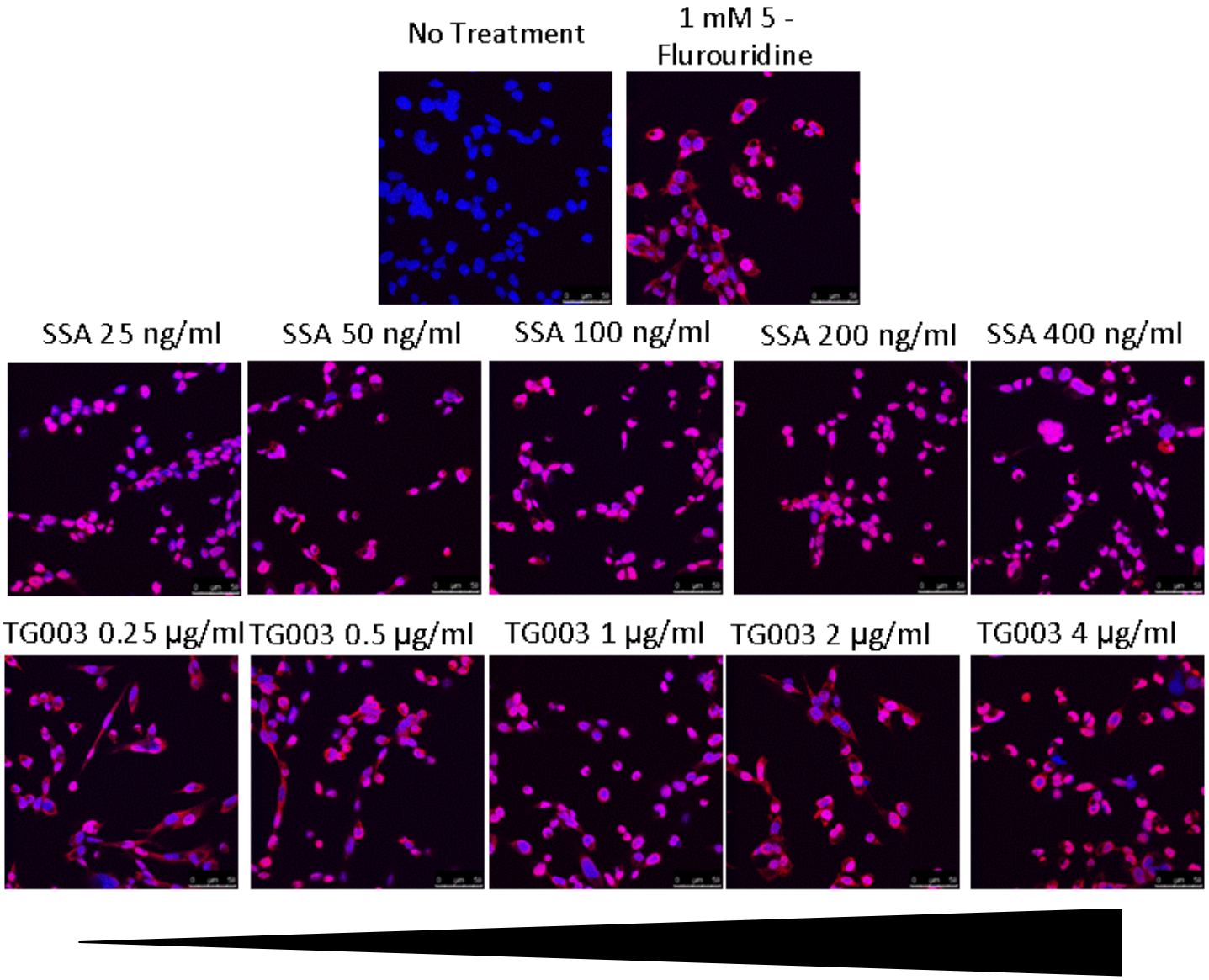


Figure 5.2.9 Titration of Spliceostatin A and TG003 concentration. Immunocytochemistry analysis of HepG2 cells, treatments of each sample is described above the picture. Varying concentration of SSA and TG003 were added to cells 30 minutes before the addition of 1 mM 5-Flurouridine for 2 hours. Cells were then stained by immunocytochemistry using either anti-BrdU and Alexa Fluor 555 antibody or DAPI. Photographs were taken on an SP5 confocal.

5.2.4 More pre-mRNA can be detected in the nucleus of SSA treated samples compared to the cytoplasm indicating possible nuclear retention of RNA

In order to elucidate the effect of SSA on nuclear RNA retention, nascent RNAs from the nucleus and the cytoplasm were isolated. Subsequently RT-PCRs of genes found commonly in HepG2 cells and rRNAs were performed (fig 5.2.10). The RT-PCR of the βactin, HMBS and RPS3A genes verified the immunocytochemistry analysis, showing a build-up and retention of RNAs in the nucleus of SSA-treated cells. Figure 5.2.10 shows that upon the addition of SSA there is a visible increase in pre-mRNA (unspliced transcripts) in the nucleus. In addition this is confirmed by a build-up of pre-mRNA also seen in the nucleus of cells treated with TG003; again confirming the immunocytochemistry it can be seen to a lesser extent than with SSA treatment (fig 5.2.10 C-F). Contrary to our initial hypothesis RT-PCR of 18S and 5.8S rRNAs indicated that neither SSA nor TG003 affected their nuclear export; therefore rRNAs may not be the major component of RNAs retained in the nucleus upon SSA treatment.

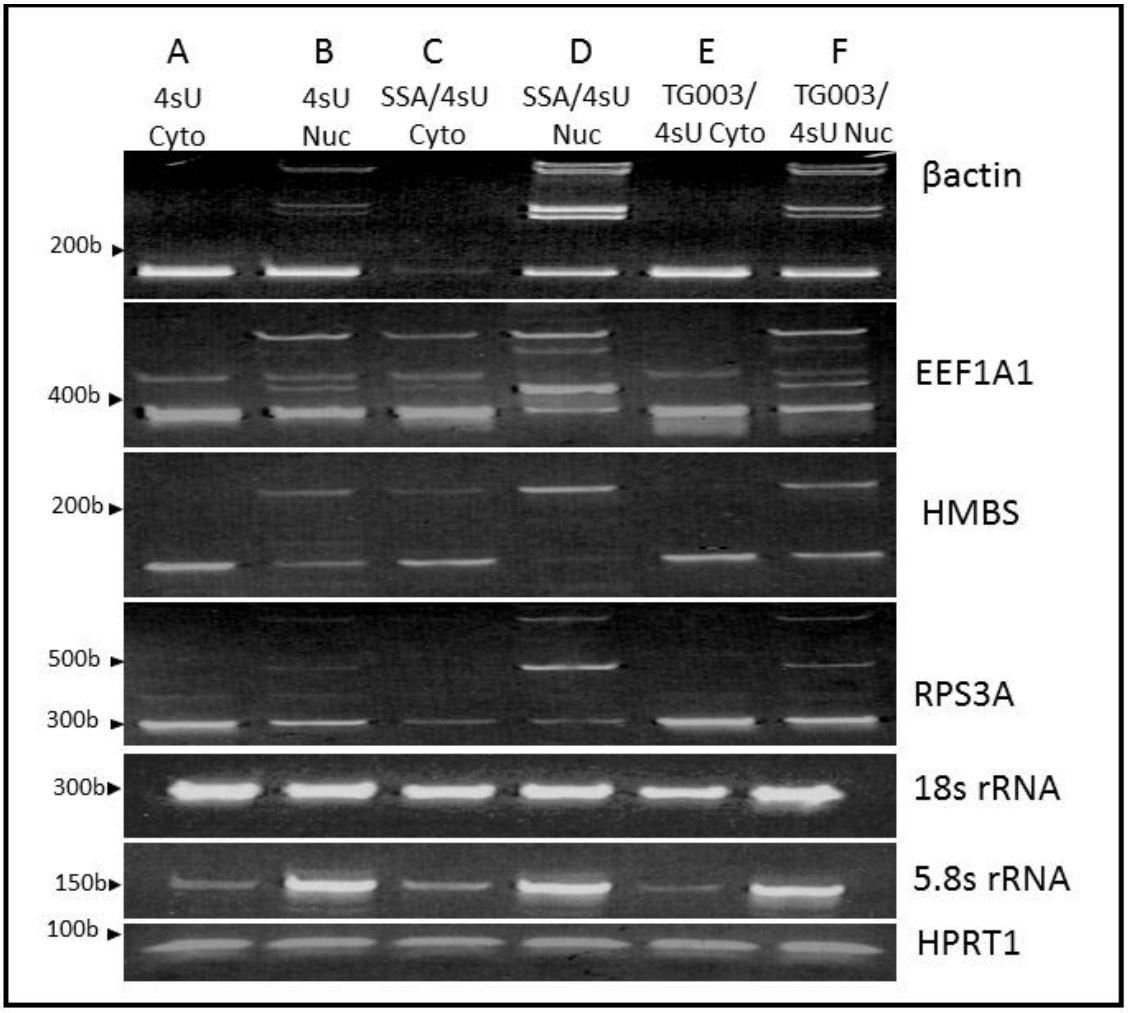


Figure 5.2.10 Spliceostatin A causes a build-up of nascent pre-mRNA in the nucleus. cDNA synthesis and RT-PCR of nascent RNA taken from cytoplasmic and nuclear portions of human HepG2 cells under varied conditions A) Cytoplasmic RNA treated with 200 µM thiouridine (4sU) base analogue for 2 hours. B) Nuclear RNA treated with 200 µM thiouridine (4sU) base analogue for 2 hours C) Cytoplasmic RNA pre-treated with SSA (200 ng/ml) for 30 mins and then incubated with 200 µM 4sU for 2 hours D) Nuclear RNA pre-treated with SSA (200 ng/ml) for 30mins and then incubated with 200 µM 4sU for 2 hours E) Cytoplasmic RNA pre-treated with TG003 (2 µg/ml) for 30mins and then incubated with 200 µM 4sU for 2 hours F) Nuclear RNA pre-treated with TG003 (2 µg/ml) for 30 mins and then incubated with 200 µM 4sU for 2 hours. Primers that span only an exon, so mature mRNA can be detected, are used as a control (HPRT1 gene). In all other samples the primers bind to adjacent exons flanking an intron.

5.3 DISCUSSION

5.3.1 The effects of splicing inhibitors, SSA and TG003 can be seen more clearly in nascent RNAs

This chapter addresses the difficulties of assessing the effects of splicing inhibitors on RNA transcripts. It was not possible to see the effects of SSA on a majority of pre-mRNAs by RT-PCR because RNA transcripts with long half-lives obscured detection of RNAs affected by SSA. The average time of RNA turnover in mammalian cells is ~ 7 hours where <0.5% of RNA transcripts have half-lives less than 1 hour (Sharova et al., 2009). A "short" half-life of RNA is defined at <4 hours so the effects of treatment with splicing inhibitors after 2 hours is unlikely to be seen (Tani et al., 2012). We have shown that by isolating nascent RNAs from cells pre-treated with SSA, it is possible to visualise the global effects of SSA. This novel method for analysing the effects of SSA will be useful for future studies on the spliceosome and the potential of SSA as therapy for cancers.

Once a working assay for SSA had been established, it became obvious that SSA had different effects on individual RNAs. This observation has been corroborated by a number other of studies (Kaida et al., 2007, Lo et al., 2007, Corrionero et al., 2011). It is thought that targeting a single aspect of the splicing process can give rise to diverse effects on different RNA transcripts and it has also been noted that some genes are more sensitive to SSA than others, inducing alternative splicing and often inhibiting splicing completely (Bonnal et al., 2012).

This can be verified *in vivo* where mutations in the SF3b1 subunit protein cause specific developmental defects in neural crest formation in zebrafish, supporting

the theory that different cell types and genes have different requirements of SF3b1 (An and Henion, 2012).

Other examples substantiating the differential effects of SSA can be seen when observing the action of SSA on yeast intron-containing transcripts; which are simple and synonymous to one another compared to human mRNA. The effects *in vitro* showed that there is an unknown mechanism by which the spliceosome can distinguish between RNA transcripts, most likely due to co-factors (Kaida et al., 2007). Knocking down core spliceosomal components in flies results in differential effects on alternative splicing (Pleiss et al., 2007). Mutations of different spliceosomal components cause different effects on RNAs and can be validated in our results as some differences in the effect of TG003 and SSA on RNA transcripts can be observed (fig 5.2.2, fig 5.2.4 and fig 5.2.6).

To conclude, a majority of nascent RNA transcripts are affected by SSA with the minority that are unaffected showing no change in expression (figure 5.2.6). Of the transcripts that were affected by SSA differential effects could be seen. A deeper understanding of the effect of different cofactors of the core splicing machinery could help achieve modulation of mis-splicing that leads to disease.

5.3.2 The effects of SSA and TG003 differ

During this study, we detected a difference between the effects of SSA and TG003 on RNA processing. When observing the action of SSA and TG003 on RNA transcripts some transcripts were affected to the same degree by both inhibitors,

some differ depending on which inhibitor is used and some transcripts show no effect with one or both inhibitors. A majority of these results could be repeated confirming the role of each drug in causing different effects on RNAs. In addition immunocytochemistry analysis revealed that although both SSA and TG003 caused nuclear retention of RNA, it occurred to significantly different degrees with SSA being 1.5 times more potent.

Pleiss et al 2007 investigated 24 different mutations of the spliceosome core components and mRNA processing in flies to observe their effect on intron removal in mRNA on a global scale. They discovered that splicing mutations have an effect on a broad range of transcripts resulting in accumulation of unspliced pre-mRNA; however not all transcripts followed the same pattern. Transcripts can be affected equally by all the mutations, some are affected to differing degrees and some transcripts that are only affected by specific mutations (Pleiss et al., 2007). This has implications that individual splicing components have different effects depending on the RNA substrate.

Furthermore TG003 splicing inhibitor targets SR proteins presumably having an effect on both U2- and U12-dependent spliceosomes whereas SSA is thought to target the SF3b subunit of the U2 snRNP that specifically inhibits U2-dependent spliceosome.

5.3.3 Splicing inhibition causes nuclear retention of RNA transcripts

Immunofluorescence studies show the effect of the splicing inhibitors on localisation of RNAs. RNA retention can be can be detected in cells treated with either SSA or TG003 to varying degrees; SSA-treated samples showed 1.5 times more RNA within the nucleus than TG003-treated samples. In SSA treated cells there was 56% colocalisation of nascent RNAs in the nucleus compared to control samples which show 12% colocalisation (Figure 5.2.8). This shows high levels of nuclear RNA retention in SSA-treated samples. Current literature similarly report accumulation of multiple transcripts with partial intron retention upon treatment with SSA (Bonnal et al., 2012).

Accumulated pre-mRNAs after splicing inhibition can be very stable despite containing termination signals in introns which would normally result in degradation by the NMD pathway, therefore nuclear accumulation of unspliced RNAs is possible (Kaida et al., 2007). However, reports from Kaida et al., 2007 show that the SF3b subunit has a dual function in the splicing and nuclear retention of pre-mRNA, therefore inhibition of this component actually causes pre-mRNA leakage. Furthermore they show accumulation of unspliced proteins upon SSA treatment, bypassing all checks of unspliced RNA including NMD (Kaida et al., 2007). The class of RNA that constitute the majority of RNAs retained in the nucleus upon treatment with SSA is still not known and initial hypotheses alluding to retention of intronless ribosomal RNAs have not yet been corroborated by RT-PCR.

CHAPTER 6 - GENERAL DISCUSSION

6.1 Post-transcriptional regulation of microRNAs in ESCs

Our main focus was to characterise the mechanism behind the differences in global miRNA expression in ESCs and differentiated NPSCs. We discovered that the majority of miRNAs (~60%) investigated were regulated by transcription. Of the remaining ~40% of miRNAs, all of these were post-transcriptionally regulated in ESCs. Our data presents a lack of correlation between primary and mature miRNA expression levels in ESCs confirming that Drosha often regulates or blocks miRNA expression in early developmental stages. The prevalence of post-transcriptional miRNA regulation observed in ESCs has been previously reported in mouse embryonic development (Thomson et al., 2006).

It is possible that observed transcription of primary miRNAs that are not processed is a potential hallmark of ESCs, which contributes to their plasticity and allows for rapid and discrete expression of crucial miRNAs during differentiation. This idea is supported when investigating expression of primary miRNAs located on chromosome 1 and 2 (fig 3.2.7). Sixteen out of 37 miRNAs analysed were actively transcribed and turned-over in ESCs however further processing into mature miRNAs is inhibited in ESCs.

The potential mechanisms behind extensive post-transcriptional miRNA regulation in ESCs include regulation by RNA binding proteins (RNA-bps) that are specifically expressed in ESCs e.g. Lin28. In future we aim to establish if RNA binding proteins regulate those miRNAs that we discovered to be post-transcriptionally regulated (Chapter 3).

MicroRNA deep sequencing has contributed to the discovery of huge numbers of miRNAs. Out of 2578 human miRNAs recorded in miRBASE, only 300 miRNAs are highly sequenced with more than 1000 reads per million (RPM) (miRBase and miRGator). This suggests that only 19% of existing miRNAs are likely to have functional biological role. With this data we are able to infer that the highly expressed miRNAs which are post-transcriptionally regulated in ESCs are likely to have a biological role, potentially in early development.

6.2 Alternative polyadenylation of miRNA clusters

Our investigation of miRNA regulation led to the discovery of alternatively polyadenylated primary miRNA transcripts. Primary miR-let-7a/f transcripts are alternatively polyadenylated between ESCs and NPSCs, which could play a role in the differential expression of mature miRNA (fig 4.2.4). Further study showed that each alternatively polyadenylated transcript was able to undergo Drosha processing into precursor miRNA however this occurred in the absence of Lin28 (fig 4.2.5).

The effect of APA can be further investigated by generating primary transcripts with mutated PASs rendering them unrecognisable by the cleavage and polyadenylation specificity factor (CPSF). Altered primary miRNA transcripts could then be transfected into cell lines and the resultant mature miRNA expression can be quantitated. We hypothesise that in the case of miR-let-7a/f, mutating the PAS positioned between pre-miR-let-7a and pre-miR-let-7f would result in decreased production of both mature miRNAs.

Post-transcriptional regulation of miRNAs is very apparent in miRNA clusters; although clusters are transcribed together individual mature miRNA expression can be different within a specific tissue type or between tissues. Recent evidence points to the role of tertiary RNA structure as a scaffold for accessory RNA-bps which can cause structural remodelling, exposing new binding sites/motifs (Chakraborty et al., 2012, Chaulk et al., 2011). In mice, 3' truncation of the pri-miR 17-92 can lead to an overall increase in mature miRNA expression (He et al., 2005) which was later shown to be caused by structural disruption of the cluster (Chaulk et al., 2011). In addition, Chakraborty et al. (2012) showed that changing the order of miRNA hairpins in the pri-miR-20a-19a cluster results in impaired tertiary structure. This leaves the pri-miR-20a-19a cluster exposed to RNA-bp, MPC, resulting in altered processing efficiency. APA could work to alter tertiary structures of primary transcripts potentially revealing or excluding binding sites for RNA-bps that can promote or impede miRNA processing.

Chakraborty et al. (2012) show that organisation of the tertiary structure of primary miRNAs may be the first point of post-transcriptional regulation. We show that clustered miRNA transcripts undergo extensive alternative processing which has the potential to contribute to altering RNA tertiary structure thereby influencing RNA-bp-mediated regulation of miRNA processing.

6.3 Effect of Spliceostatin A on nascent RNAs

Like previous reports we found it difficult to assess the effects of splicing inhibitors because RNA transcripts with long half-lives obscured detection of

RNAs affected by SSA. We overcame this by isolating nascent RNAs from cells allowing us to visualise global effects of SSA. This novel method for analysing the effects of SSA will be useful for future studies on splicing inhibitors.

Nuclear and cytoplasmic nascent RNAs treated with SSA and TG003 are currently undergoing mass RNA sequencing and analysis (in collaboration with Dr Rosenstiel at IKMB, University of Kiel). We hope that this sequencing data will help us uncover which RNA transcripts are retained in the nucleus under treatment. We would predict that inhibiting the spliceosome will have an effect on rRNAs and other intronless genes, however we are not certain if this would be due to a direct effect of inhibiting the spliceosome or an indirect effect on transcription and export. We will be able to analyse the effects of the two drugs and any differences might give insight into the function of the SF3b subunit of the spliceosome and the role of CLK1 and CLK7 in splicing which are inhibited by SSA and TG003. Furthermore, identifying RNA transcripts that are not affected by the splicing inhibitors could lead to further studies into the role of gene sequence and associated binding proteins on spliceosome recognition. We will also establish if SSA can affect genes with U12-type introns as proteins of the SF3b complex is present in both spliceosomes (Turunen et al., 2012). In fact SSA action and binding to the different subcomponents of SF3b has not yet been established and perhaps any similarities with the U12-dependant spliceosome could help clarify SSA binding.

The spliceosome is present in *Saccharomyces cerevisiae* spp, despite less than 5% of their genome containing introns (Parenteau et al., 2008). In addition deletion

of many of these introns proved they were not essential for growth (Parenteau et al., 2008). This implies that the spliceosome may have other potential functions in addition to splicing, such as mRNA export. Splicing inhibitors are a useful tool to elucidate any ulterior roles of the spliceosome, particularly in organisms where conditional splicing mutations are not available (Pleiss et al., 2007). Understanding the role of individual subunits of the macromolecular spliceosome will help uncover the exact role the spliceosome plays in the transcriptome; including effects on transcription, mRNA export, miRNA biogenesis and expression.

6.4 Splicing and microRNAs

There are many aspects of splicing and miRNA biogenesis that are physiologically linked. The fact that a majority of human miRNAs are located in the introns of other genes implies coordinated function of the spliceosome and the microprocessor (Rodriguez et al., 2004). Intronic pre-miRNA processing is known to occur prior to splicing catalysis (Kim and Kim 2006) and recently it has been shown that pre-miRNA transcripts can be immunoprecipitated with the spliceosome and Drosha can be immunoprecipitated with splicing intermediate introns (Kataoka et al., 2009).

There is a strong evolutionary connection between the RNA splicing machinery and miRNA mediated gene silencing (Dumesic et al., 2013, Tabach et al., 2013) including a correlation between the expression of miRNA components and the number of introns within a species (Tabach et al., 2013). Based on evolutionary

analysis it is also thought that a number of proteins involved in intron removal are orthologues of proteins involved in miRNA processing (Tabach et al., 2013).

As the microprocessing machinery and the spliceosome are linked in function and share components, inhibiting the spliceosome might cause inhibition of miRNA processing, potentially resulting in the rise of pri-miRNA expression and a decline in mature miRNAs (Dumesic et al., 2013). Janas et al., 2011 show that knockdown of the U1 splicing factors can in fact reduce intronic miRNA expression, again substantiating the link between splicing and miRNAs. A direct consequence of inhibition of the spliceosome will also be seen in mirtrons which are directly spliced by the splicing machinery, bypassing Drosha processing. In addition we observed alternative splicing of the miR-302 cluster, the consequence for which is unknown. It is possible that treatment of ESCs with SSA could reveal a potential change in miR-302 expression as a result of inhibiting alternative splicing of the transcript.

Advances in sequencing technologies have led to the discovery of an abundance of circular RNAs (ciRNA) that arise by splicing of an exon which joins its 5' and 3' ends to form a lariat. CDR1as is a ciRNA that is known to target miR-7 and act as a miRNA sponge causing knockdown of miRNA expression that is 10 times greater than any known transcript (Memczak et al., 2013). It is thought that the role of ciRNAs could go beyond binding just miRNAs, with potential to also act on RNA-bps. Inhibiting splicing may result in a reduction of ciRNAs which would directly affect miRNA activity (Hansen et al., 2013, Memczak et al., 2013).

Alteration of tertiary structure by alternative polyadenylation or otherwise can have a direct impact on the spliceosome (Chakraborty et al., 2012). As previously discussed, the spliceosome stalls at introns containing miRNAs to allow time for RNA to fold (Chakraborty et al., 2012). MPC is an RNA-bp that binds to the tertiary structure of pre-20a-19a in mice and inhibits its expression (Chakraborty et al., 2012). Chakraborty et al., 2012 show that MPC recruitment is linked to the spliceosome therefore it is possible that altered tertiary structure could lead to increased binding of MPC leading to increased splicing thereby impeding pri-miRNA processing. It would be interesting to investigate the effect of inhibiting splicing on miRNA targeting; with the inclusion of introns it is possible that many miRNAs are able to target the intronic sequences for transcript degradation, potentially preventing the production of aberrant proteins.

6.5 Conclusion

We established a novel method to distinguish primary miRNA expression from confounding overlapping genes in ESCs and NPSCs. Using the method of 3'RACE we were able to identify a significant number of primary miRNAs that were inhibited from further processing in ESCs. Furthermore we revealed that miRNA gene clusters generated several primary transcripts by alternative polyadenylation. This may serve to regulate individual miRNAs independently from the rest of the cluster.

MicroRNAs have a huge role in early development across many species therefore post-transcriptional regulation of miRNAs is likely to be an important biological mechanism in human ESCs (Lee et al., 1993, Olsen and Ambros, 1999, Aravin et al., 2003, Bernstein et al., 2003). Subsequently further research into the control of miRNA expression promises a greater understanding of how the differentiation of stem cells can be regulated for the purpose of regenerative medicine.

Human ESCs and cancer cell types are known to share miRNA expression profiles adding to their repertoire of other shared properties namely, self-renewal and a block in differentiation (Kim and Orkin, 2011, Thomson et al., 2006). As a result, further study into the regulation of ESC-specific miRNAs may prove useful in treating disease.

Finally, we implemented a novel method for isolating nascent RNAs in order to establish an improved assay of the splicing inhibitor, SSA. This is likely to contribute to our understanding of splicing mechanisms and regulation, with potential for aiding the discovery of novel splicing drugs for cancer therapy.

References

ALBERTS, B. (2002) *Molecular biology of the cell,* New York, Garland Science.

ALTUVIA, Y., LANDGRAF, P., LITHWICK, G., ELEFANT, N., PFEFFER, S., ARAVIN, A., BROWNSTEIN, M. J., TUSCHL, T. & MARGALIT, H. (2005) Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res,* 33**,** 2697-706.

AMBROS, V. (2004) The functions of animal microRNAs. *Nature,* 431**,** 350-5.

AMBS, S., PRUEITT, R. L., YI, M., HUDSON, R. S., HOWE, T. M., PETROCCA, F., WALLACE, T. A., LIU, C. G., VOLINIA, S., CALIN, G. A., YFANTIS, H. G., STEPHENS, R. M. & CROCE, C. M. (2008) Genomic profiling of microRNA and messenger RNA reveals deregulated microRNA expression in prostate cancer. *Cancer Res,* 68**,** 6162-70.

AN, M. & HENION, P. D. (2012) The zebrafish sf3b1b460 mutant reveals differential requirements for the sf3b1 pre-mRNA processing gene during neural crest development. *Int J Dev Biol,* 56**,** 223-37.

ANOKYE-DANSO, F., TRIVEDI, C. M., JUHR, D., GUPTA, M., CUI, Z., TIAN, Y., ZHANG, Y., YANG, W., GRUBER, P. J., EPSTEIN, J. A. & MORRISEY, E. E. (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell,* 8**,** 376-88.

ARAVIN, A. A., LAGOS-QUINTANA, M., YALCIN, A., ZAVOLAN, M., MARKS, D., SNYDER, B., GAASTERLAND, T., MEYER, J. & TUSCHL, T. (2003) The small RNA profile during Drosophila melanogaster development. *Dev Cell,* 5**,** 337-50.

BABIARZ, J. E. & BLELLOCH, R. (2008) Small RNAs - their biogenesis, regulation and function in embryonic stem cells.

BAGGA, S., BRACHT, J., HUNTER, S., MASSIRER, K., HOLTZ, J., EACHUS, R. & PASQUINELLI, A. E. (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell,* 122**,** 553-63.

BALLARINO, M., PAGANO, F., GIRARDI, E., MORLANDO, M., CACCHIARELLI, D., MARCHIONI, M., PROUDFOOT, N. J. & BOZZONI, I. (2009) Coupled RNA processing and transcription of intergenic primary microRNAs. *Mol Cell Biol,* 29**,** 5632-8.

BARBATO, C., ARISI, I., FRIZZO, M. E., BRANDI, R., DA SACCO, L. & MASOTTI, A. (2009) Computational challenges in miRNA target predictions: to be or not to be a true target? *J Biomed Biotechnol,* 2009**,** 803069.

BARROSO-DELJESUS, A., ROMERO-LOPEZ, C., LUCENA-AGUILAR, G., MELEN, G. J., SANCHEZ, L., LIGERO, G., BERZAL-HERRANZ, A. & MENENDEZ, P. (2008) Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. *Mol Cell Biol,* 28**,** 6609-19.

BAULCOMBE, D. (2003) Overview of RNA interference and related processes. *Curr Protoc Mol Biol,* Chapter 26**,** Unit 26 1.

BECKER, K. A., GHULE, P. N., THERRIEN, J. A., LIAN, J. B., STEIN, J. L., VAN WIJNEN, A. J. & STEIN, G. S. (2006) Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J Cell Physiol,* 209**,** 883-93.

BERARDI, E., PUES, M., THORREZ, L. & SAMPAOLESI, M. (2012) miRNAs in ESC differentiation. *Am J Physiol Heart Circ Physiol,* 303**,** H931-9.

BEREZIKOV, E. (2011) Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet,* 12**,** 846-60.

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.

BEYER, A. L. & OSHEIM, Y. N. (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev,* 2**,** 754-65.

BOLDRUP, L., COATES, P. J., WAHLGREN, M., LAURELL, G. & NYLANDER, K. (2012) Subsite-based alterations in miR-21, miR-125b, and miR-203 in squamous cell carcinoma of the oral cavity and correlation to important target proteins. *J Carcinog,* 11**,** 18.

BONNAL, S., VIGEVANI, L. & VALCARCEL, J. (2012) The spliceosome as a target of novel antitumour drugs. *Nat Rev Drug Discov,* 11**,** 847-59.

BOUTET, S. C., CHEUNG, T. H., QUACH, N. L., LIU, L., PRESCOTT, S. L., EDALATI, A., IORI, K. & RANDO, T. A. (2012) Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell Stem Cell,* 10**,** 327-36.

BRAY, I., TIVNAN, A., BRYAN, K., FOLEY, N. H., WATTERS, K. M., TRACEY, L., DAVIDOFF, A. M. & STALLINGS, R. L. (2011) MicroRNA-542-5p as a novel tumor suppressor in neuroblastoma. *Cancer Lett,* 303**,** 56-64.

BRODY, Y. & SHAV-TAL, Y. (2011) Transcription and splicing: when the twain meet. *Transcription,* 2**,** 216-20.

BUENO, M. J., PEREZ DE CASTRO, I., GOMEZ DE CEDRON, M., SANTOS, J., CALIN, G. A., CIGUDOSA, J. C., CROCE, C. M., FERNANDEZ-PIQUERAS, J. & MALUMBRES, M. (2008) Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. *Cancer Cell,* 13**,** 496-506.

BUVOLI, M., BUVOLI, A. & LEINWAND, L. A. (2007) Interplay between exonic splicing enhancers, mRNA processing, and mRNA surveillance in the dystrophic Mdx mouse. *PLoS One,* 2**,** e427.

CACERES, J. F., STAMM, S., HELFMAN, D. M. & KRAINER, A. R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science,* 265**,** 1706-9.

CALIN, G. A., FERRACIN, M., CIMMINO, A., DI LEVA, G., SHIMIZU, M., WOJCIK, S. E., IORIO, M. V., VISONE, R., SEVER, N. I., FABBRI, M., IULIANO, R., PALUMBO, T., PICHIORRI, F., ROLDO, C., GARZON, R., SEVIGNANI, C., RASSENTI, L., ALDER, H., VOLINIA, S., LIU, C. G., KIPPS, T. J., NEGRINI, M. & CROCE, C. M. (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med,* 353**,** 1793-801.

CARD, D. A., HEBBAR, P. B., LI, L., TROTTER, K. W., KOMATSU, Y., MISHINA, Y. & ARCHER, T. K. (2008) Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol Cell Biol,* 28**,** 6426-38.

CASTELLANO, L., GIAMAS, G., JACOB, J., COOMBES, R. C., LUCCHESI, W., THIRUCHELVAM, P., BARTON, G., JIAO, L. R., WAIT, R., WAXMAN, J., HANNON, G. J. & STEBBING, J. (2009) The estrogen receptor-alpha-induced microRNA signature regulates itself and its transcriptional response. *Proc Natl Acad Sci U S A,* 106**,** 15732-7.

CHAKRABORTY, S., MEHTAB, S., PATWARDHAN, A. & KRISHNAN, Y. (2012) Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing. *Rna,* 18**,** 1014-28.

CHAPMAN, E. J. & CARRINGTON, J. C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet,* 8**,** 884-96.

CHAULK, S. G., THEDE, G. L., KENT, O. A., XU, Z., GESNER, E. M., VELDHOEN, R. A., KHANNA, S. K., GOPING, I. S., MACMILLAN, A. M., MENDELL, J. T., YOUNG, H. S., FAHLMAN, R. P. & GLOVER, J. N. (2011) Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biol,* 8**,** 1105-14.

CHEKULAEVA, M. & FILIPOWICZ, W. (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol.*

CHEN, J. F., MANDEL, E. M., THOMSON, J. M., WU, Q., CALLIS, T. E., HAMMOND, S. M., CONLON, F. L. & WANG, D. Z. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet,* 38**,** 228-33.

CHEN, K. & RAJEWSKY, N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet,* 8**,** 93-103.

CHEN, X., BA, Y., MA, L., CAI, X., YIN, Y., WANG, K., GUO, J., ZHANG, Y., CHEN, J., GUO, X., LI, Q., LI, X., WANG, W., ZHANG, Y., WANG, J., JIANG, X., XIANG, Y., XU, C., ZHENG, P., ZHANG, J., LI, R., ZHANG, H., SHANG, X., GONG, T., NING, G., WANG, J., ZEN, K., ZHANG, J. & ZHANG, C. Y. (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res,* 18**,** 997-1006.

CHENDRIMADA, T. P., FINN, K. J., JI, X., BAILLAT, D., GREGORY, R. I., LIEBHABER, S. A., PASQUINELLI, A. E. & SHIEKHATTAR, R. (2007) MicroRNA silencing through RISC recruitment of eIF6. *Nature,* 447**,** 823-8.

CHO, W. C. (2010) MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int J Biochem Cell Biol,* 42**,** 1273-81.

CLANCY, J. L., WEI, G. H., ECHNER, N., HUMPHREYS, D. T., BEILHARZ, T. H. & PREISS, T. (2011) mRNA isoform diversity can obscure detection of miRNA-mediated control of translation. *RNA,* 17**,** 1025-31.

COLGAN, D. F. & MANLEY, J. L. (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev,* 11**,** 2755-66.

COOPER, T. A., WAN, L. & DREYFUSS, G. (2009) RNA and disease. *Cell,* 136**,** 777-93.

CORRIONERO, A., MINANA, B. & VALCARCEL, J. (2011) Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev,* 25**,** 445-59.

CRICK, F. (1970) Central dogma of molecular biology. *Nature,* 227**,** 561-3.

CUI, E. H., LI, H. J., HUA, F., WANG, B., MAO, W., FENG, X. R., LI, J. Y. & WANG, X. (2012a) Serum microRNA 125b as a diagnostic or prognostic biomarker for advanced NSCLC patients receiving cisplatin-based chemotherapy. *Acta Pharmacol Sin*.

CUI, Y., XIAO, Z., HAN, J., SUN, J., DING, W., ZHAO, Y., CHEN, B., LI, X. & DAI, J. (2012b) MiR-125b orchestrates cell proliferation, differentiation and migration in neural stem/progenitor cells by targeting Nestin. *BMC Neurosci,* 13**,** 116.

DAHM, R. (2005) Friedrich Miescher and the discovery of DNA. *Dev Biol,* 278**,** 274-88.

DAVIS, B. N. & HATA, A. (2009) Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Commun Signal,* 7**,** 18.

DAVIS, B. N., HILYARD, A. C., LAGNA, G. & HATA, A. (2008) SMAD proteins control DROSHA-mediated microRNA maturation. *Nature,* 454**,** 56-61.

DER, S. D., ZHOU, A., WILLIAMS, B. R. & SILVERMAN, R. H. (1998) Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A,* 95**,** 15623-8.

DI GIAMMARTINO, D. C., NISHIDA, K. & MANLEY, J. L. (2011) Mechanisms and consequences of alternative polyadenylation. *Mol Cell,* 43**,** 853-66.

DIEDERICHS, S. & HABER, D. A. (2007) Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell,* 131**,** 1097-108.

DOLKEN, L., RUZSICS, Z., RADLE, B., FRIEDEL, C. C., ZIMMER, R., MAGES, J., HOFFMANN, R., DICKINSON, P., FORSTER, T., GHAZAL, P. & KOSZINOWSKI, U. H. (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *Rna,* 14**,** 1959-72.

DU, L., SCHAGEMAN, J. J., SUBAUSTE, M. C., SABER, B., HAMMOND, S. M., PRUDKIN, L., WISTUBA, II, JI, L., ROTH, J. A., MINNA, J. D. & PERTSEMLIDIS, A. (2009) miR-93, miR-98, and miR-197 regulate expression of tumor suppressor gene FUS1. *Mol Cancer Res,* 7**,** 1234-43.

DU, T. & ZAMORE, P. D. (2005) microPrimer: the biogenesis and function of microRNA. *Development,* 132**,** 4645-52.

DUAN, R., PAK, C. & JIN, P. (2007) Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum Mol Genet,* 16**,** 1124-31.

DUMESIC, P. A., NATARAJAN, P., CHEN, C., DRINNENBERG, I. A., SCHILLER, B. J., THOMPSON, J., MORESCO, J. J., YATES, J. R., 3RD, BARTEL, D. P. & MADHANI, H. D. (2013) Stalled Spliceosomes Are a Signal for RNAi-Mediated Genome Defense. *Cell,* 152**,** 957-68.

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.

ESQUELA-KERSCHER, A. & SLACK, F. J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer,* 6**,** 259-69.

ESTELLER, M. (2011) Non-coding RNAs in human disease. *Nat Rev Genet,* 12**,** 861-74.

FENG, B., JIANG, J., KRAUS, P., NG, J. H., HENG, J. C., CHAN, Y. S., YAW, L. P., ZHANG, W., LOH, Y. H., HAN, J., VEGA, V. B., CACHEUX-RATABOUL, V., LIM, B., LUFKIN, T. & NG, H. H. (2009) Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol,* 11**,** 197-203.

FIRE, A. Z. (2007) Gene silencing by double-stranded RNA (Nobel Lecture). *Angew Chem Int Ed Engl,* 46**,** 6966-84.

FRIEL, R., VAN DER SAR, S. & MEE, P. J. (2005) Embryonic stem cells: understanding their history, cell biology and signalling. *Adv Drug Deliv Rev,* 57**,** 1894-903.

FUKUDA, T., YAMAGATA, K., FUJIYAMA, S., MATSUMOTO, T., KOSHIDA, I., YOSHIMURA, K., MIHARA, M., NAITOU, M., ENDOH, H., NAKAMURA, T., AKIMOTO, C., YAMAMOTO, Y., KATAGIRI, T., FOULDS, C., TAKEZAWA, S., KITAGAWA, H., TAKEYAMA, K., O'MALLEY, B. W. & KATO, S. (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat Cell Biol,* 9**,** 604-11.

GALEJ, W. P., OUBRIDGE, C., NEWMAN, A. J. & NAGAI, K. (2013) Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature,* 493**,** 638-43.

GANGARAJU, V. K. & LIN, H. (2009) MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol,* 10**,** 116-25.

GARZON, R., CALIN, G. A. & CROCE, C. M. (2009) MicroRNAs in Cancer. *Annu Rev Med,* 60**,** 167-79.

GEBESHUBER, C. A., ZATLOUKAL, K. & MARTINEZ, J. (2009) miR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis. *EMBO Rep,* 10**,** 400-5.

GERRARD, L., RODGERS, L. & CUI, W. (2005) Differentiation of human embryonic stem cells to neural lineages in adherent culture by blocking bone morphogenetic protein signaling. *Stem Cells,* 23**,** 1234-41.

GIRALDEZ, A. J., MISHIMA, Y., RIHEL, J., GROCOCK, R. J., VAN DONGEN, S., INOUE, K., ENRIGHT, A. J. & SCHIER, A. F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science,* 312**,** 75-9.

GREGORY, R. I., YAN, K. P., AMUTHAN, G., CHENDRIMADA, T., DORATOTAJ, B., COOCH, N. & SHIEKHATTAR, R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature,* 432**,** 235-40.

GRUBER, A. R., MARTIN, G., KELLER, W. & ZAVOLAN, M. (2012) Cleavage Factor Im is a key regulator of 3' UTR length. *RNA Biol,* 9.

GRUBER, J. J., ZATECHKA, D. S., SABIN, L. R., YONG, J., LUM, J. J., KONG, M., ZONG, W. X., ZHANG, Z., LAU, C. K., RAWLINGS, J., CHERRY, S., IHLE, J. N., DREYFUSS, G. & THOMPSON, C. B. (2009) Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell,* 138**,** 328-39.

GRZYBOWSKA, E. A. Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. (2012) *Biochem Biophys Res Commun,* 424**,** 1-6.

GUIL, S. & CACERES, J. F. (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol,* 14**,** 591-6.

GUO, H., INGOLIA, N. T., WEISSMAN, J. S. & BARTEL, D. P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature,* 466**,** 835-40.

HAGIWARA, M. (2005) Alternative splicing: a new drug target of the post-genome era. *Biochim Biophys Acta,* 1754**,** 324-31.

HAJ KHELIL, A., DEGUILLIEN, M., MORINIERE, M., BEN CHIBANI, J. & BAKLOUTI, F. (2008) Cryptic splicing sites are differentially utilized in vivo. *Febs J,* 275**,** 1150-62.

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., PEDERSEN, J. S., KWON, S. C., BELAIR, C. D., KIM, Y. K., YEOM, K. H., YANG, W. Y., HAUSSLER, D., BLELLOCH, R. & KIM, V. N. (2009) Posttranscriptional crossregulation between Drosha and DGCR8. *Cell,* 136**,** 75-84.

HANSEN, T. B., JENSEN, T. I., CLAUSEN, B. H., BRAMSEN, J. B., FINSEN, B., DAMGAARD, C. K. & KJEMS, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature,* 495**,** 384-8.

HARISMENDY, O., NOTANI, D., SONG, X., RAHIM, N. G., TANASA, B., HEINTZMAN, N., REN, B., FU, X. D., TOPOL, E. J., ROSENFELD, M. G. & FRAZER, K. A. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature,* 470**,** 264-8.

HARPER, S. J. & BATES, D. O. (2008) VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat Rev Cancer,* 8**,** 880-7.

HE, L. & HANNON, G. J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet,* 5**,** 522-31.

HE, L., HE, X., LOWE, S. W. & HANNON, G. J. (2007) microRNAs join the p53 network--another piece in the tumour-suppression puzzle. *Nat Rev Cancer,* 7**,** 819-22.

HE, L., THOMSON, J. M., HEMANN, M. T., HERNANDO-MONGE, E., MU, D., GOODSON, S., POWERS, S., CORDON-CARDO, C., LOWE, S. W., HANNON, G. J. & HAMMOND, S. M. (2005) A microRNA polycistron as a potential human oncogene. *Nature,* 435**,** 828-33.

HEO, I., JOO, C., CHO, J., HA, M., HAN, J. & KIM, V. N. (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell,* 32**,** 276-84.

HINTERSEHER, I., TROMP, G. & KUIVANIEMI, H. (2011) Genes and abdominal aortic aneurysm. *Ann Vasc Surg,* 25**,** 388-412.

HIRIART, E., VAVASSEUR, A., TOUAT-TODESCHINI, L., YAMASHITA, A., GILQUIN, B., LAMBERT, E., PEROT, J., SHICHINO, Y., NAZARET, N., BOYAULT, C., LACHUER, J., PERAZZA, D., YAMAMOTO, M. & VERDEL, A. (2012) Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. *EMBO J,* 31**,** 2296-308.

HUA, Y., SAHASHI, K., HUNG, G., RIGO, F., PASSINI, M. A., BENNETT, C. F. & KRAINER, A. R. (2010) Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev,* 24**,** 1634-44.

HUDDER, A. & NOVAK, R. F. (2008) miRNAs: effectors of environmental influences on gene expression and disease. *Toxicol Sci,* 103**,** 228-40.

HUMPHREYS, D. T., WESTMAN, B. J., MARTIN, D. I. & PREISS, T. (2005) MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A,* 102**,** 16961-6.

HUTVAGNER, G., MCLACHLAN, J., PASQUINELLI, A. E., BALINT, E., TUSCHL, T. & ZAMORE, P. D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science,* 293**,** 834-8.

HWANG, H. W., WENTZEL, E. A. & MENDELL, J. T. (2007) A hexanucleotide element directs microRNA nuclear import. *Science,* 315**,** 97-100.

JAZDZEWSKI, K., MURRAY, E. L., FRANSSILA, K., JARZAB, B., SCHOENBERG, D. R. & DE LA CHAPELLE, A. (2008) Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc Natl Acad Sci U S A,* 105**,** 7269-74.

JI, Z. & TIAN, B. (2009) Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One,* 4**,** e8419.

JUDSON, R. L., BABIARZ, J. E., VENERE, M. & BLELLOCH, R. (2009) Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol,* 27**,** 459-61.

KAHAI, S., LEE, S. C., LEE, D. Y., YANG, J., LI, M., WANG, C. H., JIANG, Z., ZHANG, Y., PENG, C. & YANG, B. B. (2009) MicroRNA miR-378 regulates nephronectin expression modulating osteoblast differentiation by targeting GalNT-7. *PLoS One,* 4**,** e7535.

KAIDA, D., BERG, M. G., YOUNIS, I., KASIM, M., SINGH, L. N., WAN, L. & DREYFUSS, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature,* 468**,** 664-8.

KAIDA, D., MOTOYOSHI, H., TASHIRO, E., NOJIMA, T., HAGIWARA, M., ISHIGAMI, K., WATANABE, H., KITAHARA, T., YOSHIDA, T., NAKAJIMA, H., TANI, T., HORINOUCHI, S. & YOSHIDA, M. (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol,* 3**,** 576-83.

KAIKKONEN, M. U., LAM, M. T. & GLASS, C. K. (2011) Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res,* 90**,** 430-40.

KALKATAWI, M., RANGKUTI, F., SCHRAMM, M., JANKOVIC, B. R., KAMAU, A., CHOWDHARY, R., ARCHER, J. A. & BAJIC, V. B. (2012) Dragon PolyA Spotter: predictor of poly(A) motifs within human genomic DNA sequences. *Bioinformatics,* 28**,** 127-9.

KALSOTRA, A. & COOPER, T. A. (2011) Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet,* 12**,** 715-29.

KATAOKA, N., FUJITA, M. & OHNO, M. (2009) Functional association of the Microprocessor complex with the spliceosome. *Mol Cell Biol,* 29**,** 3243-54.

KAWAHARA, Y., ZINSHTEYN, B., CHENDRIMADA, T. P., SHIEKHATTAR, R. & NISHIKURA, K. (2007) RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep,* 8**,** 763-9.

KEDDE, M., STRASSER, M. J., BOLDAJIPOUR, B., OUDE VRIELINK, J. A., SLANCHEV, K., LE SAGE, C., NAGEL, R., VOORHOEVE, P. M., VAN DUIJSE, J., OROM, U. A., LUND, A. H., PERRAKIS, A., RAZ, E. & AGAMI, R. (2007) RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell,* 131**,** 1273-86.

KEFAS, B., GODLEWSKI, J., COMEAU, L., LI, Y., ABOUNADER, R., HAWKINSON, M., LEE, J., FINE, H., CHIOCCA, E. A., LAWLER, S. & PUROW, B. (2008) microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. *Cancer Res,* 68**,** 3566-72.

KENT, O. A., REAYI, A., FOONG, L., CHILIBECK, K. A. & MACMILLAN, A. M. (2003) Structuring of the 3' splice site by U2AF65. *J Biol Chem,* 278**,** 50572-7.

KENZELMANN, M., MAERTENS, S., HERGENHAHN, M., KUEFFER, S., HOTZ-WAGENBLATT, A., LI, L., WANG, S., ITTRICH, C., LEMBERGER, T., ARRIBAS, R., JONNAKUTY, S., HOLLSTEIN, M. C., SCHMID, W., GRETZ, N., GRONE, H. J. & SCHUTZ, G. (2007) Microarray analysis of newly synthesized RNA in cells and animals. *Proc Natl Acad Sci U S A,* 104**,** 6164-9.

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

KIM, D. H., SAETROM, P., SNOVE, O., JR. & ROSSI, J. J. (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A,* 105**,** 16230-5.

KIM, J. & ORKIN, S. H. (2011) Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine. *Genome Med,* 3**,** 75.

KIM, V. N., HAN, J. & SIOMI, M. C. (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol,* 10**,** 126-39.

KIM, Y. K. & KIM, V. N. (2007) Processing of intronic microRNAs. *Embo J,* 26**,** 775-83.

KIRIAKIDOU, M., TAN, G. S., LAMPRINAKI, S., DE PLANELL-SAGUER, M., NELSON, P. T. & MOURELATOS, Z. (2007) An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell,* 129**,** 1141-51.

KLOOSTERMAN, W. P. & PLASTERK, R. H. (2006) The diverse functions of microRNAs in animal development and disease. *Dev Cell,* 11**,** 441-50.

KOONIN, E. V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct,* 1**,** 22.

KORNBLIHTT, A. R., SCHOR, I. E., ALLO, M., DUJARDIN, G., PETRILLO, E. & MUNOZ, M. J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol,* 14**,** 153-65.

KUCHENBAUER, F., MORIN, R. D., ARGIROPOULOS, B., PETRIV, O. I., GRIFFITH, M., HEUSER, M., YUNG, E., PIPER, J., DELANEY, A., PRABHU, A. L., ZHAO, Y., MCDONALD, H., ZENG, T., HIRST, M., HANSEN, C. L., MARRA, M. A. & HUMPHRIES, R. K. (2008) In-depth characterization of the microRNA transcriptome in a leukemia progression model. *Genome Res,* 18**,** 1787-97.

KUO, C. H. & YING, S. Y. (2013) MicroRNA-mediated somatic cell reprogramming. *J Cell Biochem,* 114**,** 275-81.

LADEWIG, E., OKAMURA, K., FLYNT, A. S., WESTHOLM, J. O. & LAI, E. C. (2012) Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res,* 22**,** 1634-45.

LAGOS-QUINTANA, M., RAUHUT, R., LENDECKEL, W. & TUSCHL, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science,* 294**,** 853-8.

LAI, E. C. (2002) Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet,* 30**,** 363-4.

LAKSHMIPATHY, U., LOVE, B., GOFF, L. A., JORNSTEN, R., GRAICHEN, R., HART, R. P. & CHESNUT, J. D. (2007) MicroRNA expression pattern of undifferentiated and differentiated human embryonic stem cells. *Stem Cells Dev,* 16**,** 1003-16.

LAU, N. C., LIM, L. P., WEINSTEIN, E. G. & BARTEL, D. P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science,* 294**,** 858-62.

LAURENT, L. C., CHEN, J., ULITSKY, I., MUELLER, F. J., LU, C., SHAMIR, R., FAN, J. B. & LORING, J. F. (2008) Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells,* 26**,** 1506-16.

LE HIR, H., IZAURRALDE, E., MAQUAT, L. E. & MOORE, M. J. (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *Embo J,* 19**,** 6860-9.

LEE, E. J., BAEK, M., GUSEV, Y., BRACKETT, D. J., NUOVO, G. J. & SCHMITTGEN, T. D. (2008) Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *Rna,* 14**,** 35-42.

LEE, L. W., ZHANG, S., ETHERIDGE, A., MA, L., MARTIN, D., GALAS, D. & WANG, K. (2010) Complexity of the microRNA repertoire revealed by next-generation sequencing. *Rna,* 16**,** 2170-80.

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.

LEGENDRE, M., RITCHIE, W., LOPEZ, F. & GAUTHERET, D. (2006) Differential repression of alternative transcripts: a screen for miRNA targets. *PLoS Comput Biol,* 2**,** e43.

LEONARDO, T. R., SCHULTHEISZ, H. L., LORING, J. F. & LAURENT, L. C. (2012) The functions of microRNAs in pluripotency and reprogramming. *Nat Cell Biol,* 14**,** 1114-21.

LI, M. A. & HE, L. (2012) microRNAs as novel regulators of stem cell pluripotency and somatic cell reprogramming. *Bioessays,* 34**,** 670-80.

LI, Y., GUESSOUS, F., ZHANG, Y., DIPIERRO, C., KEFAS, B., JOHNSON, E., MARCINKIEWICZ, L., JIANG, J., YANG, Y., SCHMITTGEN, T. D., LOPES, B., SCHIFF, D., PUROW, B. & ABOUNADER, R. (2009) MicroRNA-34a inhibits glioblastoma growth by targeting multiple oncogenes. *Cancer Res,* 69**,** 7569-76.

LIAO, B., BAO, X., LIU, L., FENG, S., ZOVOILIS, A., LIU, W., XUE, Y., CAI, J., GUO, X., QIN, B., ZHANG, R., WU, J., LAI, L., TENG, M., NIU, L., ZHANG, B., ESTEBAN, M. A. & PEI, D. (2011) MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J Biol Chem,* 286**,** 17359-64.

LIAO, J. Y., MA, L. M., GUO, Y. H., ZHANG, Y. C., ZHOU, H., SHAO, P., CHEN, Y. Q. & QU, L. H. (2010) Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS One,* 5**,** e10563.

LICATALOSI, D. D. & DARNELL, R. B. (2010) RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet,* 11**,** 75-87.

LIN, S. L., CHANG, D. C., CHANG-LIN, S., LIN, C. H., WU, D. T., CHEN, D. T. & YING, S. Y. (2008) Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *Rna,* 14**,** 2115-24.

LISTERMAN, I., SAPRA, A. K. & NEUGEBAUER, K. M. (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol,* 13**,** 815-22.

LIU, Q. & PAROO, Z. (2010) Biochemical principles of small RNA pathways. *Annu Rev Biochem,* 79**,** 295-319.

LO, C. W., KAIDA, D., NISHIMURA, S., MATSUYAMA, A., YASHIRODA, Y., TAOKA, H., ISHIGAMI, K., WATANABE, H., NAKAJIMA, H., TANI, T., HORINOUCHI, S. & YOSHIDA, M. (2007) Inhibition of splicing and nuclear retention of pre-mRNA by spliceostatin A in fission yeast. *Biochem Biophys Res Commun,* 364**,** 573-7.

LODISH, H. (2000) *Molecular cell biology,* New York, W.H. Freeman.

LODISH, H. F. & PORTER, M. (1980) Translational control of protein synthesis after infection by vesicular stomatitis virus. *J Virol,* 36**,** 719-33.

LU, J., SHEN, Y., WU, Q., KUMAR, S., HE, B., SHI, S., CARTHEW, R. W., WANG, S. M. & WU, C. I. (2008) The birth and death of microRNA genes in Drosophila. *Nat Genet,* 40**,** 351-5.

LUCIANO, D. J., MIRSKY, H., VENDETTI, N. J. & MAAS, S. (2004) RNA editing of a miRNA precursor. *Rna,* 10**,** 1174-7.

LUI, W. O., POURMAND, N., PATTERSON, B. K. & FIRE, A. (2007) Patterns of known and novel small RNAs in human cervical cancer. *Cancer Res,* 67**,** 6031-43.

LUND, E., GUTTINGER, S., CALADO, A., DAHLBERG, J. E. & KUTAY, U. (2004) Nuclear export of microRNA precursors. *Science,* 303**,** 95-8.

LUTZ, C. S. (2008) Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol,* 3**,** 609-17.

MA, F., XU, S., LIU, X., ZHANG, Q., XU, X., LIU, M., HUA, M., LI, N., YAO, H. & CAO, X. (2011) The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-gamma. *Nat Immunol,* 12**,** 861-9.

MANIATIS, T. & REED, R. (2002) An extensive network of coupling among gene expression machines. *Nature,* 416**,** 499-506.

MARSON, A., LEVINE, S. S., COLE, M. F., FRAMPTON, G. M., BRAMBRINK, T., JOHNSTONE, S., GUENTHER, M. G., JOHNSTON, W. K., WERNIG, M., NEWMAN, J., CALABRESE, J. M., DENNIS, L. M., VOLKERT, T. L., GUPTA, S., LOVE, J., HANNETT, N., SHARP, P. A., BARTEL, D. P., JAENISCH, R. & YOUNG, R. A. (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell,* 134**,** 521-33.

MARTENS-UZUNOVA, E. S., OLVEDY, M. & JENSTER, G. (2013) Beyond microRNA - Novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett*.

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.

MAYEDA, A. & KRAINER, A. R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell,* 68**,** 365-375.

MAYR, C. & BARTEL, D. P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell,* 138**,** 673-84.

MEISTER, G., LANDTHALER, M., PATKANIOWSKA, A., DORSETT, Y., TENG, G. & TUSCHL, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell,* 15**,** 185-97.

MEMCZAK, S., JENS, M., ELEFSINIOTI, A., TORTI, F., KRUEGER, J., RYBAK, A., MAIER, L., MACKOWIAK, S. D., GREGERSEN, L. H., MUNSCHAUER, M., LOEWER, A., ZIEBOLD, U., LANDTHALER, M., KOCKS, C., LE NOBLE, F. & RAJEWSKY, N. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature,* 495**,** 333-8.

MERRITT, W. M., LIN, Y. G., HAN, L. Y., KAMAT, A. A., SPANNUTH, W. A., SCHMANDT, R., URBAUER, D., PENNACCHIO, L. A., CHENG, J. F., NICK, A. M., DEAVERS, M. T., MOURAD-ZEIDAN, A., WANG, H., MUELLER, P., LENBURG, M. E., GRAY, J. W., MOK, S., BIRRER, M. J., LOPEZ-BERESTEIN, G., COLEMAN, R. L., BAR-ELI, M. & SOOD, A. K. (2008) Dicer, Drosha, and outcomes in patients with ovarian cancer. *N Engl J Med,* 359**,** 2641-50.

MICHLEWSKI, G., GUIL, S. & CACERES, J. F. (2010) Stimulation of pri-miR-18a Processing by hnRNP A1. *Adv Exp Med Biol,* 700**,** 28-35.

MICHLEWSKI, G., GUIL, S., SEMPLE, C. A. & CACERES, J. F. (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell,* 32**,** 383-93.

MIRONOV, A. A., FICKETT, J. W. & GELFAND, M. S. (1999) Frequent alternative splicing of human genes. *Genome Res,* 9**,** 1288-93.

MIYOSHI, K., TSUKUMO, H., NAGAMI, T., SIOMI, H. & SIOMI, M. C. (2005) Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes Dev,* 19**,** 2837-48.

MIYOSHI, N., ISHII, H., NAGANO, H., HARAGUCHI, N., DEWI, D. L., KANO, Y., NISHIKAWA, S., TANEMURA, M., MIMORI, K., TANAKA, F., SAITO, T., NISHIMURA, J., TAKEMASA, I., MIZUSHIMA, T., IKEDA, M., YAMAMOTO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. (2011) Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell,* 8**,** 633-8.

MORLANDO, M., BALLARINO, M., GROMAK, N., PAGANO, F., BOZZONI, I. & PROUDFOOT, N. J. (2008) Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol,* 15**,** 902-9.

MUELLER, A. A., CHEUNG, T. H. & RANDO, T. A. (2013) All's well that ends well: alternative polyadenylation and its implications for stem cell biology. *Curr Opin Cell Biol,* 25**,** 222-32.

MUKHOPADHYAY, R., JIA, J., ARIF, A., RAY, P. S. & FOX, P. L. (2009) The GAIT system: a gatekeeper of inflammatory gene expression. *Trends Biochem Sci,* 34**,** 324-31.

MURAKI, M., OHKAWARA, B., HOSOYA, T., ONOGI, H., KOIZUMI, J., KOIZUMI, T., SUMI, K., YOMODA, J., MURRAY, M. V., KIMURA, H., FURUICHI, K., SHIBUYA, H., KRAINER, A. R., SUZUKI, M. & HAGIWARA, M. (2004) Manipulation of alternative splicing by a newly developed inhibitor of Clks. *J Biol Chem,* 279**,** 24246-54.

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,* 102**,** 12135-40.

NAM, D. K., LEE, S., ZHOU, G., CAO, X., WANG, C., CLARK, T., CHEN, J., ROWLEY, J. D. & WANG, S. M. (2002) Oligo(dT) primer generates a high

frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc Natl Acad Sci U S A,* 99**,** 6152-6.
NEWMAN, M. A. & HAMMOND, S. M. (2010) Emerging paradigms of regulated microRNA processing. *Genes Dev,* 24**,** 1086-92.
NEWMAN, M. A., MANI, V. & HAMMOND, S. M. (2011) Deep sequencing of microRNA precursors reveals extensive 3' end modification. *Rna,* 17**,** 1795-803.
NEWMAN, M. A., THOMSON, J. M. & HAMMOND, S. M. (2008) Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *Rna,* 14**,** 1539-49.
NG, B., YANG, F., HUSTON, D. P., YAN, Y., YANG, Y., XIONG, Z., PETERSON, L. E., WANG, H. & YANG, X. F. (2004) Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of untolerized epitopes. *J Allergy Clin Immunol,* 114**,** 1463-70.
NILSEN, T. W. & GRAVELEY, B. R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature,* 463**,** 457-63.
NOTTROTT, S., SIMARD, M. J. & RICHTER, J. D. (2006) Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol,* 13**,** 1108-14.
OBERNOSTERER, G., LEUSCHNER, P. J., ALENIUS, M. & MARTINEZ, J. (2006) Post-transcriptional regulation of microRNA expression. *Rna,* 12**,** 1161-7.
OLSEN, P. H. & AMBROS, V. (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol,* 216**,** 671-80.
OUNZAIN, S., CRIPPA, S. & PEDRAZZINI, T. (2013) Small and long non-coding RNAs in cardiac homeostasis and regeneration. *Biochim Biophys Acta,* 1833**,** 923-33.
PARENTEAU, J., DURAND, M., VERONNEAU, S., LACOMBE, A. A., MORIN, G., GUERIN, V., CECEZ, B., GERVAIS-BIRD, J., KOH, C. S., BRUNELLE, D., WELLINGER, R. J., CHABOT, B. & ABOU ELELA, S. (2008) Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol Biol Cell,* 19**,** 1932-41.
PAROO, Z., YE, X., CHEN, S. & LIU, Q. (2009) Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell,* 139**,** 112-22.
PASQUINELLI, A. E., REINHART, B. J., SLACK, F., MARTINDALE, M. Q., KURODA, M. I., MALLER, B., HAYWARD, D. C., BALL, E. E., DEGNAN, B., MULLER, P., SPRING, J., SRINIVASAN, A., FISHMAN, M., FINNERTY, J., CORBO, J., LEVINE, M., LEAHY, P., DAVIDSON, E. & RUVKUN, G. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature,* 408**,** 86-9.
PAWLICKI, J. M. & STEITZ, J. A. (2010) Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function. *Trends Cell Biol,* 20**,** 52-61.
PETERSEN, C. P., BORDELEAU, M. E., PELLETIER, J. & SHARP, P. A. (2006) Short RNAs repress translation after initiation in mammalian cells. *Mol Cell,* 21**,** 533-42.

PICKRELL, J. K., PAI, A. A., GILAD, Y. & PRITCHARD, J. K. (2010) Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet,* 6**,** e1001236.

PLACE, R. F., LI, L. C., POOKOT, D., NOONAN, E. J. & DAHIYA, R. (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A,* 105**,** 1608-13.

PLEISS, J. A., WHITWORTH, G. B., BERGKESSEL, M. & GUTHRIE, C. (2007) Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol,* 5**,** e90.

PONTING, C. P. & BELGARD, T. G. (2010) Transcribed dark matter: meaning or myth? *Hum Mol Genet,* 19**,** R162-8.

PREALL, J. B. & SONTHEIMER, E. J. (2005) RNAi: RISC gets loaded. *Cell,* 123**,** 543-5.

RAO, P. K., KUMAR, R. M., FARKHONDEH, M., BASKERVILLE, S. & LODISH, H. F. (2006) Myogenic factors that regulate expression of muscle-specific microRNAs. *Proc Natl Acad Sci U S A,* 103**,** 8721-6.

RIBAS, J., NI, X., CASTANARES, M., LIU, M. M., ESOPI, D., YEGNASUBRAMANIAN, S., RODRIGUEZ, R., MENDELL, J. T. & LUPOLD, S. E. (2012) A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res,* 40**,** 6821-33.

RINO, J. & CARMO-FONSECA, M. (2009) The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends Cell Biol,* 19**,** 375-84.

RODRIGUEZ, A., GRIFFITHS-JONES, S., ASHURST, J. L. & BRADLEY, A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res,* 14**,** 1902-10.

ROUSH, S. & SLACK, F. J. (2008) The let-7 family of microRNAs. *Trends Cell Biol,* 18**,** 505-16.

ROY, S. W. & GILBERT, W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet,* 7**,** 211-21.

ROYBAL, G. A. & JURICA, M. S. (2010) Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. *Nucleic Acids Res,* 38**,** 6664-72.

RYAZANSKY, S. S., GVOZDEV, V. A. & BEREZIKOV, E. (2011) Evidence for post-transcriptional regulation of clustered microRNAs in Drosophila. *BMC Genomics,* 12**,** 371.

SAKAMOTO, S., AOKI, K., HIGUCHI, T., TODAKA, H., MORISAWA, K., TAMAKI, N., HATANO, E., FUKUSHIMA, A., TANIGUCHI, T. & AGATA, Y. (2009) The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol Cell Biol,* 29**,** 3754-69.

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.

SELBACH, M., SCHWANHAUSSER, B., THIERFELDER, N., FANG, Z., KHANIN, R. & RAJEWSKY, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature,* 455**,** 58-63.

SENGUPTA, S., DEN BOON, J. A., CHEN, I. H., NEWTON, M. A., STANHOPE, S. A., CHENG, Y. J., CHEN, C. J., HILDESHEIM, A., SUGDEN, B. & AHLQUIST, P. (2008) MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating mRNAs encoding extracellular matrix proteins. *Proc Natl Acad Sci U S A,* 105**,** 5874-8.

SHAH, P. P., HUTCHINSON, L. E. & KAKAR, S. S. (2009) Emerging role of microRNAs in diagnosis and treatment of various diseases including ovarian cancer. *J Ovarian Res,* 2**,** 11.

SHALOM-FEUERSTEIN, R., SERROR, L., DE LA FOREST DIVONNE, S., PETIT, I., ABERDAM, E., CAMARGO, L., DAMOUR, O., VIGOUROUX, C., SOLOMON, A., GAGGIOLI, C., ITSKOVITZ-ELDOR, J., AHMAD, S. & ABERDAM, D. (2012) Pluripotent stem cell model reveals essential roles for miR-450b-5p and miR-184 in embryonic corneal lineage specification. *Stem Cells,* 30**,** 898-909.

SHAROVA, L. V., SHAROV, A. A., NEDOREZOV, T., PIAO, Y., SHAIK, N. & KO, M. S. (2009) Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res,* 16**,** 45-58.

SHI, Y. (2012) Alternative polyadenylation: new insights from global analyses. *RNA,* 18**,** 2105-17.

SHIOHAMA, A., SASAKI, T., NODA, S., MINOSHIMA, S. & SHIMIZU, N. (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochem Biophys Res Commun,* 304**,** 184-90.

SINGH, R. & VALCARCEL, J. (2005) Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol,* 12**,** 645-53.

SINGH, R. K. & COOPER, T. A. (2012) Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med,* 18**,** 472-82.

SIOMI, H. & SIOMI, M. C. (2010) Posttranscriptional regulation of microRNA biogenesis in animals. *Mol Cell,* 38**,** 323-32.

SLEZAK-PROCHAZKA, I., DURMUS, S., KROESEN, B. J. & VAN DEN BERG, A. (2010) MicroRNAs, macrocontrol: regulation of miRNA processing. *Rna,* 16**,** 1087-95.

SOTILLO, E. & THOMAS-TIKHONENKO, A. (2011) Shielding the messenger (RNA): microRNA-based anticancer therapies. *Pharmacol Ther,* 131**,** 18-32.

SUBRAMANYAM, D., LAMOUILLE, S., JUDSON, R. L., LIU, J. Y., BUCAY, N., DERYNCK, R. & BLELLOCH, R. (2011) Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat Biotechnol,* 29**,** 443-8.

SUH, M. R., LEE, Y., KIM, J. Y., KIM, S. K., MOON, S. H., LEE, J. Y., CHA, K. Y., CHUNG, H. M., YOON, H. S., MOON, S. Y., KIM, V. N. & KIM, K. S. (2004) Human embryonic stem cells express a unique set of microRNAs. *Dev Biol,* 270**,** 488-98.

SUMAZIN, P., YANG, X., CHIU, H. S., CHUNG, W. J., IYER, A., LLOBET-NAVAS, D., RAJBHANDARI, P., BANSAL, M., GUARNIERI, P., SILVA, J. & CALIFANO, A. (2011) An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell,* 147**,** 370-81.

SUN, M. M., LI, J. F., GUO, L. L., XIAO, H. T., DONG, L., WANG, F., HUANG, F. B., CAO, D., QIN, T., YIN, X. H., LI, J. M. & WANG, S. L. (2013) TGF-beta1 suppression of microRNA-450b-5p expression: a novel mechanism for blocking myogenic differentiation of rhabdomyosarcoma. *Oncogene*.

SUZUKI, H., MARUYAMA, R., YAMAMOTO, E. & KAI, M. (2012) DNA methylation and microRNA dysregulation in cancer. *Mol Oncol,* 6**,** 567-78.

SUZUKI, H., TAKATSUKA, S., AKASHI, H., YAMAMOTO, E., NOJIMA, M., MARUYAMA, R., KAI, M., YAMANO, H. O., SASAKI, Y., TOKINO, T., SHINOMURA, Y., IMAI, K. & TOYOTA, M. (2011) Genome-wide profiling of chromatin signatures reveals epigenetic regulation of MicroRNA genes in colorectal cancer. *Cancer Res,* 71**,** 5646-58.

SUZUKI, H. I., YAMAGATA, K., SUGIMOTO, K., IWAMOTO, T., KATO, S. & MIYAZONO, K. (2009) Modulation of microRNA processing by p53. *Nature,* 460**,** 529-33.

TABACH, Y., BILLI, A. C., HAYES, G. D., NEWMAN, M. A., ZUK, O., GABEL, H., KAMATH, R., YACOBY, K., CHAPMAN, B., GARCIA, S. M., BOROWSKY, M., KIM, J. K. & RUVKUN, G. (2013) Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature,* 493**,** 694-8.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TAKAHASHI, K. & YAMANAKA, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TANG, G. Q. & MAXWELL, E. S. (2008) Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome Res,* 18**,** 104-12.

TANI, H., MIZUTANI, R., SALAM, K. A., TANO, K., IJIRI, K., WAKAMATSU, A., ISOGAI, T., SUZUKI, Y. & AKIMITSU, N. (2012) Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res,* 22**,** 947-56.

THOMSON, J. M., NEWMAN, M., PARKER, J. S., MORIN-KENSICKI, E. M., WRIGHT, T. & HAMMOND, S. M. (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev,* 20**,** 2202-7.

THORNTON, J. E. & GREGORY, R. I. (2012) How does Lin28 let-7 control development and disease? *Trends Cell Biol,* 22**,** 474-82.

THORREZ, L. & SAMPAOLESI, M. (2011) The future of induced pluripotent stem cells for cardiac therapy and drug development. *Curr Pharm Des,* 17**,** 3258-70.

TRABUCCHI, M., BRIATA, P., GARCIA-MAYORAL, M., HAASE, A. D., FILIPOWICZ, W., RAMOS, A., GHERZI, R. & ROSENFELD, M. G. (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature,* 459**,** 1010-4.

TREIBER, T., TREIBER, N. & MEISTER, G. (2012) Regulation of microRNA biogenesis and function. *Thromb Haemost,* 107**,** 605-10.

TURUNEN, J. J., NIEMELA, E. H., VERMA, B. & FRILANDER, M. J. (2012) The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA,* 4**,** 61-76.

VALENTE, L. & NISHIKURA, K. (2005) ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. *Prog Nucleic Acid Res Mol Biol,* 79**,** 299-338.

VAN KOUWENHOVE, M., KEDDE, M. & AGAMI, R. (2011
) MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat Rev Cancer,* 11**,** 644-56.

VENABLES, J. P., KLINCK, R., KOH, C., GERVAIS-BIRD, J., BRAMARD, A., INKEL, L., DURAND, M., COUTURE, S., FROEHLICH, U., LAPOINTE, E., LUCIER, J. F., THIBAULT, P., RANCOURT, C., TREMBLAY, K., PRINOS, P., CHABOT, B. & ELELA, S. A. (2009) Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol,* 16**,** 670-6.

VIDIGAL, J. A. & VENTURA, A. (2012) Embryonic stem cell miRNAs and their roles in development and disease. *Semin Cancer Biol,* 22**,** 428-36.

VISWANATHAN, S. R., DALEY, G. Q. & GREGORY, R. I. (2008) Selective blockade of microRNA processing by Lin28. *Science,* 320**,** 97-100.

WAHL, M. C., WILL, C. L. & LUHRMANN, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell,* 136**,** 701-18.

WAKIYAMA, M., TAKIMOTO, K., OHARA, O. & YOKOYAMA, S. (2007) Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev,* 21**,** 1857-62.

WANG, E. T., SANDBERG, R., LUO, S., KHREBTUKOVA, I., ZHANG, L., MAYR, C., KINGSMORE, S. F., SCHROTH, G. P. & BURGE, C. B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature,* 456**,** 470-6.

WANG, J., LU, M., QIU, C. & CUI, Q. (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res,* 38**,** D119-22.

WANG, L. G. & GU, J. (2012) Serum microRNA-29a is a promising novel marker for early detection of colorectal liver metastasis. *Cancer Epidemiol,* 36**,** e61-7.

WANG, Y., LI, X. & HU, H. (2011) Transcriptional regulation of co-expressed microRNA target genes. *Genomics,* 98**,** 445-52.

WANG, Z., LIN, S., LI, J.J., XU, Z., YAO, H., ZHU, X., XIE, D., SHEN, Z., SZE, J., LI, K., LU, G., CHAN, D.T., POON, W.S., KUNG, H., LIN, M.C. (2011) MYC Protein Inhibits Transcription of the MicroRNA Cluster *MC-let-7a-1~let-7d*via Noncanonical E-box*. JBC 286, 39703–39714.

WANG, Y., MEDVID, R., MELTON, C., JAENISCH, R. & BLELLOCH, R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet,* 39**,** 380-5.

WIGHTMAN, B., HA, I. & RUVKUN, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell,* 75**,** 855-62.

WILL, C. L. & LUHRMANN, R. (2005) Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem,* 386**,** 713-24.

WILL, C. L. & LUHRMANN, R. (2011) Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology,* 3.

WILSON, K. D., VENKATASUBRAHMANYAM, S., JIA, F., SUN, N., BUTTE, A. J. & WU, J. C. (2009) MicroRNA profiling of human-induced pluripotent stem cells. *Stem Cells Dev,* 18**,** 749-58.

WINDHAGER, L., BONFERT, T., BURGER, K., RUZSICS, Z., KREBS, S., KAUFMANN, S., MALTERER, G., L'HERNAULT, A., SCHILHABEL, M., SCHREIBER, S., ROSENSTIEL, P., ZIMMER, R., EICK, D., FRIEDEL, C. C. & DOLKEN, L. (2012) Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res,* 22**,** 2031-42.

WINTER, J., KUNATH, M., ROEPCKE, S., KRAUSE, S., SCHNEIDER, R. & SCHWEIGER, S. (2007) Alternative polyadenylation signals and promoters

act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1. *BMC Mol Biol,* 8**,** 105.

WU, L., FAN, J. & BELASCO, J. G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A,* 103**,** 4034-9.

WU, M., JOLICOEUR, N., LI, Z., ZHANG, L., FORTIN, Y., L'ABBE, D., YU, Z. & SHEN, S. H. (2008) Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. *Carcinogenesis,* 29**,** 1710-6.

XUE, S., JIA, L. & JIA, J. (2009) Association between somatostatin gene polymorphisms and sporadic Alzheimer's disease in Chinese population. *Neurosci Lett,* 465**,** 181-3.

YAMAGATA, K., FUJIYAMA, S., ITO, S., UEDA, T., MURATA, T., NAITOU, M., TAKEYAMA, K., MINAMI, Y., O'MALLEY, B. W. & KATO, S. (2009) Maturation of microRNA is hormonally regulated by a nuclear receptor. *Mol Cell,* 36**,** 340-7.

YANG, W., CHENDRIMADA, T. P., WANG, Q., HIGUCHI, M., SEEBURG, P. H., SHIEKHATTAR, R. & NISHIKURA, K. (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol,* 13**,** 13-21.

YI, R., DOEHLE, B. P., QIN, Y., MACARA, I. G. & CULLEN, B. R. (2005) Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *Rna,* 11**,** 220-6.

YI, R., QIN, Y., MACARA, I. G. & CULLEN, B. R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev,* 17**,** 3011-6.

YOUNIS, I., BERG, M., KAIDA, D., DITTMAR, K., WANG, C. & DREYFUSS, G. (2010) Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol Cell Biol,* 30**,** 1718-28.

YU, B. & CHEN, X. (2010) Analysis of miRNA Modifications. *Methods Mol Biol,* 592**,** 137-48.

YUAN, Z., SUN, X., JIANG, D., DING, Y., LU, Z., GONG, L., LIU, H. & XIE, J. (2010) Origin and evolution of a placental-specific microRNA family in the human genome. *BMC Evol Biol,* 10**,** 346.

ZHANG, B., PAN, X., COBB, G. P. & ANDERSON, T. A. (2006) Plant microRNA: a small regulatory molecule with big impact. *Dev Biol,* 289**,** 3-16.

ZHANG, Y., ZHANG, R. & SU, B. (2009) Diversity and evolution of MicroRNA gene clusters. *Sci China C Life Sci,* 52**,** 261-6.

ZHAO, J., NI, N., LIU, C., ZHANG, P., YU, Y., CHEN, Z. & CHEN, W. (2011) Cloning and characterization of the 5' flanking region of microRNA let-7a-1/let-7f-1 gene cluster in human lung cancer cell. African Journal of Biotechnology, 10, 9233-9242.

ZHU, J., MAYEDA, A. & KRAINER, A. R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell,* 8**,** 1351-61.

Supplementary Material

Supplementary Table S1 ESC and NPSC miRNAs and their host mRNAs. List of all microRNAs cloned from ESCs and NPSCs. Overlapping host mRNA (Host) ID and average expression levels (ave. signal) as detected by microarray are given where relevant, na = not applicable, U = unclassified EST. Cloning, sequencing and microarray were performed by Dr Elcie Chan.

miRNA	ES1		HG -1	ave. signal	miRNA	NS		HG -1	ave. signal
302a*	2370		U		let-7f	4356		na	
130a	1006		na					HUWE1	238.20
182	905		na		let-7a	3701		U	
423-5p	717		CCDC55	352.30	let-7a	3701		na	
151-3p	663		PTK2	3410.40				na	
103	539		PANK3	416.50	21	3162		na	
			PANK2	1826.90	423-5p	2418		CCDC55	385.30
378	488		PPARGC1B	183.50	9*	1299		na	
340	474		RNF130	1660.90	9*	1299		C1ORF61	34425.20
21	445		na					na	
320a	413		na		92b	885		na	
101	397		na		9	870		na	
			RCL1	652.40	9	870		C1ORF61	34425.20
183	350		na					na	
130b	321		PPIL2	118.50	101	768		na	
335	176		MEST	619.50				RCL1	413.50
191	156		DALRD3	249.10	151-3p	761		PTK2	3408.00
30e*	151		NFYC	704.60	130a	689		na	
148a	142		na		340	682		RNF130	1683.20
221*	117		na		29a	559		na	
302b	115		U		let-7c	481		C21ORF34	665.50
30a*	109		C6ORF155	98.90	320a	448		na	
302a	105		U		103	427		PANK3	168.50
25	104		MCM7	2891.90				PANK2	835.60
93	89		MCM7	2891.90	26a	393		CTDSPL	478.50
31	83		U					CTDSP2	2810.60
302d	75		U		let-7b	371		U	
106b	71		MCM7	2891.90	let-7e	326		na	
302c	68		U		125b	297		na	
17	66		C13ORF25	264.00				C21ORF34	665.50
140-3p	79		WWP2	618.70	24	297		C9ORF3	196.90
1323	58		na					na	
503	48		MGC16121	190.40	378	280		PPARGC1B	149.40
28-3p	47		LPP	4342.80	130b	268		PPIL2	110.70
99b	47		na		30e*	222		NFYC	456.10
302d*	46		U		210	209		na	
367	45		U		30d	207		na	
30a	44		C6ORF155	98.90	let-7g	199		WDR82	2258.40
1308	43		na		100	197		na	
532	42		CLCN5	138.00	let-7d	190		na	
26a	38		CTDSPL	1321.90	92b*	188		na	
			CTDSP2	1969.00	16	175		DLEU2	131.10
185	35		C22ORF25	221.90				SMC4	410.60
20b	33		na		27b	170		C9ORF3	196.90
25*	33		MCM7	2891.90	140-3p	169		WWP2	227.00
708	31		na		181a	168		U	
374b	31		na					U	
16	30		DLEU2	152.30	25	166		MCM7	1275.70
			SMC4	874.60	let-7i	153		na	
744	30		MAP2K4	477.20	7	150		HNRPK	2797.30
29a	30		na					na	
30d	30		na		7	150		C19ORF30	251.10
222	28		na		744	135		MAP2K4	359.70
1270	27		ZNF826	108.60	221*	134		na	
1275	26		na		30a*	131		C6ORF155	88.70
124	25		na		99b	124		na	
141	25		na		30a	116		C6ORF155	88.70
92b	25		na		191	114		DALRD3	294.70
24	24		C9ORF3	147.10	335	101		MEST	107.50
			na		146b	88		na	
374a	24		U		106b	88		MCM7	1275.70

548f	24	PCDH15	97.10
		DMD	166.50
		ERBB4	123.70
		na	
		na	
221	22	na	
92b*	22	na	
219-2-3p	20	na	
		na	
let-7a	20	na	
		na	
92a	18	C13ORF25	264.00
151-5p	17	PTK2	3410.40
30c-2*	16	C6ORF155	98.90
708*	16	na	
106b*	15	MCM7	2891.90
193a-3p	15	na	
28-5p	13	LPP	4342.80
128	12	R3HDM1	215.30
		ARPP-21	103.90
363	12	na	
769-5p	11	na	
107	10	PANK1	602.20
210	10	na	
1260	10	na	
19b	10	C13ORF25	264.00
		na	
96	9	na	
421	9	na	
200c	9	na	
1	8	C20ORF166	84.90
192	8	na	
424	8	U	
20a	8	C13ORF25	264.00
186	7	ZRANB2	2422.40
941	7	DNAJC5	229.40
1298	7	HTR2C	141.20
27b	7	C9ORF3	147.10
92a-1*	6	C13ORF25	264.00
375	5	na	
1826	5	na	
181a	5	U	
18a	5	C13ORF25	264.00
302b*	5	U	

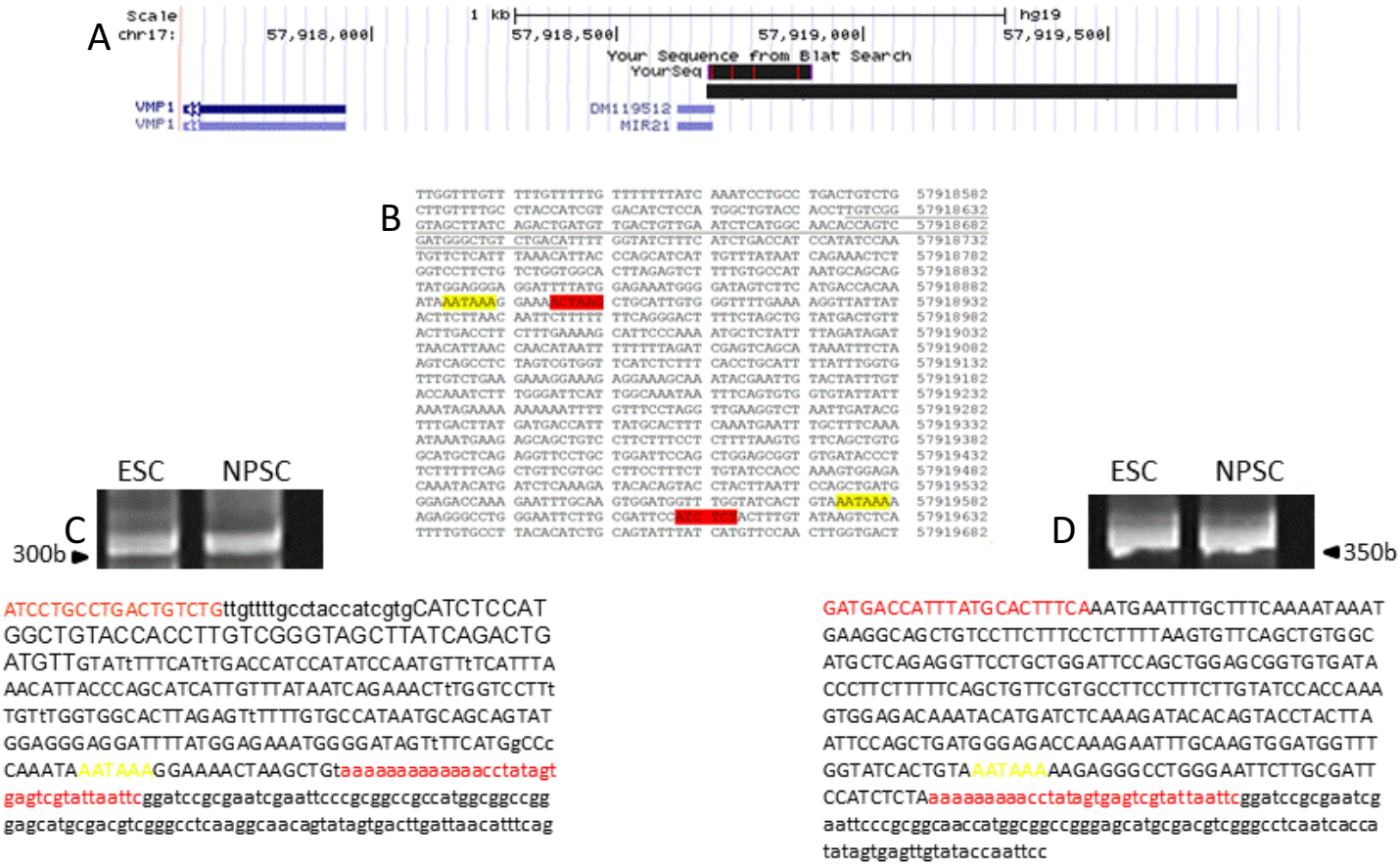
148a	85	na	
93	72	MCM7	1275.70
532	68	CLCN5	135.20
28-3p	58	LPP	2796.30
151-5p	57	PTK2	3408.00
185	54	C22ORF25	456.30
23a	53	na	
128	49	R3HDM1	220.00
		ARPP-21	101.30
99a	44	C21ORF34	665.50
27a	43	na	
222	42	na	
17	41	C13ORF25	180.90
25*	39	MCM7	1275.70
374b	38	na	
374a	35	U	
26b	34	CTDSP1	618.50
92a	34	C13ORF25	180.90
		na	
181b	32	U	
		U	
769-5p	29	na	
941	27	DNAJC5	244.20
941	27	DNAJC5	244.20
15a	24	DLEU2	131.10
124	23	na	
		na	
1274b	22	na	
107	21	PANK1	449.60
379	21	na	
135b	20	LEMD1	1779.80
148b	20	COPZ1	2367.80
23b	20	C9ORF3	196.90
20a	20	C13ORF25	180.90
30c-2*	19	C6ORF155	88.70
1260	18	na	
1275	18	na	
98	18	HUWE1	238.20
195	17	U	
138	17	na	
		na	
221	17	na	
542-3p	16	na	
29c	15	na	
720	14	na	
10a	14	na	
30e	13	NFYC	456.10
450b	12	na	
186	11	ZRANB2	1837.30
153	11	PTPRN	91.90
		PTPRN2	228.60
19b	11	C13ORF25	180.90
		na	
28-5p	9	LPP	2796.30
125b-2*	8	C21ORF34	665.50
421	7	na	
30b	7	na	
22	7	C17ORF91	451.90
424	6	U	
324-5p	6	na	
532-5p	6	CLCN5	135.20
125a-5p	5	na	
342-3p	5	EVL	8128.60

Supplementary Table S2 List of verified poly(A) signal motifs (Kalkatawi et al., 2012)

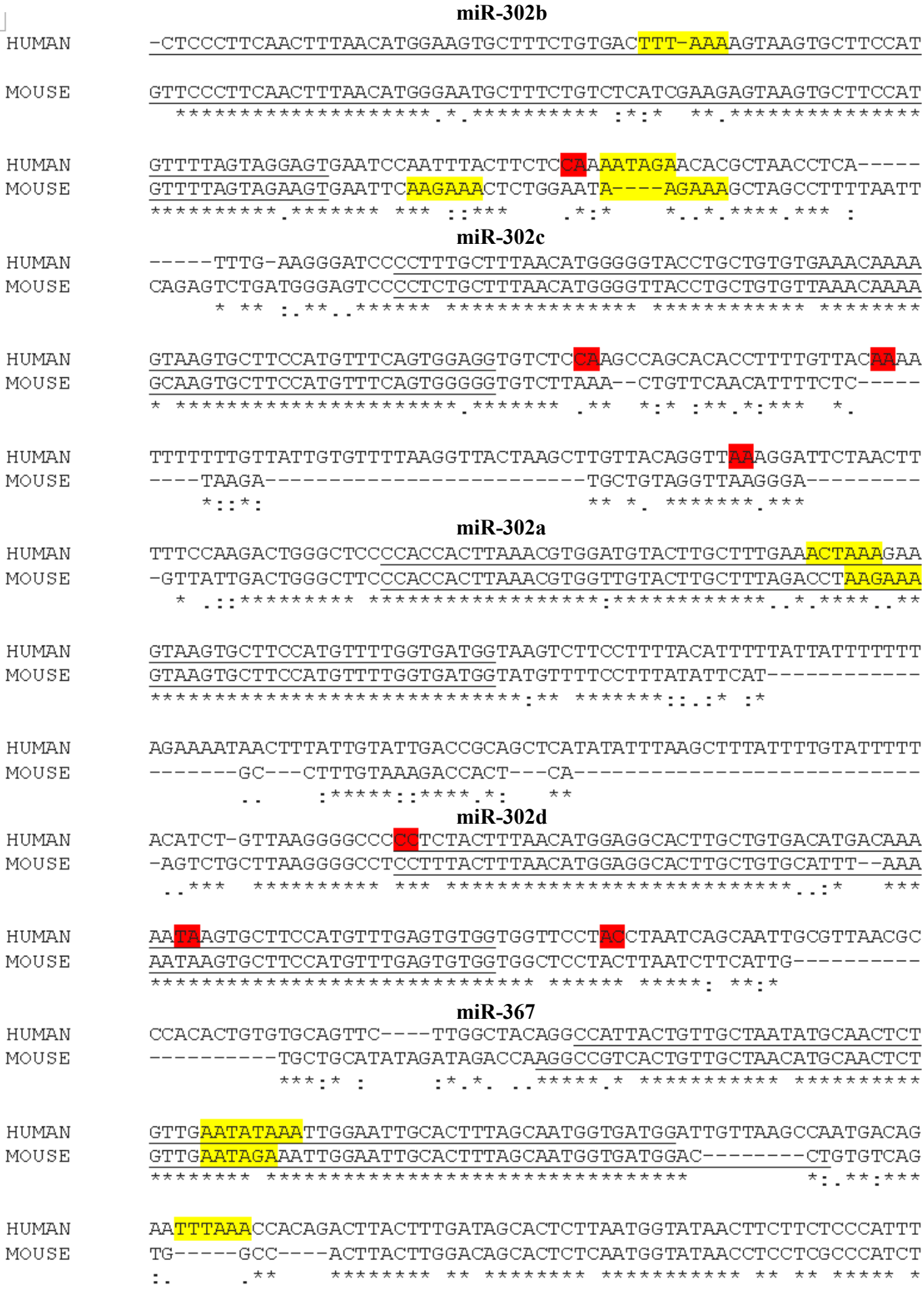
Poly(A) signal motifs
AATAAA
ATTAAA
AGTAAA
TATAAA
CATAAA
GATAAA
AATACA
AATATA
AATAGA
AAAAAG
ACTAAA
AAGAAA
AATGAA
TTTAAA
AAAACA
GGGGCT

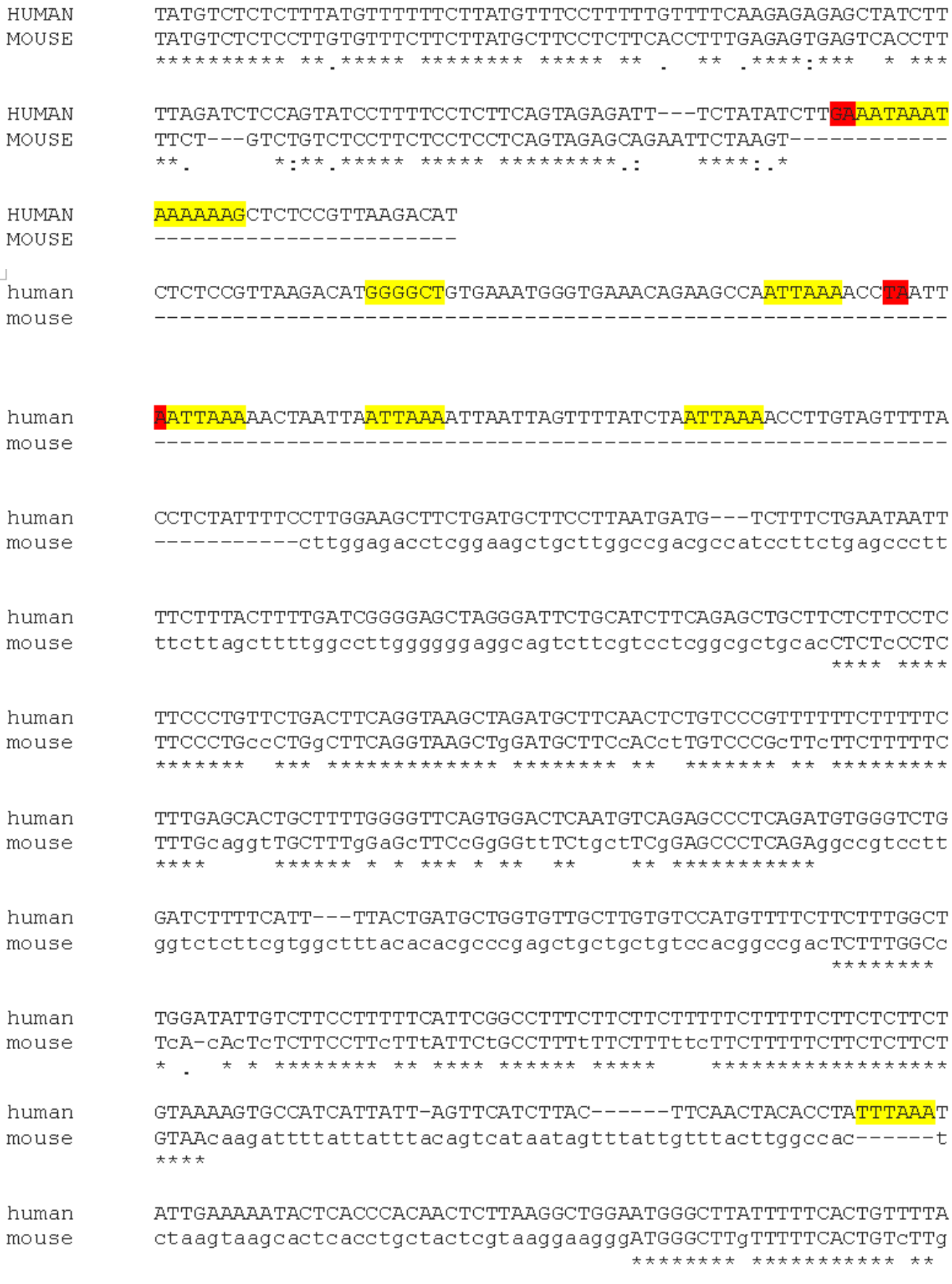
Supplementary Table S3 Genes upregulated by IFNa. The arbitrary numbers represent expression levels of the genes in untreated HT1080 cells (Untreated expression) and IFNa treated HT1080 cell (IFNa induced expression) Fold difference shows how many times more the genes are expressed when induced by IFNa. Identified by GENECHIP array analysis by Der et al., 1988.

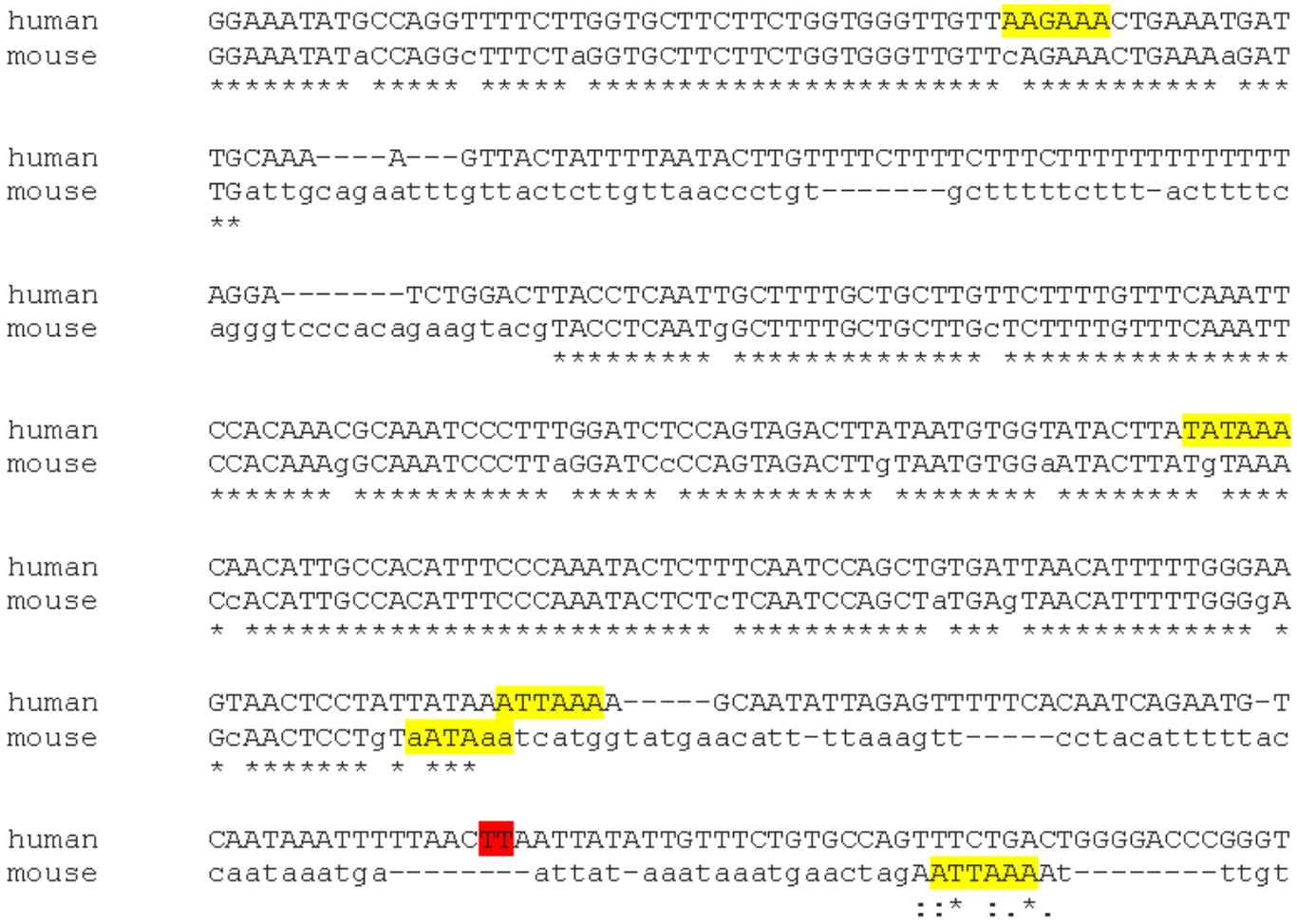
IFN induced gene	Untreated expression	IFNa induced expression	Fold difference
RING4	29	800	28
9.27	62	1,431	23.2
MxA	19	426	>21.3
6.16	−12	405	>20.2
p48	50	499	10.1
OAS1	18	584	>29.2
ISGI51	148	3,054	20.7
TRIM21	−34	204	>10.2



Supplementary figure S1 Alternative polyadenylation of miR-21. A) A Schematic of the genomic position of pre-miR-21 (blue line) and the 3'RACE sequencing results illustrating the two 3'ends of pri-miR-21 (black lines). B) Sequence of pre-miR-21 (underlined), the sequences in red indicate the position of the two poly(A) tails and the yellow sequence represents the PASs that aid the use of each poly(A) site. C) and D) the gels of the two 3'RACE reactions using different primers – show transcripts with two different poly(A) sites in primary miR-21 (the sequence of the bands is shown below the gels – the red sequence represents the primers used for the 3'RACE and the yellow sequence highlights the putative PAS which could cause polyadenylation).


miR-302b
HUMAN -CTCCCTTCAACTTTAACATGGAAGTGCTTTCTGTGACTTT-AAAAGTAAGTGCTTCCAT
MOUSE GTTCCCTTCAACTTTAACATGGGAATGCTTTCTGTCTCATCGAAGAGTAAGTGCTTCCAT
HUMAN GTTTTAGTAGGAGTGAATCCAATTTACTTCTCCAAAATAGAACACGCTAACCTCA-----
MOUSE GTTTTAGTAGAAGTGAATTCAAGAAACTCTGGAATA----AGAAAGCTAGCCTTTTAATT
miR-302c
HUMAN -----TTTG-AAGGGATCCCCTTTGCTTTAACATGGGGGTACCTGCTGTGTGAAACAAAA
MOUSE CAGAGTCTGATGGGAGTCCCCTCTGCTTTAACATGGGGTTACCTGCTGTGTTAAACAAAA
HUMAN GTAAGTGCTTCCATGTTTCAGTGGAGGTGTCTCCAAGCCAGCACACCTTTTGTTACAAAA
MOUSE GCAAGTGCTTCCATGTTTCAGTGGGGGTGTCTTAAA--CTGTTCAACATTTTCTC-----
HUMAN TTTTTTTGTTATTGTGTTTTAAGGTTACTAAGCTTGTTACAGGTTAAAGGATTCTAACTT
MOUSE ----TAAGA-----------------------TGCTGTAGGTTAAGGGA---------
miR-302a
HUMAN TTTCCAAGACTGGGCTCCCCACCACTTAAACGTGGATGTACTTGCTTTGAAACTAAAGAA
MOUSE -GTTATTGACTGGGCTTCCCACCACTTAAACGTGGTTGTACTTGCTTTAGACCTAAGAAA
HUMAN GTAAGTGCTTCCATGTTTTGGTGATGGTAAGTCTTCCTTTTACATTTTTATTATTTTTTT
MOUSE GTAAGTGCTTCCATGTTTTGGTGATGGTATGTTTTCCTTTATATTCAT-----------
HUMAN AGAAAATAACTTTATTGTATTGACCGCAGCTCATATATTTAAGCTTTATTTTGTATTTTT
MOUSE -------GC---CTTTGTAAAGACCACT---CA-------------------------
miR-302d
HUMAN ACATCT-GTTAAGGGGCCCCCTCTACTTTAACATGGAGGCACTTGCTGTGACATGACAAA
MOUSE -AGTCTGCTTAAGGGGCCTCCTTTACTTTAACATGGAGGCACTTGCTGTGCATTT--AAA
HUMAN AATAAGTGCTTCCATGTTTGAGTGTGGTGGTTCCTACCTAATCAGCAATTGCGTTAACGC
MOUSE AATAAGTGCTTCCATGTTTGAGTGTGGTGGCTCCTACTTAATCTTCATTG----------
miR-367
HUMAN CCACACTGTGTGCAGTTC----TTGGCTACAGGCCATTACTGTTGCTAATATGCAACTCT
MOUSE ----------TGCTGCATATAGATAGACCAAGGCCGTCACTGTTGCTAACATGCAACTCT
HUMAN GTTGAATATAAATTGGAATTGCACTTTAGCAATGGTGATGGATTGTTAAGCCAATGACAG
MOUSE GTTGAATAGAAATTGGAATTGCACTTTAGCAATGGTGATGGAC--------CTGTGTCAG
HUMAN AATTTAAACCACAGACTTACTTTGATAGCACTCTTAATGGTATAACTTCTTCTCCCATTT
MOUSE TG-----GCC----ACTTACTTGGACAGCACTCTCAATGGTATAACCTCCTCGCCCATCT

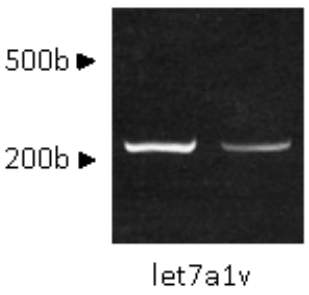

HUMAN TATGTCTCTCTTTATGTTTTTTCTTATGTTTCCTTTTTGTTTTCAAGAGAGAGCTATCTT
MOUSE TATGTCTCTCCTTGTGTTTCTTCTTATGCTTCCTCTTCACCTTTGAGAGTGAGTCACCTT
HUMAN TTAGATCTCCAGTATCCTTTTCCTCTTCAGTAGAGATT---TCTATATCTTGAAATAAAT
MOUSE TTCT---GTCTGTCTCCTTCTCCTCCTCAGTAGAGCAGAATTCTAAGT----------
HUMAN AAAAAAGCTCTCCGTTAAGACAT
MOUSE ----------
human CTCTCCGTTAAGACATGGGGCTGTGAAATGGGTGAAACAGAAGCCAATTAAAACCTAATT
mouse ----------
human AATTAAAAACTAATTAATTAAAATTAATTAGTTTTATCTAATTAAAACCTTGTAGTTTTA
mouse ----------
human CCTCTATTTTCCTTGGAAGCTTCTGATGCTTCCTTAATGATG---TCTTTCTGAATAATT
mouse ----------cttggagacctcggaagctgcttggccgacgccatccttctgagcccttt
human TTCTTTACTTTTGATCGGGGAGCTAGGGATTCTGCATCTTCAGAGCTGCTTCTCTTCCTC
mouse ttcttagcttttggccttgggggggaggcagtcttcgtcctcggcgctgcacCTCTcCCTC
human TTCCCTGTTCTGACTTCAGGTAAGCTAGATGCTTCAACTCTGTCCCGTTTTTTCTTTTTC
mouse TTCCCTGccCTGgCTTCAGGTAAGCTgGATGCTTCcACctTGTCCCGcTTcTTCTTTTTC
human TTTGAGCACTGCTTTTGGGGTTCAGTGGACTCAATGTCAGAGCCCTCAGATGTGGGTCTG
mouse TTTGcaggtTGCTTTgGaGcTTCcGgGGttTCtgctTCgGAGCCCTCAGAggccgtcctt
human GATCTTTTCATT---TTACTGATGCTGGTGTTGCTTGTGTCCATGTTTTCTTCTTTGGCT
mouse ggtctcttcgtggctttacacacgcccgagctgctgctgtccacggccgacTCTTTGGCc
human TGGATATTGTCTTCCTTTTTCATTCGGCCTTTCTTCTTCTTTTTCTTTTTCTTCTCTTCT
mouse TcA-cAcTcTCTTCCTTcTTtATTCtGCCTTTtTTCTTTttcTTCTTTTTCTTCTCTTCT
human GTAAAAGTGCCATCATTATT-AGTTCATCTTAC------TTCAACTACACCTATTTAAAT
mouse GTAAcaagattttattatttacagtcataatagtttattgtttacttggccac------t
human ATTGAAAAATACTCACCCACAACTCTTAAGGCTGGAATGGGCTTATTTTTCACTGTTTTA
mouse ctaagtaagcactcacctgctactcgtaaggaagggATGGGCTTgTTTTTCACTGTcTTg



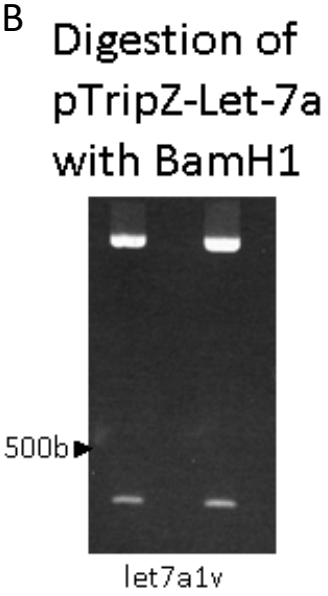
Supplementary figure S2 Human and mouse miR-302 cluster sequence alignment. Matching conserved sequences are indicated by the "*" below the nucleotides. The underlined sequences indicate the precursor miR-302b, 302c, 302a, 302d and 367 (top to bottom respectively). Sequences in yellow represent potential PASs (predicted by web-based tool: https://github.com/coggy/signals_in_sequence) and the red sequence annotate the multiple 3'ends established by 3'RACE and ABI 3730xl DNA sequencing.

Supplementary figure S3 PCR of miR-let-7a primary transcript, digestion of this insert post ligation into pTRIPz and sequencing of the insert and vector A) shows a PCR gel with miR-let-7a product, this DNA is then expressed in pTRIPz vector B) Shows the gel after digestion of the pTRIPz vector with BamH1, the miR-let-7a insert is represented by the smaller band. C) The sequence of a portion of the pTRIPz vector and the DNA insert.

A



B



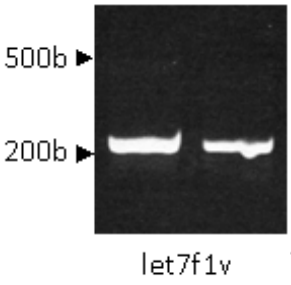
C

pTripZ-Let-7a sequence

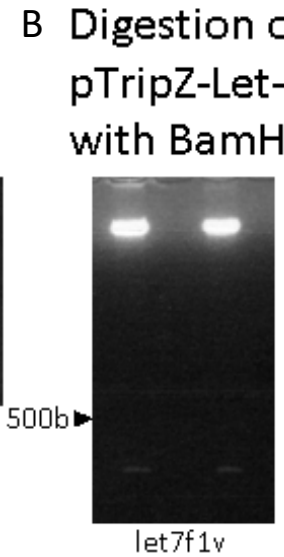
caccatgagcgagctgatcaaggagaacatgcacatgaagctgtacatggagggcaccgtga
acaaccaccacttcaagtgcacatccgagggcgaaggcaagccctacgagggcacccagacc
atgaagatcaaggtggtcgagggcggccctctcccccttcgccttcgacatcctggctaccagctt
catgtacggcagcaaagccttcatcaaccacacccagggcatccccgacttctttaagcagtcct
tccctgagggcttcacatgggagagaatcaccacatacgaagacgggggcgtgctgaccgcta
cccaggacaccagcttccagaacggctgcatcatctacaacgtcaagatcaacggggtgaactt
cccatccaacggccctgtgatgcagaagaaaacacgcggctgggaggccaacaccgagatgct
gtaccccgctgacggcggcctgagaggccacagccagatggccctgaagctcgtgggcggggg
ctacctgcactgctccttcaagaccacatacagatccaagaaacccgctaagaacctcaagatg
cccggcttccacttcgtggaccacagactggaaagaatcaaggaggccgacaaagagacctac
gtcgagcagcacgagatggctgtggccaagtactgcgacctccctagcaaactggggcacaga
taatcgatagttCATAAACAAATGCTGCACTACATCTCTTTAAGACAAGTTAAT
CTTAAATATTGTACAATTAGTTAACTGACTTTCTATCAGACCGCCTGGATG
CAGACTTTTCTATCACGTTAGGAAAGACAGTAGATTGTATAGTTATCTCCC
AGTGGTGGGTGTGACCCTAAAACTATACAACCTACTACCTCATCCCACAG
TGAAGAGAACATCCAGGGTGAATGGTGAAAAGGAATCACAGTTGAGCAC
CACAGGAAGGCTTTTTTTCCTCCTCGGTAATCCTGGTTTCCTGTGTGAGTG
ATAAGAAAAATAGTATT

Supplementary figure S4 PCR of miR-let-7f primary transcript, digestion of this insert post ligation into pTRIPz and sequencing of the insert and vector A) shows a PCR gel with miR-let-7f product, this DNA is then expressed in pTRIPz vector B) Shows the gel after digestion of the pTRIPz vector with BamH1, the miR-let-7f insert is represented by the smaller band. C) The sequence of a portion of the pTRIPz vector and the DNA insert.

A



B



C

pTripZ-Let-7f sequence

Caccatgagcgagctgatcaaggagaacatgcacatgaagctgtacatggagggcaccgtga
acaaccaccacttcaagtgcacatccgagggcgaaggcaagccctacgagggcacccagacc
atgaagatcaaggtggtcgagggcggccctctcccccttcgccttcgacatcctggctaccagctt
catgtacggcagcaaagccttcatcaaccacacccagggcatccccgacttctttaagcagtcc
ttccctgagggcttcacatgggagagaatcaccacatacgaagacgggggcgtgctgaccgct
acccaggacaccagcttccagaacggctgcatcatctacaacgtcaagatcaacggggtgaac
ttcccatccaacggccctgtgatgcagaagaaaacacgcggctgggaggccaacaccgagatg
ctgtaccccgctgacggcggcctgagaggccacagccagatggccctgaagctcgtgggcggg
ggctacctgcactgctccttcaagaccacatacagatccaagaaacccgctaagaacctcaag
atgcccggcttccacttcgtggaccacagactggaaagaatcaaggaggccgacaaagagac
ctacgtcgagcagcacgagatggctgtggccaagtactgcgacctccctagcaaactggggca
cagataatcgatagttGTCTTATTTCTCTGTGTCAGAATAAATGTATTTTTTTAT
TTTGATTTATGCTGATAATTTTATGTTGAAATTTTCTTTCGAAAGAGATTG
TACTTTCCATTCCAGAAGAAAACATTGCTCTATCAGAGTGAGGTAGTAGA
TTGTATAGTTGTGGGGTAGTGATTTTACCCTGTTCAGGAGATAACTATAC
AATCTATTGCCTTCCCTGAGGAGTAGACTTGCTGCATTATTTTCTTTTTATT
TAGATGATATTAAAACTCAGAAGAATTAATTTTGACATTTTGTATTTACAG
TTTATCAGT

Supplementary figure S5 PCR of miR-let-7af primary transcript, digestion of this insert post ligation into pTRIPz and sequencing of the insert and vector A) shows a PCR gel with miR-let-7af product, this DNA is then expressed in pTRIPz vector B) Shows the gel after digestion of the pTRIPz vector with BamH1, the miR-let-7af insert is represented by the smaller band. C) The sequence of a portion of the pTRIPz vector and the DNA insert.

