Characterisation of isomiRs in stem cells

Geok Chin Tan

Institute of Reproductive and Developmental Biology
Department of Surgery and Cancer
Faculty of Medicine
Imperial College London

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

Statement of Originality

All experiments included in this thesis were performed by me unless otherwise stated in the text.

Copyright Statement

Acknowledgements

I would like to thank my supervisor Dr Nicholas Dibb for giving me the opportunity to work in his lab and for all of his guidance and support throughout my PhD, without which this project would not have been possible. I am also very grateful to Dr Wei Cui for teaching me the technique of stem cell culture, her comments on my project related to stem cells and as a wonderful co-supervisor. I would like to also thank Professor Malcolm Parker for his supports and advise on academic and non-academic related subjects.

Many thanks to Elcie Chan for the generation of all the stem cell libraries which forms the platform for my project. My sincere thanks also to Gunter Meister for supplying the Argonaute antibodies, Leandro Castellano for the help in the design of RNA sponges, Laki Buluwela for the pTRIPz lentiviral vector and last but not least Alywn Dart from Charlotte Bevan group for the prostate cancer cell lines. To all IRDB members, especially the past and current members of the 5th and 1st floors, many thanks for all your help and advice over the years and for making my time at the IRDB so enjoyable.

My special thanks go to my wife and children, Pauline, Jonathan and Grace, for their continuous encouragement and supports. A special acknowledgement goes to the Malaysian Ministry of Higher Education, National University of Malaysia and Genesis Research Trust for funding my PhD scholarship and research grant, and for giving me the opportunity to continue exploring my scientific interest at a higher level.

Abstract

Since the inception of deep sequencing, isomiRs are consistently observed to be produced by most miRNA genes in a variety of cell types. Here I use northern blotting to show that isomiRs are not a sequencing artefact and I also observed that different cell lines and tissue types expressed distinctive isomiR patterns. All tested isomiRs could be immunoprecipitated with Argonaute proteins 1 or 2, indicating that they are functional.

IsomiRs with differences at the 5' end have a different seed sequence compared to the canonical/ annotated microRNA. Bioinformatics analysis predicts that 5' isomiRs will target large numbers of different mRNAs compared to their canonical counterpart and *vice versa*. These predictions were supported by my *in vitro* luciferase assays, which I used to establish that isomiR-9 has gained the ability to target DNMT3B and NCAM2 mRNA but has lost the ability to target CDH1. During this study I identified a number of new targets of miRNAs *in vitro*, all of which were confirmed by mutagenesis of the predicted target sites.

Moreover, I have made RNA sponge vectors that can distinguish between miR-9 and isomiR-9. The "isomiR-9 sponge" could specifically sequester isomiR-9 at a better efficiency than the canonical miR-9, which has just one base difference at the 5' end, and *vice-versa*. This adds further assurance that isomiRs can recognise different targets to canonical/ annotated microRNAs and also establishes a useful research tool for future studies.

Taken together, this study shows that isomiRs are capable of targeting 3' UTRs, can associate with Argonaute proteins and may have different target mRNAs to canonical mRNAs. I also discuss some examples of miRNA genes whose evolution is likely to have been influenced by isomiR production, which adds further support to the view that isomiRs are of biological and evolutionary importance.

Presentations and Publications

Tan GC, Sarkar R, Chandrashekran A, Cui W, Dibb NJ. "MicroRNA mediated reprogramming of fibroblasts to iPSCs by inducible lentiviral system" was presented as a poster in the Surgery and Cancer skill afternoon at Imperial College London on 11 January 2012.

Tan GC, Chan E, Sarkar R, Robinson S, Cui W, Dibb NJ. "IsomiRs are functional and have different set of target genes from their canonical microRNA" was presented as a poster in Cell Symposium: Functional RNAs in Spain on 2-4 December 2012.

Tan GC, Chan E, Sarkar R, Cui W, Molnar A, Meister G, Baulcombe D, Dibb NJ. "Canonical microRNAs and their isomiRs have different target genes" was presented as a poster in the "Biogenesis and turnover of small RNAs" meeting at the Royal Society, Edinburgh, UK on 15-17 January 2013. It was awarded the best poster prize by the Biochemical Society Transactions.

Tan GC, Dibb NJ. MicroRNA-induced pluripotent stem cells. Malays J Pathol 2012;34(2):167-168.

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).

Content

Abbreviations

ADAR	adenosine deaminases acting on RNA
Ago	Argonaute protein
APS	ammonium persulfate
bFGF	basic fibroblast growth factor
BMP	bone morphogenetic protein
bp	base pair
BSA	bovine serum albumin
BTG1	B-cell translocation gene 1
cDNA	complementary deoxyribonucleic acid
CDH1	Cadherin-1
CM	conditioned medium
CMV	cytomegalovirus
CNS	central nervous system
DAPI	4,6-diamidino-2-phenylindole
DGCR8	DiGeorge syndrome critical region 8
DMEM	Dulbecco's Modified Eagle Medium
DMNT3B	DNA (cytosine-5-)-methyltransferase 3 beta
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsRNA	double stranded RNA
ECL	enhanced chemiluminescence
EF1-α	elongation factor 1-alpha
EGF	epidermal growth factor
eIF4E	eukaryotic translation initiation factor subunit 4E
EMBL	European Molecular Biology Laboratory
FACS	fluorescent-activated cell sorting
FBS	fetal bovine serum
FGF	fibroblast growth factor
GFP	green fluorescent protein
HEK	human embryonic kidney
HMGA2	High-mobility group AT-hook2
hESCs	human embryonic stem cells
hMSCs	human mesenchymal stem cells
hNSCs	human neural progenitor stem cells
ICM	inner cell mass
iPSCs	induced pluripotent stem cells
kb	kilobase
klf4	kruppel-like factor 4
KO-DMEM	Knockout Dulbecco's Modified Eagle Medium
KSR	knock out serum replacement
Lefty1	Left-right determination factor 1
MBS	multiple binding sites
MEF	mouse embryonic fibroblast
miRNA	microRNA
miRNP	microribonucleoprotein complex
mRNA	messenger RNA

NCAM	neural cell adhesion molecule
NEAA	non-essential amino acids
NPSC	neural progenitor/stem cell
NSCs	neural stem cells
nt	nucleotide
Oct-4	octamer-binding transcription factor 4
OSKM	Oct4, Sox2, Klf4 and Myc
PAGE	polyacrylamide gel electrophoresis
PAR-CLIP	Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprepicitation
PAX6	paired box gene 6
P&S	penicillin and streptomycin
PBS	phosphate buffer saline
piRNA	PIWI interacting RNA
PLL	poly-L-Lysine
PTEN	Phosphatase and tensin homolog
PTGS	Posttranscriptional gene silencing
RFP	red fluorescent protein
RIPA	radioimmunoprecipitate assay
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
Rock1	Rho-associated, coiled-coil containing protein kinase 1
RT	reverse transcription
SDS	sodium dodecyl sulfate
siRNA	small interfering RNA
Sox2	SRY (sex determining region Y)-box 2
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
TGF	transforming growth factor
TRBP	trans-activator RNA (Tar) binding protein
TRE	tetracycline response element
UTR	untranslated region

List of Figures

List of Tables

List of supplementary tables

List of supplementary figures

Chapter 1 Introduction

1.1 The discovery of RNA interference

In 1998, Fire and colleagues reported that the injection of double stranded RNA caused the degradation of mRNA encoded by the gene *unc-22* of the small nematode *Caenorhabditis elegans*. They termed this effect RNA interference and showed that it was sequence specific and effective at concentrations far lower than the target *unc-22* mRNA, indicating the involvement of amplification effect in this process (Fire et al., 1998).

In parallel, David Baulcombe's group (Hamilton et al., 1999) discovered posttranscriptional gene silencing (PTGS) in plants. Both viral infection and transgenic expression in plants can induce PTGS, which targets both cellular and viral mRNA. It was inferred that PTGS operates through the generation of small RNA molecules of 21 to 25 nucleotides, which are now known as small interfering RNAs (siRNAs). Consequently, work on both animals and plants together revealed a highly conserved mechanism of RNA interference that had evolved at least in part for combating viral infection.

Two biochemistry groups were able to recapitulate RNA interference in cell free extracts from Drosophila cells, which led to the establishment of three phases of the RNA interference reaction: cleavage of a long dsRNA into shorter dsRNA segments by Dicer; the loading of single stranded RNA into the RISC (RNA-induced silencing complex) and the targeting and degradation of mRNA by this complex (Tuschl et al., 1999; Hammond et al., 2000; see below). Subsequently mutations of C.*elegans* that

confirm resistance to RNA interference were found to disrupt genes that encoded components of RISC (Fire, 2007).

1.2 The discovery of microRNAs

Lee et al., (1993) identified two overlapping transcripts of the *lin-4* gene of *C. elegans*, of approximately 22 and 61 nts that inhibited the expression of *lin-14* through complementarity to the 3' untranslated region (UTR) of lin-14 mRNA. The 61 nucleotides molecule can also fold into a double-stranded "hairpin" (Lee et al., 1993). They suggested that lin-4 inhibits translation of lin-14 through an antisense RNA-RNA interaction. A similar conclusion was made by another group who reported that there were 7 conserved sites in the 3' UTR of lin-14, which were complementary to a portion of lin-4 RNA (Wightman et al., 1993).

Subsequently, it was shown that *lin-4* and a second gene *let-7* acted in a sequential stage specific expression pattern that regulates the timing of *C.elegans* development. The *let-7* gene encodes a 21-nucleotide RNA that is complementary to the 3' UTR of genes *lin-14, lin-28, lin-41, lin-42* and *daf-12*. *Let-7* is expressed at the adult but not embryonic stage. *Let-7* was also identified in humans, fruit flies, chickens, frogs, zebrafish, molluscs and sea urchins and the binding site in its target was conserved in some of these organisms (Reinhart et al., 2000; Pasquinelli et al., 2000). In 2001, three groups published their discovery of large number of similar small RNA molecules, referred to as microRNAs, in *C.elegans* and subsequently mouse (Lee et al., 2001; Lau et al., 2001; Lagos-Quintana et al., 2001, 2002). Remarkably, the cytoplasmic cellular machinery that mediates RNA interference is also responsible for

the generation of microRNAs (miRNAs). As described below, there are additional processing steps for miRNA generation that take place in the nucleus.

1.3 Non-coding RNAs

MicroRNAs belong to one of the classes of non-coding RNAs. Non-coding RNAs are functional RNAs that do not translate into protein. They comprise: transfer RNA (tRNA); ribosomal RNA (rRNA); small nucleolar RNA (snoRNA); microRNA; small interfering RNA (siRNAs); small nuclear RNA (snRNA); piwi-interacting RNA (piRNA), and long ncRNA (Figure 1.1; Morin et al., 2008). MicroRNAs are the most abundant and to date, approximately 1600 pre-miRNAs and 2040 mature human miRNAs have been identified (miRBase, August 2012, Griffith-Jones et al., 2004; Kozomara et al., 2011). MicroRNAs are about 19-25 nucleotides in length and are now known to have important post-transcriptional roles in almost every cellular process in eukaryotes. These processes include the regulation of developmental timing and signalling pathways, apoptosis, metabolism, myogenesis and cardiogenesis, brain development, and human pathologies like viral diseases, genetic disorders and cancer (Esquela-Kerscher et al., 2006; Kloosterman et al., 2006a; Shi et al., 2008).

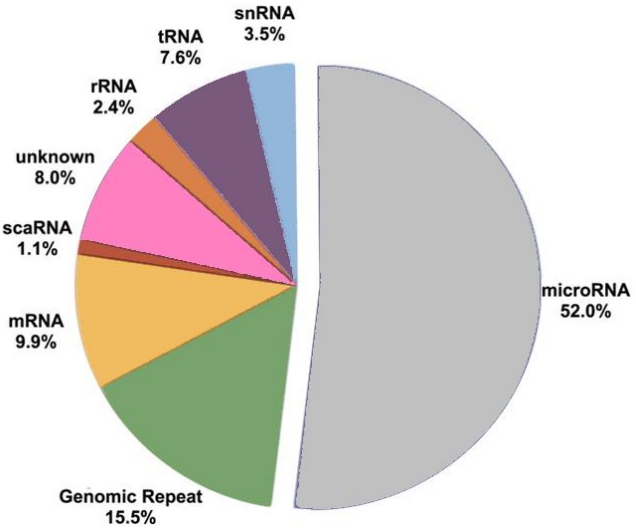


Figure 1.1
Distribution of sequence count of 8 major classes of small RNAs in the hESCs small RNA library. They are represented as a fraction of the total sequences that has at least one perfect alignment to the human reference genome. MiRNA represents the most abundantly expressed class, i.e., >50% of the 8 classes of small RNAs. snRNA – small nuclear RNA; tRNA – transfer RNA; rRNA – ribosomal RNA; scaRNA – Small Cajal body-specific RNA; mRNA – messenger RNA. Reproduced from Morin et al., (2008).

1.4 MicroRNA

1.4.1 Biogenesis

MicroRNA genes can be located between genes as well as within the intron or exon regions of other genes in the human genome (Figure 1.2). The miRNA genes are transcribed into primary miRNA (pri-miRNA) by RNA polymerase II (Lee et al., 2004; Cai et al., 2004) or in some instances polymerase III (Borchert et al., 2006). These primary transcripts range from hundreds to thousands of nucleotides in length and can encode multiple precursor miRNAs (Breving et al., 2010).

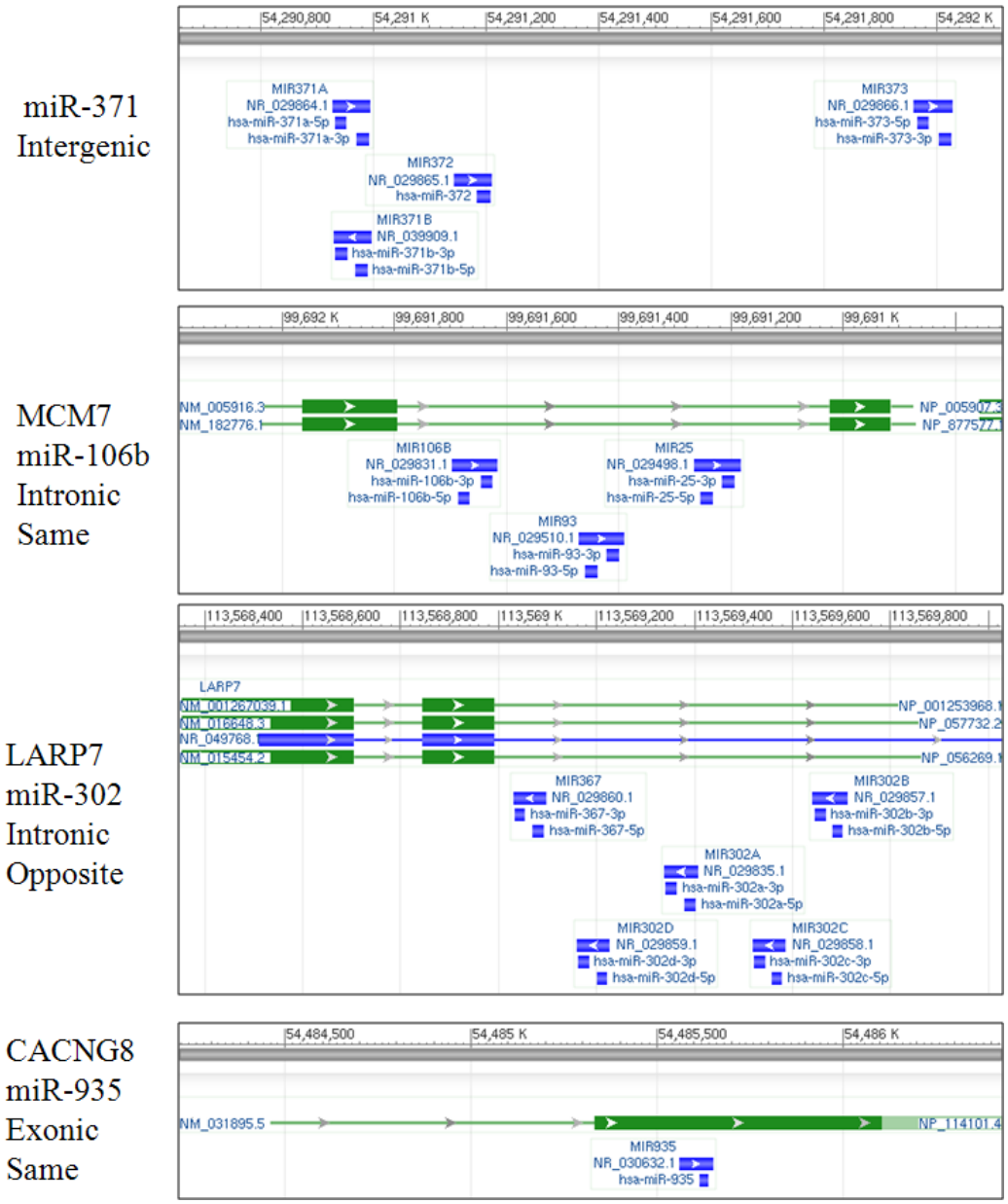


Figure 1.2 Examples of genomic location of miRNA genes, also indicated is whether the direction of transcription is the same or opposite of the host gene.

Pri-miRNA undergoes processing by Drosha, an RNase III endonuclease (Lee et al., 2003). Drosha forms a microprocessor complex with DGCR8 (DiGeorge syndrome critical region gene 8), which is called Pasha in Drosophila and PASH-1 in C. *elegans* (Yeom et al., 2006; Han et al., 2009; Breving et al., 2010). This complex binds to stem loops within pri-miRNA and can excise and release precursor miRNA (pre-

miRNA) (Basyuk et al., 2003; Lee et al., 2003) (Figure 1.3). DGRC8 assists Drosha to cleave ~ 11 bp away from the ssRNA-dsRNA junction (Han et al., 2006). The hairpin of pre-miRNA is ~ 70 nt in length. Not all miRNAs are dependent upon Drosha-mediated processing, these include miRNAs called mirtrons that are processed by splicing (Berezikov et al., 2007; Chan et al., 2007).

The pre-miRNA is then transported into the cytoplasm by Exportin-5 (Yi et al., 2003; Murchison et al., 2004). Here, the pre-miRNA is further processed by Dicer, also an RNase III endonuclease, resulting in the generation of a ~ 22 nt miRNA-miRNA* duplex (Grishok et al., 2001; Ketting et al., 2001), leaving the 5' phosphate and 2 nt 3' overhang characteristic of processing by an RNase III.

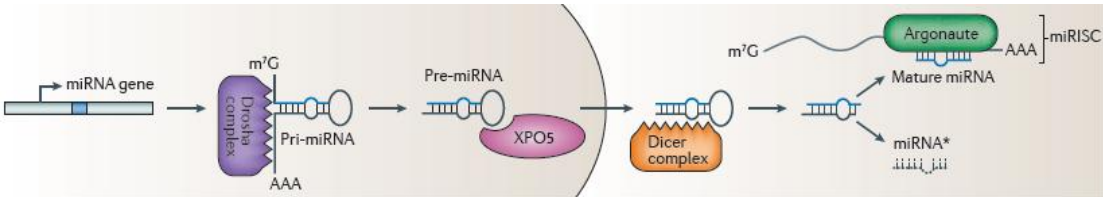


Figure 1.3 Biogenesis of miRNA.
MiRNA genes are transcribed into primary miRNA (pri-miRNA) transcripts that undergo processing by Drosha. The resulting hairpin precursor miRNAs (pre-miRNAs) are transported to the cytoplasm by exportin 5 (XPO5 on figure). The Dicer complex removes the loop region from pre-miRNAs, and one strand of the resulting duplex is bound by Argonaute to form an miRNA-induced silencing complex (miRISC), which targets mRNAs for regulation. While the other strand, which is often called the star strand (miRNA*), is degraded. Taken from Pasquinelli, 2012. Copyright permission obtained from author and Nature Publishing Group.

1.4.2 The complexity of miRNA regulation

The regulation of miRNA biogenesis is under transcriptional, post-transcriptional and feedback loops controls. Studies have shown that miRNAs expression differ in developmental stages and tissue types. Therefore, precise control of miRNA biogenesis is crucial in the maintenance normal cellular function (Kim et al., 2009).

Transcriptional control

Various Pol II-associated transcription factors are involved in the control of miRNA transcription. For instance, in the studies of myogenesis, Rao et al., (2006) found that myogenin and myoD1 bind to regions upstream to muscle specific miRNAs (miR-1 and miR-133 cluster) and likely to regulate their expression. Meanwhile, Chen et al., (2006) showed that miR-1 and miR-133 are regulated by serum response factor (SRF).

Post-transcriptional control

The primary miRNA let-7 is found in both undifferentiated stem cells and differentiated cells. However, interestingly mature let-7 is only identified in differentiated cells, because it is under post-transcriptional control. Many studies have established that the RNA binding protein Lin28 is responsible for the inhibition of let-7 maturation (Nam et al., 2008; Kim et al., 2009; Lehrbach et al., 2010). RNA editing is another post-transcriptional control mechanism, where adenine is altered to inosine by adenine deaminases (Yang et al., 2006; Kawahara et al., 2007). In addition, there are other proteins that are involved in the post-transcriptional regulation of miRNAs (Siomi et al., 2010; Guil et al., 2007; Davis et al., 2008; Trabucchi et al., 2009).

Feedback loop control

Two types of feedback loops have been observed: (1) single negative feedback and (2) double negative feedback. Martinez and colleagues analysed the transcription factors that are associated with miRNAs and predicted targets of miRNAs which are also transcription factors and found that many of the transcription factors are repressed by the same miRNA that activated it (Martinez et al., 2008). Let-7 and lin28 are an

example of a double negative feedback loop. Lin28 blocks let-7 biogenesis, whereas let-7 suppresses lin28 protein synthesis (Kim et al., 2009).

1.4.3 Mechanism of target selection

Mature miRNA is held at both ends by an Argonaute protein in the RNA-induced silencing complex (RISC) that guides the miRNA towards target mRNAs resulting in reduced protein production, via mechanisms that are still under investigation, namely mRNA destabilisation, deadenylation or translational repression. In animals, miRNAs usually form incomplete complementary duplexes with their mRNA targets, which are normally located at the 3' UTR. The canonical site of target recognition is the "seed region" which is located at nucleotides 2 to 7 or 2 to 8 at the 5' end of the miRNA and often has perfect complementarity pairing to the target mRNA (Bartel, 2009). Atypical sites have also been described such as the interaction between let-7 and lin-41 in C. elegans. In this example, imperfect pairing of the seed region at the 5' end is compensated for by pairing at the 3' end (Vella et al., 2004; Bartel, 2009; Pasquinelli, 2012). Recently, central pairing (nucleotides 4 to 15) been shown to lead to Ago2 mediated target cleavage (Shin et al., 2010).

1.4.4 Argonaute protein

Human Ago1, Ago3 and Ago4 genes are located on chromosome 1, whereas the Ago2 gene is on chromosome 8 (Hock et al., 2008). In addition, Ago2 is the only one with 'slicer' activity and therefore capable of cleaving target mRNA (Liu et al., 2004). Ago2 also mediates the action of interfering RNA. Argonaute protein consists of 3 domains: PIWI, MID and PAZ. The PAZ domain recognises the 2 nucleotides at the 3' overhang of the miRNA duplex that is produced by Dicer (Cenik et al., 2011). The

5' monophosphate of the miRNA is buried within the MID domain, while the 3' end is exposed at the PAZ domain. Based on the structure of Ago2, it favours binding of small RNAs that begin with an adenosine (A) or uridine (U) at the 5' end. These features of Ago proteins might be important in the loading of miRNA into the RISC complex.

1.4.5 Star strands

In general, one strand of the RNA duplex denoted as the guide strand preferentially accumulates (Schwarz et al., 2003). This strand is often assumed to be the dominant functional product that is incorporated into the RISC to direct translational repression or degradation of mRNA (Hutvagner, 2005). The opposite strand is referred to as the passenger or star strand (miR*) and usually is less frequently sequenced (Lagos-Quintana et al., 2002; Aravin et al., 2003; Lim et al., 2003). In some instances both miR and miR* are equally expressed (Kloosterman et al., 2006b; Stark et al., 2007).

In the cloning and sequencing data of Chan (unpublished; see chapter 5) of human embryonic stem cells (hESCs), miR-302a* was detected as the dominant strand, in fact, it was sequenced 20 times more than the guide strand of miR-302a, suggesting that it might be a functional strand. Indeed, recent studies have demonstrated that the star strand of a precursor miRNA can be associated with Argonaute proteins (Ghildiyal et al., 2010) and the inhibitory effect of miR* has been shown in cultured cell and transgenic animals (Okamura et al., 2008).

Many deep sequencing studies indicate that the dominant strand of the mature miRNA can switch in different tissues and at different developmental times (Ro et al., 2007;

Ruby et al., 2007; de Wit et al., 2009; Chiang et al., 2010). The predicted targets of miR and miR* differ significantly (Griffiths-Jones et al., 2011). The process of switching between miR and miR* in different tissues is referred to as arm switching and is suggested to be a fundamental mechanism in the evolution of miRNA function (de Wit et al., 2009; Griffith-Jones et al., 2011).

In Drosophila, reports showed that star strands are associated with Ago proteins (Ghildiyal et al. 2010; Okamura et al., 2008). Recently, miR-24-2* was found to be preferentially expressed in MCF7 breast cancer cells where it might have a tumour suppression role. Ectopic expression of miR-24-2* resulted in reduced cell survival through the suppression of protein kinase C alpha (PKCα) (Martin et al., 2012). Interestingly, in a study where either miR-10a or miR-10a* was transfected into Group B coxsackievirus (RLuc-CVB3) infected HEK cells, only miR-10a* was found to up-regulate the biosynthesis of CVB3. The authors suggested that miR-10a* might be involved in viral pathogenesis (Tong et al., 2013). Bioinformatics analysis showed that a substantial fraction of miRNA* species are stringently conserved over vertebrate evolution, with greatest conservation in their seed regions (Yang et al., 2011). It was also found that the 3' UTR target sites that match the seed sequence of miRNA* species are under demonstrable selective conservation (Okamura et al., 2008).

1.5 IsomiRs

1.5.1 The identification of isomiRs

The advent of high-throughput deep sequencing has led to the detection of large numbers of miRNAs (Morin et al., 2008; Lee et al., 2010; Cloonan et al., 2011). In these miRNA libraries, miRNAs encoded by the same gene frequently exhibited variation in length from the canonical sequence annotated in miRBase, as a result of an addition or deletion at the 5' or 3' ends or both. These variants were termed as isomiRs (Neilsen et al., 2012). They can be categorised into 5' isomiRs, 3' isomiRs and mixed 5' and 3' isomiRs.

To date, isomiRs have been detected in a variety of cell lines and cancers such as hESCs, endothelial cells, 293T cells, prostate cancer, gastric cancer, breast cancer and leukemic cells (Morin et al., 2008; Bar et al., 2008; Kuchenbauer et al., 2008; Lipchina et al., 2011; Voellenkle et al., 2012; Watahiki et al., 2011; Li et al., 2012; Chang et al., 2012). Currently, the isomiR databases that are available in the web include miRBase (Griffith-Jones et al., 2004), YM500 (Cheng et al., 2012), Hood lab (Institute of Systems Biology, 2012), miRGator v3.0 (Cho et al., 2012; Narry Kim lab) and SeqBuster (Pantano et al., 2010) (Table 1.1).

No	Name of the database	Web-link	Reference
1	miRBase	http://www.mirbase.org/	Griffith-Jones et al., 2004
2	miRGator v3.0	http://mirgator.kobic.re.kr/	Cho et al., 2012
3	SeqBuster	http://code.google.com/p/seqbuster/ (need to download software)	Pantano et al., 2010
4	Hood lab	http://hood.systemsbiology.net/cgi-bin/isomir/find.pl	Institute of systems biology (ISB) 2012
5	YM500	http://ngs.ym.edu.tw/ym500/	Cheng et al., 2012

Table 1.1 List of isomiR databases

Despite the large number of isomiRs that have been detected, there is relatively little experimental proof that they are functional (Fernandez-Valverde et al., 2010; Burroughs et al., 2011; Fukunaga et al., 2012; Humphreys et al., 2012; Lloren et al., 2013). Dominant isomiRs were found to be differentially expressed across *Drosophila melanogaster* development and tissues (Fernandez-Valverde et al., 2010). Burrough et al., (2011) showed that isomiRs can associate with Ago protein. Using an assay that shows Ago2 cleaves target mRNA at nucleotide positions 10 and 11 from the 5' end of small RNA (Beitzinger et al., 2007), Azuma-Mukai et al., (2008) showed that 5' isomiRs were able to participate in Ago2-mediated RNA cleavage.

It was subsequently shown that altering the Dicer partner proteins could change the choice of the cleavage site, producing isomiRs with different target specificities and function in Drosophilia (Fukunaga et al., 2012). Recently, Lloren et al., (2013),

analysed gene expression by microarray after transfecting miR-101 and isomiR-101 into SH-SY5Y cells. They found that isomiR-101 has an overall weaker inhibitory effect than miR-101 and largely targeted the same set of genes. Only two of the genes that were found to be down-regulated in isomiR-101 transfected cells were not regulated by miR-101.

1.5.2 Origin of isomiRs

There has been some concern that isomiRs are simply sequencing artefacts. However, "spike in" synthetic RNA oligonucleotide experiments indicate that isomiR identification far exceeds error rates (Wyman et al., 2011). 3' isomiRs are the most frequently observed isomiRs (Wyman et al., 2011; Lee et al., 2010; Burroughs et al., 2011; Newman et al., 2011). Although not as frequent, 5' isomiRs were also detected. This heterogeneity in length is thought to arise in part from imprecise cleavage by Drosha or Dicer, which would be expected to give rise to equivalent numbers of 3' or 5' isomiRs that will otherwise match the parent gene and for this reason are referred to as templated (Neilsen et al., 2012). Non-templated refers to post-transcriptional modifications such as A to I editing that may not match the parent gene. The excess of 3' isomiRs that are observed are thought to arise by trimming, adenylation or uridylation (Han et al., 2011; Liu et al., 2011; Wyman et al., 2011; Heo et al., 2012). In addition, Liu et al., (2011) showed that knockdown of *Nibbler* (a 3' to 5' exoribonuclease) was accompanied by loss of some 3' isomiRs.

The 3' ends of miRNA extend from the PAZ domain of the Argonaute protein and are therefore available to exonucleolytic attack (Schirle et al., 2012; Elkayam et al., 2012), whereas the 5' ends of miRNAs are buried within the MID domain and are protected

(Neilsen et al., 2012). Wu et al., (2009) showed that alternative processing of primary miRNA by Drosha and DGCR8 can generate precursor miRNA with or without 5' end variation. Eventually, these precursor miRNAs may undergo 3' end modification which produces mature miRNAs having 5', 3' or mix variations (Wu et al., 2009).

In principle, 5' isomiRs have different seed regions to their canonical miRNA and therefore could have a different subset of target genes. Although miRBase (August 2012) has included isomiRs in their database, miRNAs are still annotated as a single mature miRNA sequence.

1.6 Target prediction programs

As a result of the use of cloning and high throughput deep sequencing, thousands of miRNAs have been discovered. Target prediction programs have been created to attempt to generate predictions of miRNA targets based on genome wide computational search for miRNA and mRNA UTR complementary sites. The initial clue came from the observation that lin-4 complementarity to multiple conserved sites to the 3' UTR of lin-14 mRNA is required for the repression of lin-14 (Lee et al., 1993; Wightman et al., 1993; Bartel., 2009).

The most significant contribution to target recognition was the identification of Watson-Crick miRNA-mRNA perfect complementarity of 6 to 8 bp at the 5' end of miRNA and 3' UTR of mRNA (Lewis et al., 2003; Rakewsky et al., 2004). As a result, the initial method of target prediction was based on complementarity of the miRNA to the target site and the predicted free energy of the miRNA-mRNA duplex (Rakewsky

et al., 2004; Rakewsky et al., 2006). Subsequently, a new generation of miRNA target prediction programs emerged in 2005 that are based on more extensive bioinformatics analysis using cross-species comparison (Lewis et al., 2005).

In TargetScan (Lewis et al., 2005), miRNA targets are predicted by searching for Watson-Crick base pairing matches between the seed region and 3' UTRs that are conserved via whole genome alignment. Based on a prediction study, more than 5300 human genes were predicted targets of miRNA, which represented 30% of the human gene set (Lewis et al., 2005). Figure 1.4 illustrated the conserved predicted miRNA target sites in the 3' UTR of NCAM2 (long red arrows). Intriguingly, there are a few other conserved sites (short yellow arrows) that are not predicted target sites of any canonical/ annotated miRNA (Figure 1.4). These sites could be undiscovered target sites of isomiRs or perhaps targets of RNA binding proteins. Another related example is the mRNA encoded by the BACE2 gene, there are 3 highly conserved sites and one of these is a predicted target site of let-7. Notably, one of the remaining two conserved sites is a target site of isomiR-9 (Figure 1.4, see Chapter 3).

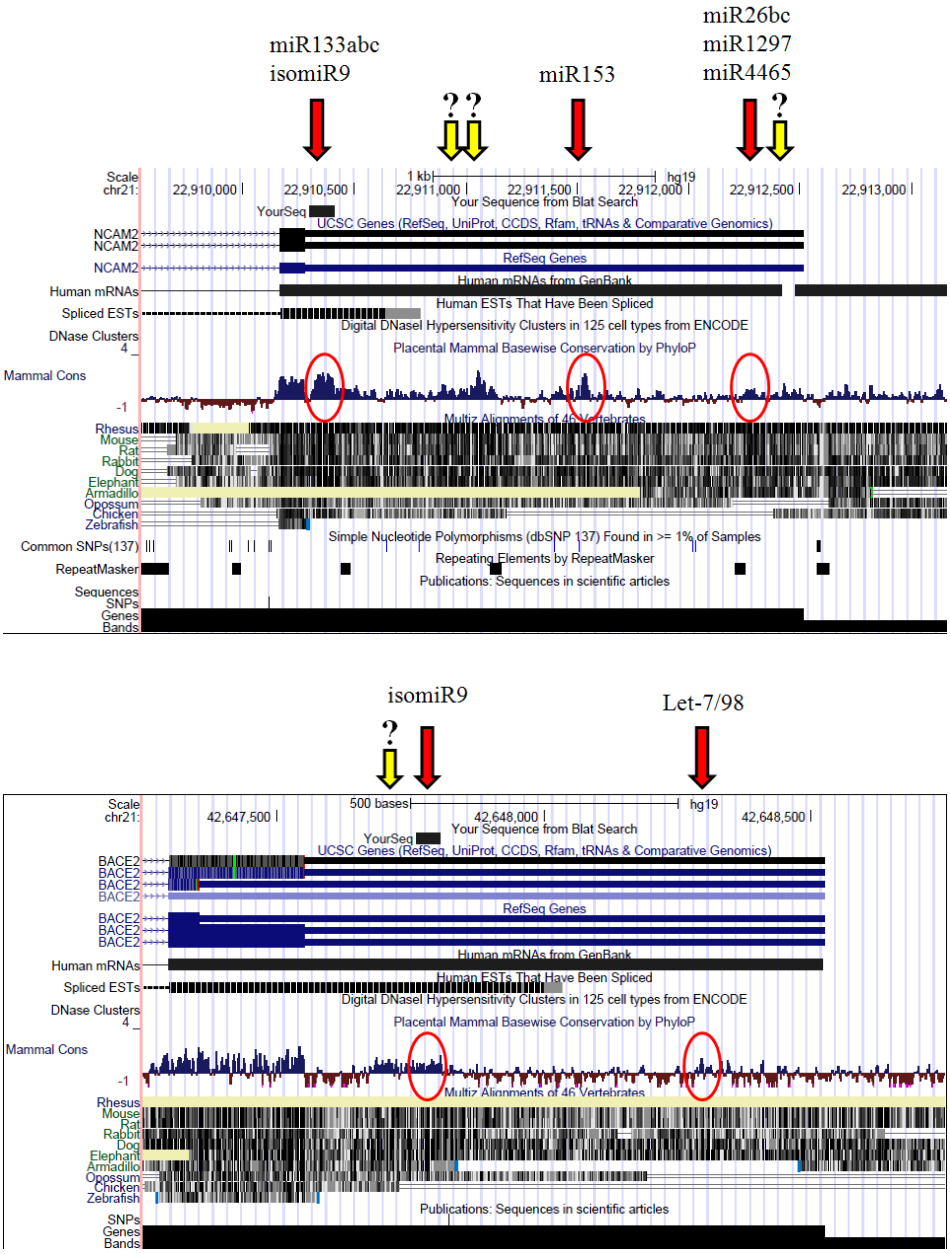


Figure 1.4 Conserved miRNA target sites in the 3' UTR of NCAM2 and BACE2. Long red arrows represent known miRNA target sites. Short yellow arrows denote conserved sites that are not known to be a target of any canonical/ annotated miRNA. Reproduced and modified from USCS genome browser.

Table 1.2 lists some of the miRNA target prediction programs that are available on the web. These programs differ in their selection criteria like the stringency of seed complementarity and measurement of base pairing stability and selection of different UTR sequence (Bartel, 2009; Ritchie et al., 2009). Different prediction databases

predict different sets of target genes, for example the predicted targets generated from TargetScan and MiRanda overlap by 39.5% only (Ritchie et al., 2009). The differences in prediction might result from the used of different 3' UTR sequence in the prediction programs (Bartel, 2009). So far, only a handful of these predictions have been experimentally validated (Rosa et al., 2009; Barroso-delJesus et al., 2011).

No	Databases	Description	Website	References
1	TargetScan	Stringent seed pairing and conservation ranking based on target interaction types	http://targetscan.org	Lewis et al., 2005
2	RNAhybrid	A tool for finding minimum free energy hybridisations of a long (target) and a short (query) RNA	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/welcome.html	Rehmsmeier et al., 2004
3	PITA	Moderately stringent seed pairing, predicted pairing stability by measuring free energy between miRNA and target UTR of your choice	http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html	Kertesz et al., 2007
4	PicTar	Stringent seed pairing Overall predicted pairing stability UTR from UCSC database	http://pictar.mdc-berlin.de	Krek et al., 2005
5	miRNA - Target Gene Prediction at EMBL	Stringent seed pairing Genome comparison across insect species (Drosophilia)	http://www.russell.embl.de/miRNAs	Stark et al.2003, Brennecke et al., 2005
6	miRanda	Moderate stringent seed pairing The target sites predicted by miRanda are scored for likelihood of mRNA downregulation using mirSVR, a regression model that is trained on sequence and contextual features of the predicted miRNA:mRNA duplex	http://www.microrna.org	Enright et al., 2003, John et al., 2004
7	MicroTar	Based on mRNA sequence complementarity (3'-UTR seed matches) and RNA duplex energy prediction and uses the RNAlib library from the Vienna RNA package	http://tiger.dbs.nus.edu.sg/microtar/	Thadani et al., 2006
8	DIANA-microT	Complementarity, conservation human and mouse	http://diana.cslab.ece.ntua.gr/microT/	Maragkakis et al., 2009

Table 1.2 List of target prediction tools. All target prediction programs depend on seed target complementarity. In addition, some include conservation and/or free energy measurement. Adapted from Bartel, (2009).

A large scale approach using mass spectrometry to measure protein level reduction after miRNA transfection has revealed that a 7mer-A1 match, which has only 6 complementary base pairs (Figure 1.5) was more effective than a complete 1-7mer Watson-Crick match (Baek et al., 2008). Therefore, an A in the UTR which aligns with miRNA nucleotide 1 favours miRNA-mediated protein down-regulation, even when the A in the UTR does not participate in a Watson-Crick interaction. This also explains the preferential conservation of an A at position 1 of UTR target sites (Lewis et al., 2005). In animals, the currently recognised canonical types of miRNA target sites that involve the seed region include 7mer-A1, 7mer-m8 and 8mer (Figure 1.5). Other atypical types include 3' supplementary and compensatory sites that have pairing at the seed region as well as additional pairing at the 3' end to enhance target recognition, usually at position 13 to 16 nucleotides from 5' end (Bartel, 2009). Recently, other atypical sites have also been discovered, known as central pairing that has 11 to 12 contiguous Watson-Crick pairing at the centre but lacks pairing at both the seed region and 3' end (Shin et al., 2010).

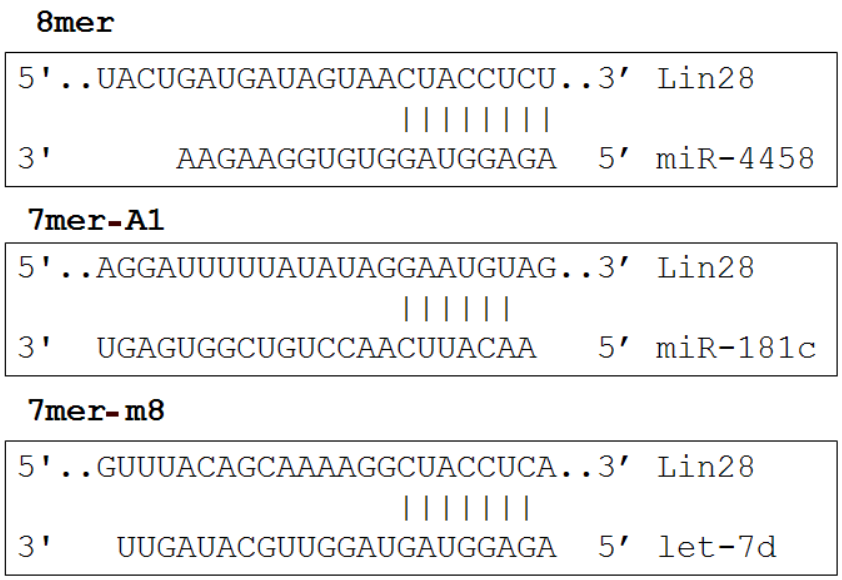


Figure 1.5 Examples of the canonical type of miRNA – mRNA target interaction. Vertical line represents Watson-Crick base pairing.

Farh and colleagues reported that the predicted non-conserved target sites outnumbered the conserved ones by ten to one (Farh et al., 2005). Using reporter assays, the authors showed that a large proportion of the non-conserved target sites can function. However, analysis of mRNA and miRNA expression profile revealed that 3' UTR with non-conserved target sites are most often found in genes that are expressed in tissue where the complementary miRNA is absent (Farh et al., 2005).

Grimson et al., (2007) showed there were other features in the 3' UTR that increased target site efficacy such as: (1) target sites that are positioned away from the centre of long UTRs. One possible explanation is that sites at the centre would have the opportunity to fold from segments at either sides but sites near the end would not (Bartel, 2009). (2) An AU-rich nucleotide composition near the site and (3) positioning within the 3' UTR at least 15 nt from the stop codon. In fact, by at large conserved 7-mer target sites were preferentially found in the above mentioned areas (Gaidatzis et al., 2007; Grimson et al., 2007; Majoros et al., 2007).

UTRs may contain multiple targeting sites. The repression response in a 3' UTR with multiple sites is nearly the same as that observed in the sum of each site independently (Grimson et al., 2007; Nielsen et al., 2007), showing that there is an additive effect in these cases (Doench et al., 2003). In theory, miRNAs might also act synergistically. Bartel, (2009) suggested that cooperative miRNA function could provide a mechanism where repression can become more sensitive to small changes in miRNA expression levels, which greatly enhances their regulatory effect. It was also found that repression was enhanced when the distance between two target sites was between 13 and 35 nucleotides (Saetrom et al., 2007).

Other factors that could influence targeting include (1) the presence of naturally occurring decoy mRNA that might compete with 3' UTR in miRNA binding, thus reducing the amount of free miRNA (Franco-Zorrilla et al., 2007; Poliseno et al. 2010). (2) Competing RNA-binding proteins like deadend 1 (DND1) that might shield target sites from miRNA RNA-induced silencing complex (miRISC) binding (Kedde et al., 2007). (3) There are factors that might associate with the RISC and influence its regulation either positively or negatively, for example NHL2 and meiotic P26 (Mei-P26) (Neumüller et al., 2008; Hammell et al., 2009).

In addition to 3' UTRs, experiments showed that miRNA targeting can also occur at the 5' UTR and open reading frame (ORF) (Kloosterman et al., 2004; Lytle et al., 2007). Indeed, a large number of targets in the ORF were observed by computational genome wide analyses (Farh et al., 2005; Lewis et al., 2005; Lim et al., 2005; Easow et al., 2007: Grimson et al., 2007; Baek et al., 2008). However, ORF targeting is probably less frequent and less effective than 5' UTR and 3' UTR targeting, probably due to displacement of the silencing complex at this position by the translation machinery (Bartel, 2004; Bartel, 2009).

In addition, Argonaute protein could also influence the processing and loading of miRNAs. For example Ago2 favours the binding of small RNAs that begin with an adenosine (A) or uridine (U) at the 5' end (Frank et al., 2010; Cenik et al., 2011).

1.7 MicroRNA sponges

MicroRNA sponges were first described by Ebert et al., (2007) and Franco-Zorrilla et al., (2007). These sponges are decoy mRNAs that compete with endogenous mRNA for base pairing with miRNAs (Figure 1.6). The first naturally occurring RNA sponge was discovered in a plant (Franco-Zorrilla et al., 2007). The authors reported that a non–protein coding gene called induced by phosphate starvation (IPS1) from *Arabidopsis thaliana* contained a motif with sequence complementarity to miR-399. Interestingly, the pairing is interrupted by a mismatched loop at the expected miRNA cleavage site (nucleotides 10-11), which protected it from cleavage by Argonaute. Hence, IPS1 RNA is not cleaved but instead sequesters miR-399. In addition, IPS1 overexpression resulted in increased accumulation of the miR-399 target PHO2 mRNA which encodes an E2 ubiquitin conjugase–related protein.

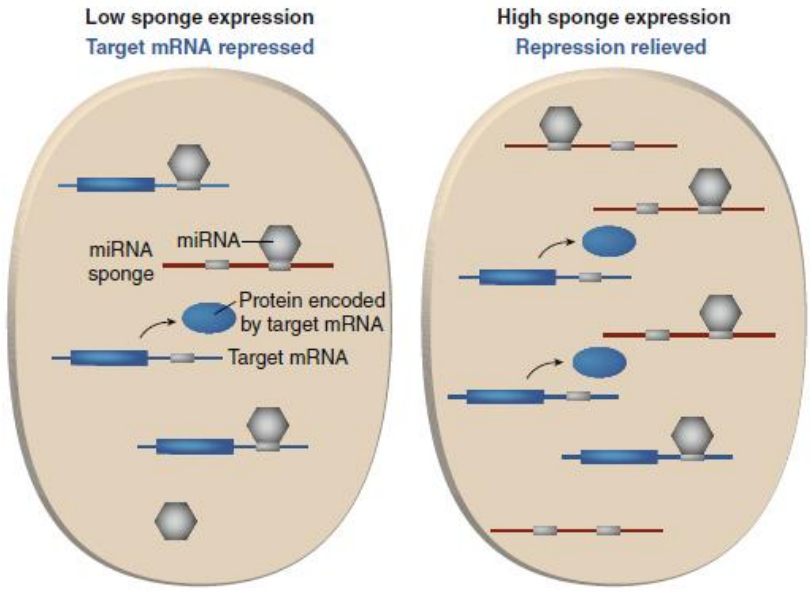


Figure 1.6 RNA sponge competes with target mRNA for binding with miRNA. Sponge RNAs (in red) contain binding sites (small grey rectangle) for miRNA of interest (grey hexagons). Target mRNAs are in blue. Blue oval represents protein output. Taken from Ebert et al., (2010a). Copyright permission obtained from author and Elsevier.

Naturally occurring miRNA sponges have also been reported in mammalian cells. These sponges are transcribed from pseudogenes (Poliseno et al., 2010). PTENP1 is an example of a naturally occurring RNA sponge. Its 3' UTR has similar conserved binding sites with that of the 3' UTR of PTEN. PTENP1 was found to be selectively lost in human cancer and appears to act as a decoy for miRNAs that target PTEN. It was found that knockdown of PTENP1, increased the abundance of PTEN targeting miRNAs, which led to a reduction of PTEN mRNA and protein levels (Poliseno et al., 2010). Based on alignment studies, other possible pseudogenes have been identified that could act as decoys for miR-145, the miR-1 family, miR-182, miR-143 and let-7 which are thought to regulate OCT4, CX43, FOXO3B and KRAS1P respectively (Poliseno et al., 2010). These natural miRNA decoys have been termed as "competitive endogenous RNAs" (ceRNAs) (Poliseno et al., 2010; Cesana et al., 2011; Karreth et al., 2011; Sumazin et al., 2011).

For cells that are difficult to transfect, viral vectors can be used to stably express RNA sponges (Ebert et al., 2010a, b). Haraguchi et al., (2009) described another type of transgenic expression RNA sponge termed TuD RNAs or ''tough decoy''. Their prototype decoy consisted of a stem loop hairpin with the miRNA binding site located at the single stranded loop region. After comparing various models of their tough decoys, the most effective one has two multiple binding sites which are flanked by two stem structures (Figure 1.7). It was also found that these RNA decoys were stable and could achieve long term suppression of miRNA (Haraguchi et al., 2009). As some miRNAs have been found to be very stable and to have *in-vivo* half lifes of more than a week (van Rooij et al., 2007; Bail et al., 2010), RNA sponges might be an effective way to sequester and thereby inhibit miRNA activity.

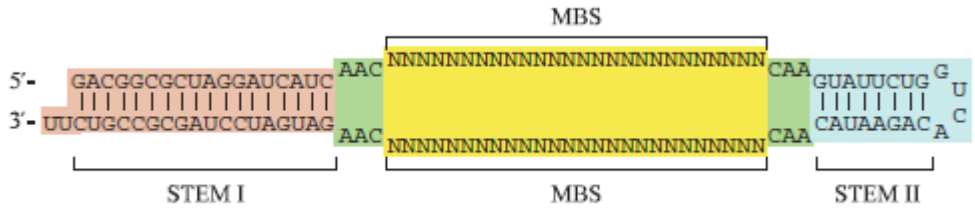


Figure 1.7 TuD RNA or tough decoy.This decoy RNA contains two multiple binding site (MBS) regions, which are flanked by two stem structures through 3-nt linker. Taken from Haraguchi et al., (2009).

1.8 MicroRNA and Stem Cells

1.8.1 MicroRNA and human embryonic stem cells

Embryonic stem cells (ESCs) are derived from the inner cell mass of the blastocyst stage of an embryo (Martin, 1981; Thomson et al., 1998). They are pluripotent and therefore have the ability to differentiate into any type of specialised cells in the three embryonic germ layers (endoderm, ectoderm and mesoderm) and can also replicate indefinitely in the undifferentiated state (Martinez et al., 2010). More importantly, hESCs have a normal karyotype, maintain high telomerase activity, and exhibit remarkable unlimited expansion in culture. Hence, hECSs are a useful *in-vitro* system to study the mechanisms underlying human development (Odorico et al., 2001).

A network of transcriptional factors and RNA binding proteins have been identified, known as "stemness factors" that are involved in the maintenance of stem cell identity. These factors include Oct4, Sox2, Nanog, Lin28 and Klf4 (Marson et al., 2008). Card et al., (2008) showed that Oct4 and Sox2 bind to the conserved promoter region of the miR-302 cluster and regulate its expression. In addition, the miR-302 cluster might also be involved in cell cycle regulation by its repression of cyclin D1 (Card et al., 2008). Similarly to Oct4, the miR-302 cluster was found to be down-regulated upon

differentiation (Card et al., 2008). Inhibition of miR-302 resulted in downregulation of pluripotency markers, whereas overexpression of miR-302 lead to upregulation of these genes (Rosa et al., 2009). Similarly, Barroso-delJeus et al., (2008) also found that there were conserved binding sites for Oct4, Sox2 and Rex1 upstream to the miR-302 cluster indicating that they might be regulators of the cluster.

It was also found that let-7 induced stem cell differentiation through the repression of multiple stemness factors include Lin28, Sall4 and c-Myc, where let-7 binding sites were found in their 3' UTRs (Melton et al., 2010). The global loss of miRNAs in DGCR8 deficient ESCs resulted in defects in proliferation and differentiation (Gangaraju et al., 2009). In addition, Dicer-deficient mice die at early stages of development (Martinez et al., 2010). Therefore, maintenance of self-renewal and induction of differentiation of ESCs is tightly regulated by miRNAs (Figure 1.8).

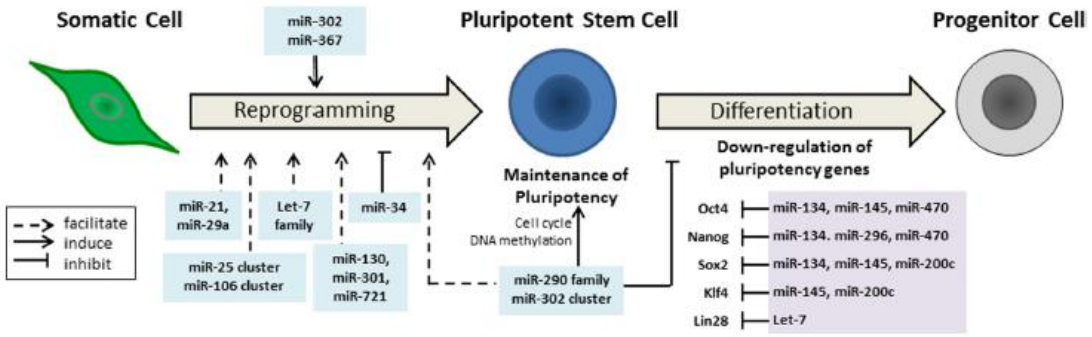


Figure 1.8 Regulation of self-renewal and differentiation by miRNAs. The left part of the figure shows miRs that facilitate, directly induce, or inhibit reprogramming of iPS cells. The right part summarizes miRs that were shown to control ESC maintenance and differentiation. Taken from Heinrich et al., (2012). Copyright permission obtained from author and Wolters Kluwer Health.

1.8.2 Reprogramming using the miR-302 cluster

Cloning and deep sequencing of hESCs consistently identified the miR-302 cluster as the most abundant and specific miRNA in stem cells (Suh et al., 2004, Bar et al., 2008, Morin et al., 2008; Lipchina et al., 2011; Chan, unpublished). MiR-302 is a polycistronic cluster that houses 5 precursor miRNAs, i.e. miR-302b, miR-302c, miR-302a, miR-302d and miR-367. In the human genome, it is 688 nt in length and located in intron 8 of the Larp7 gene in chromosome 4 (see Chapter 5). Interestingly, it is possible to reprogram a differentiated cell back to its unspecialised state, also known as an induced pluripotent stem cell (iPSC). This can be achieved by introducing stemness genes, namely *oct4, sox2, klf4* and *c-myc* (OSKM) transcription factors (Takahashi et al., 2006). Recently, it was found that the miR-302 cluster alone can reprograme both mouse and human fibroblasts to iPSCs with high efficiency (1-10%) (Anokye-Danso et al., 2011; see Chapter 5).

Several groups have addressed the mechanism of somatic cell reprogramming by miRNAs (Rosa et al., 2011; Lin et al., 2011; Hu et al., 2013). Lin et al., (2011) suggested that the mechanism of reprogramming by the miR-302 cluster involves targeted suppression of four epigenetic regulators including Lysine-specific demethylase 1 (LSD1 also known as KDM1 or AOF2), Lysine-specific histone demethylase 2 (AOF1), MECP1-p66 and methyl CpG binding protein 2 (MECP2), leading to global demethylation. As global demethylation naturally occurs in 2 stages of development, i.e., (1) during early embryogenesis and (2) at the initial stage of gametogenesis, the authors suggested that global demethylation can reset the cell back to its pluripotent state (Lin et al., 2011).

Studies have also shown that the miR-302 cluster is involved in the maintenance of pluripotency and is involved in the regulation of a number of cell signalling pathways including TGFb/nodal signalling and cyclin D1 regulation (Rosa et al., 2009; Lipchina et al., 2011; Wang et al., 2008; Card et al., 2008; Subramanyam et al., 2011; Sun et al., 1999).

The combination of target prediction, miRNA perturbation and PAR-CLIP experiments has given great insight into the targets of the mir-302 cluster. Using these assays, Lipchina and colleagues (2011) identified 146 high confidence targets of miR-302 cluster. Furthermore, inhibition of the miR-302 cluster reduced proliferation whereas overexpression increased proliferation of stem cells (Lipchina et al., 2011).

Generating disease-specific or patient-specific iPSCs provides the opportunity to study the diseases in an *in vitro* situation with greater flexibility and to gain mechanistic insight into the disease (Bellin et al., 2012). iPSCs can also be used for drug screening and development of patient-specific therapy and for the study of rare genetic disorders. In addition, it might be useful for the exploration of cell-based and gene repair therapies (Figure 1.9; Robinton et al., 2012). For example, iPSC models have been used to study cardiomyocytes with Type 1 long QT syndrome and using this model it was found that treatment with propanolol, a β-adrenergic receptor blocker, attenuated the QT phenotype. Meanwhile, Itzhaki et al., (2011) found that nifedipine, a calcium channel blocker improved Type 2 long QT syndrome phenotype. Agarwal et al., (2010) studied the biology of telomerase using an iPSC model of dyskeratosis congenital, a disorder of telomere maintenance. In addition, other iPSC disease models that have been reported include Alzheimer's disease (Israel et al.,

2012), Huntington's disease (Camnasio et al., 2012), Parkinson's disease (Soldner et al., 2009), Timothy syndrome (Yazawa et al., 2011), Pompe's disease (Huang et al., 2011), spinal muscular atrophy (Ebert et al., 2009) and familial dysautonomia (Lee et al., 2009).

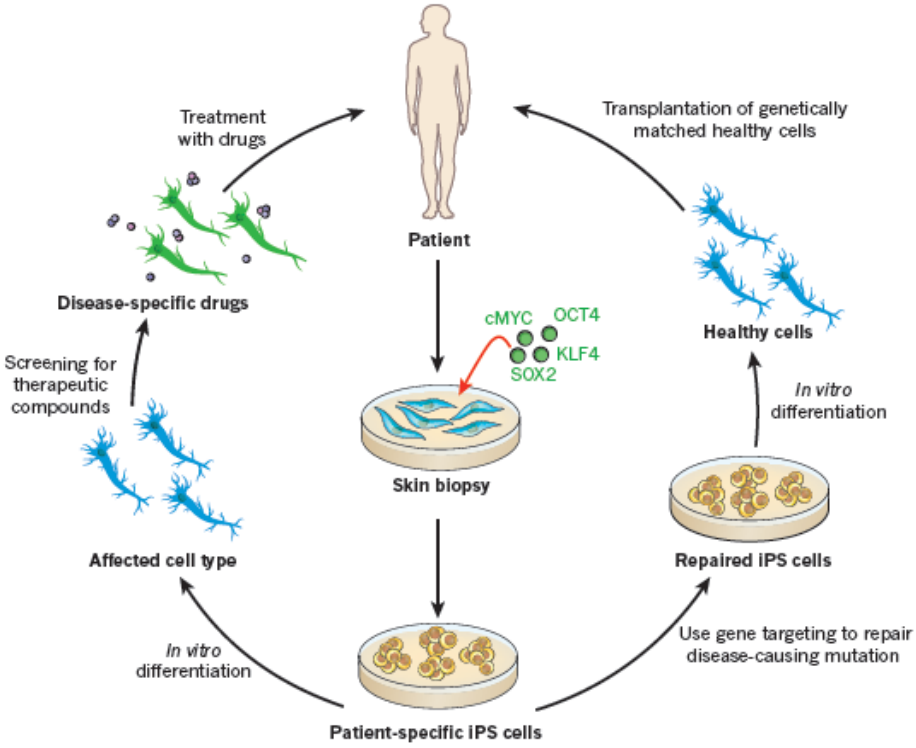


Figure 1.9 Application of iPSCs in a patient specific model. Patient-specific iPS cells - in this case derived by ectopic co-expression of transcription factors in cells isolated from a skin biopsy and used in one of the two pathways. Taken from Robinton et al., (2012). Copyright permission obtained from author and Nature Publishing Group.

All of the above studies would be helped by improvements in iPSC reprogramming efficiency. Therefore, studies are necessary to find better techniques to improve reprogramming efficiency. My aim was to determine if miR-302 cluster can reprogram human fibroblast back to its pluripotent state and to identify targets of miR-302a*.

1.8.3 MicroRNAs and neural progenitor/ stem cells

Neural stem cells (NSC) are multipotent cells that can differentiate into cells of the central nervous system (CNS) such as neurons, astrocytes, and oligodendrocytes (Alvarez-Buylla et al., 2002). NSCs can be derived from embryonic stem cells (Gerrard et al., 2005) or adult nervous system and can be cultivated *in-vitro* (Bonnamain et al., 2012), in the presence of growth factors like bFGF (Vescovi et al., 1993) and EGF (Reynolds et al., 1992). Transplantation of NSCs or recruitment of endogenous adult NSCs might be a potential strategy for the treatment of spinal cord injury (Ronaghi et al., 2010). Moreno-Manzano et al., (2009) reported that transplantation of ependymal stem progenitor cells that were derived from adult rat spinal cord that suffered a traumatic lesion lead to a functional motor recovery.

In cloning and deep sequencing of NSC, a subset of miRNAs was noted to be highly expressed and some were highly specific to NSC (Lipchina et al., 2011; Chan, unpublished). In both Lipchina et al., (2011) and Chan sequencing results, miR-9 was one of the top 3 most abundant and specific miRNAs in NSC. MiR-9 is a highly conserved miRNA and expressed primarily in the CNS (Kapsimali et al., 2007). The human genome has three miR-9 genes termed hsa-miR-9-1, hsa-miR-9-2 and hsa-miR-9-3, which encode an identical mature miR-9 (5p) and miR-9* (3p). The hsa-miR-9-1 gene is located in the intron 2 of C1orf61 gene in chromosome 1. The hsa-miR-9-2 is located in the exon of the LINC00461 gene in chromosome 5. The hsa-miR-9-3 gene is located in an intergenic region, although it partially overlaps with a non-coding RNA (LOC254559) on chromosome 15.

Remarkably, expression of miR-9/9*, miR-124 and neuroD2 converted human fibroblasts to neurons (Yoo et al., 2011). Expressing neuroD2 alone did not produce neurons but its inclusion enhanced this process. Expression of miR-9 and miR-124 was enough to cause a reduction in proliferation and to induce neuron-like morphology, but the efficiency was low (Yoo et al., 2011). Bonev et al., (2011) reported that depletion of miR-9 reduced neuronal differentiation, both at the forebrain and hindbrain in *Xenopus Tropicalis*.

Le and colleagues (2009) used retinoid acids to induce the differentiation of neuroblastoma cells into neuron-like cells. In the process, they measured miRNA levels by microarray and northern blotting, and identified 6 miRNAs that were consistently upregulated during the process of differentiation, namely miR-7, miR-124a, miR-125b, miR-199a, miR-199a* and miR-214. Subsequently, the authors showed that ectopic expression of miR-124a and miR-125b significantly increased the percentage of differentiated cells with neurite outgrowth. Using a microarray, they found 388 genes that were repressed by ectopic expression of miR-125b. Out of these 388 genes, 164 were targets of prediction programs. Ten target genes of miR-125b were validated by reporter assays. These genes are involved in metabolism, proliferation and apoptosis (Le et al., 2009).

MiR-9 was reported to have an association with neurological disorders (Yuva-Aydemir et al., 2011). It has been reported that alcohol increased miR-9 expression in supraoptic nucleus neurons and striatal neurons in an adult rat brain (Pietrzykowski et al., 2008). Increased levels of miR-9 were also found post-mortem, in the brains of patients with Alzheimer disease (Lukiw et al., 2007). In contrast, miR-9 was

downregulated in cerebral ischemia due to middle cerebral artery occlusion in rats (Jeyaseelan et al., 2008).

It is certain that miRNAs play an essential role in the maintenance of pluripotency and differentiation of human embryonic stem cells. Identifying the specific targets of miRNAs during hESC differentiation will help to elucidate the regulation of this complex mechanism.

1.9 Project Aims

1. To determine if isomiRs are functional and whether 5’ isomiRs can inhibit the expression of different mRNAs compared to the canonical/ annotated miRNA.

2. To test if it is possible to inhibit specific isomiRs by using sponge vectors.

3. To identify targets of miR-302a* and to determine if miR-302a* is important for the induction of pluripotent stem cells from somatic cells.

Hypothesis

As 5' isomiRs have different seed region to their canonical or annotated counterparts, my hypothesis is that 5' isomiRs could have different sets of target genes.

Chapter 2 Materials and Methods

2.1 Cell culture

2.1.1 General cell culture

All culture dishes (Corning, Costar), flasks (Corning), and serological plugged pipettes (Corning, Costar) used were suitable for sterile tissue culture. Unless otherwise stated, all cell lines used in the experiment were cultured in D10 media (Dulbeco's Modified Medium (DMEM) (Invitrogen, Gibco) supplemented with 10% (v/v) heat inactivated Foetal Bovine Serum (FBS) (PAA Laboratories), 50 U/ml penicillin/streptomycin (Invitrogen, Gibco) and 200 μM glutamine (Invitrogen, Gibco). Experiments were carried out in sterile condition, in a Class II flow cabinet and all cells were maintained at 37°C in 5% CO_2.

Generally, cells were passaged twice weekly or once they reached approximately 80% confluency. Prior to incubating with 0.25% trypsin (Invitrogen, Gibco) for 5 minutes at 37^{o}C in 5% CO_2, cells were washed in PBS. Subsequently, the cells were re-suspended in D10 media and centrifuged at 1000 rpm for 5 minutes. Then, the cell pellet was resuspended in D10 and plated to the required density.

2.1.2 Freezing cell lines

After washing with PBS, adherent cells were incubated with 1 ml of 0.25% trypsin in a 6-well plate. Then, 9 mls of D10 media was added to inactivate the trypsin. The cells were centrifuged at 1000 rpm for 5 minutes and re-suspended in D10 media with 10% dimethyl sulfoxide and promptly aliquoted into 1 ml cryotubes. The cells were

frozen slowly in a cryo freezing container containing isopropyl alcohol at -80°C for at least 48 hrs. Subsequently, they were stored either in -80^{o}C or in liquid nitrogen.

2.1.3 Production of mouse embryonic fibroblast-conditioned medium (MEF-CM) for hESCs culture

Mouse embryonic fibroblasts (MEF) were grown and expanded in D10 media to passages 3 or 4 depending on the speed of cell growth. Cells were then trypsinised and collected into 50ml Falcon tubes and counted. Followed by irradiation at 40 Grays (4000 rads), centrifuged at 800rpm for 4 minutes and plated at 18.8 x 10^6 cells into gelatin-coated T225 flasks with D10 media. Subsequently, 150 ml KSR media (KO DMEM and KO serum replacement, supplemented with 4ng/ml Fibroblast Growth Factor basic (FGF2) (Peprotech))(Knockout™ DMEM is a basal medium from Invitrogen, optimized for growth of undifferentiated embryonic and induced pluripotent stem cells and Knockout™ Serum Replacement from Invitrogen, is a defined, serum-free formulation optimized to grow and maintain undifferentiated ES cells in culture) was added to replace D10 media the next day. Collection was carried out continuously for 7 days. The collected media, called mouse embryonic fibroblast-conditioned media (MEF-CM) was stored at -80°C. Upon use, MEF-CM was thawed in water-bath at 37^{o}C and L-glutamine and P/S before added being filtered and kept at 4°C.

2.1.4 Preparation of matrigel coated plates

5 mls of stock matrigel (Invitrogen, Gibco) was slowly thawed at 4°C overnight and diluted with 5 mls of cold KO-DMEM. This solution was aliquoted, 1 ml working volume into 15-ml tube on ice and stored at -20°C. Upon use, matrigel was slowly

defrosted at 4°C and finally diluted with 14 mls of cold KO-DMEM. 1 ml per well of the diluted matrigel was plated onto a 6-well plate, and incubated overnight at 4°C. In case of urgency, the incubation time could be shortened to an hour at room temperature before use.

2.1.5 Human embryonic stem cells culture

H1, H7 and T5 (transgenic H1 with Oct4-EGFP) hESCs were cultured in MEF-CM supplemented with 4-8 ng/ml FGF2 in matrigel coated plates and media changed daily. Cells were routinely passaged at a 1:3 dilution after treatment with 200U/ml collagenase IV (Invitrogen) and mechanic dissection.

2.1.6 Freezing and resuscitating hESCs

hESCs were harvested similar as routine propagation except during mechanical dissection step 1x freezing mix (KO serum replacement and 10% DMSO) was added instead of MEF-CM. 1 ml aliquots were frozen overnight at -80°C in cryo freezing container before being transferred to liquid nitrogen for long-term storage. Frozen cells were revived by thawing rapidly at 37°C and resuspended in 10 mls of MEF-CM. Cells were centrifuged at 800rpm for 5 minutes to removed the DMSO, and then plated on a matrigel coated 6-well plate containing MEF-CM with FGF2.

2.1.7 Neural progenitor/stem cell differentiation from hESCs

hESCs was differentiated to neural lineage following the published protocol (Gerrard et al., 2005). Briefly, confluent hESCs were split with EDTA/ PBS in 1:5 ratios into culture dishes coated with poly-L-lysine/laminin and cultured in N2B27 media (DMEM/F12 / neurobasal media (1:1) (Invitrogen), 100x N-2 supplement (Invitrogen),

50x B-27 supplement (Invitrogen), L-glutamine and P/S) supplemented with 100ng/ml mouse recombinant noggin (R&D systems). At this stage, cells were defined as passage 1 (P1). Medium was changed every other day. Cells of P1 and P2 were split by collagenase IV into small clumps and continuously cultured in N2B27 medium plus noggin until formation of neural progenitor/ stem cells at P3.

2.1.8 Culture of neural progenitor/stem cells (NSCs)

Preparation of poly-L-lysine/laminin

Poly-L-Lysine (PLL) was diluted 1 in 6 with phosphate buffer solution (PBS) and then 1ml was added per well of a 6-well plate. Subsequently, it was incubated in hood at room temperature for 1 hour. Mouse laminin (Sigma) was diluted with 6mls of PBS to a final concentration of 20 μg/ml, plated 1ml per well of a 6-well plate and incubated overnight at 4°C.

Culture of neural progenitor/stem cells

NSCs were disassociated into single cell by TrypLE express (Invitrogen) and cultured in N2B27 media supplemented with 20ng/ml FGF2 and/or 20ng/ml Epidermal Growth Factor (EGF) (Peprotech).

2.2 Luciferase assay

All the constructs that were used for luciferase assays were listed in supplementary figure 2.3, and a full list of the primers used for cloning are given in table 2.6. Reporter vectors were constructed by the insertion of predicted 3'UTR containing miRNA target sites downstream of the gene encoding for firefly luciferase in the pGL3-Control vector (Promega). Assays were performed in 24-well plates using

HEK293 cells seeded at 50,000 cells per well a day prior to the time of transfection in D10 media without phenol red, and incubated for 48 hrs. Renilla luciferase was used as the internal control. A green fluorescent protein (GFP) expressing vector driven by the EF-1 alpha promoter was used to enable visualization of transfection efficiency. All experiments were performed in triplicates, as follows:

For each well of a 24-well plate, the following 5 components were added to give a final transfection volume of 50 µl: 200-400 ng of the reporter firefly luciferase construct; 25 ng of the renilla luciferase vector; 1-2 µl of HiPerfect transfection reagent (Qiagen) and synthetic miRNAs (known as miRNA miScript mimics, Qiagen) diluted to a range of 1 – 40 nmol, all diluted in Opti-MEM (Invitrogen). The mixture was then incubated at room temperature for 20 minutes before being added to the wells dropwise gently. The cells were incubated at 37^{o}C with 5% CO_2 for 48 hours.

After incubation, the cells were lysed by adding 100 µl of Glo lysis buffer (Promega) and incubated for 5 minutes. Two equal amounts of lysates (50 µl) were transferred to a white wall 96-well plate and added an equal volume of Bright Glo reagent (Promega) to one of the well and Renilla Glo reagent (Promega) to the other well. Next the firefly and renilla luciferase reading were taken using the Partha Luminescence program on a plate reader (Wallac 1420 Victor2, PerkinElmer). The two datasets were combined to allow the standardization of the firefly luciferase reading against the renilla luciferase reading for the final result.

2.3 Plasmid preparation

2.3.1 Recovery of plasmid from bacterial stab culture

Plasmid construct with full length DNMT3B gene was obtained from Addgene (Plasmid 35522: pcDNA3/Myc-DNMT3B1) (Chen et al., 2005), and received as bacterial stab culture. LB agar plate with 100 μg/ml ampicillin was used to grow the bacteria. The bacteria growing within the punctured area of the stab culture was obtained by a sterile pipette tip and run lightly over the agar plate, and then spread evenly over the entire surface of the plate using a sterile spreader. The plate was incubated overnight in a 37^{o}C incubator.

A sterile pipette was used to pick up a single colony from the plate the next morning, and inoculated to 5 mls of 2x YT media containing 100 μg/ml ampicillin. 2xYT (Yeast Extract Tryptone) medium is nutritionally rich and developed for growth of recombinant strains of *Escherichia coli*. YT medium was prepared using 16 gms tryptone, 10 gms bacto-yeast extract, 5 gms NaCl and added water up to 1 litre. This was incubated at 37^{o}C with constant shaking. In the evening, it was entire transferred to a 500ml flask contain 150-200 mls 2x YT media and returned to the incubator with constant shaking. The bacteria were harvested the next day by spinning at 3,000 *g* at 4^{o}C for 10 minutes. Plasmid was extracted using Mini, Midi or Maxi Prep (Qiagen).

2.3.2 Plasmid isolation

Plasmid DNA was purified by HiSpeed plasmid kit (Qiagen) following the manufacturer's protocol. Briefly, the bacteria pellet was resuspended in buffer P1 (Mini - 250μl, Midi – 4 mls and Maxi – 10 mls) and mixed well by vortexing. The cell mixture was added buffer P2, mixed vigorously by inverting (until the mixture

appear uniformly blue due to Lyseblue reagent) and incubated at room temperature for 5 minutes prior to the addition 4 mls of pre-chilled buffer P3 which was mixed vigorously until it was completely colourless. Lysate was then poured into the QIAfilter cartridge and incubated at room temperature for 10 minutes. Meanwhile, the QIAGEN-tip was equilibrated by applying 4 mls (Midi) or 10 mls (Maxi) with QBT Buffer. The lysate was then filtered into the equilibrated QIAGEN-tip and allowed to flow through by gravity flow. The column was washed by 20 mls (Midi) or 60 mls (Maxi) QC washing buffer. Subsequently, the plasmid DNA was eluted from the filter by 5 mls (Midi) or 15 mls (Maxi) QF buffer. To precipitate the plasmid DNA, 3.5 mls (Midi) or 10.5 mls (Maxi) of isopropanol was added and incubated for 5 minutes. The mixture was then transferred to a syringe fitted with QIAprecipitator. Using constant pressure the mixture was filtered through the QIAprecipitator. The DNA was then washed with 2 mls of 70% ethanol. Finally, 1 ml of TE buffer (TE buffer, pH 8.0 or 10 mM Tris-HCl, pH 8.5) was used to elute the DNA.

2.3.3 Ligation

The ligation reaction was prepared as followed; 2 µl 5X ligation buffer, 3:1 inserts**:** vector ratio, 1 unit of T4 DNA ligase and toped up with dH_2O to 10µl. The reaction mix was incubated at room temperature for 2 hours, and was ready for transformation.

2.3.4 Plasmid transformation

Plasmid (1-5 µl) and 25 µl competent cells (DH5α) (New England Biolabs) were mixed and incubated on ice for 20 minutes prior to heat shock at 42°C for 45 seconds. The mix was put back on ice for 2 minutes before adding 200 µl 2x YT media and plated on LB-agar plate with appropriate selection antibiotics.

2.4 Total RNA extraction

Total RNA was extracted from cells using Trizol (Invitrogen) as per manufacturer's instructions. A sub-confluent to confluent cells in a 6-well plate generated approximately 20-90 µg of total RNA. 1 ml of trizol reagent was used per well in a 6-well plate. Total RNA was precipitated by isopropanol and washed with 75% ethanol. Finally, the RNA pellet was resuspended in 20 - 30 µl of water treated with 0.1% diethylpyrocarbonate (DEPC). The samples were quantified by measuring 1 µl on a nanodrop. All RNA samples were promptly stored at -80^{o}C. 0.5 µl of the sample was removed for quality check on a 1.5% 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.

2.5 First strand cDNA synthesis

100 ng of sample RNA and 0.1-0.5 µg of random hexamers were made up to 4.5 µl with dH_2O and denatured at 65^{o}C for 10 mins. The mixture was then added 1 µl of 100 mM DTT, 2 µl of 5 x first strand buffer, 0.5 µl of RNase out (40 U/µL) (Invitrogen), 1 µl of 10 mM dNTPs, and lastly 1 µl of SuperScript™ III reverse transcriptase enzyme (200U/µL) (Invitrogen). The sample was incubated at 42^{o}C for 30 minutes and then 50^{o}C for 30 minutes. The sample was stored at -20^{o}C.

2.6 Primer design and alignment

A Primer-BLAST was conducted using the NCBI search engine, with the sequence of designed primers as a query sequence. Database of human genome and reference sequence for RNA were compared to identify the specificity and size of PCR products.

2.7 PCR reaction

The PCR reaction was set up using the gene specific primers (refer to table 2.3 for the list of primers). Briefly, 1.25 µl of the RT sample was added to 1.25 µl of mix forward and reverse primers (10 uM), 1 µl of 10 mM dNTPs, 2 µl of 10 x buffer and 1 µl of Taq polymerase (Promega) and made up to 25 µl with water. The PCR reaction was set at 94^{o}C for 2 minutes, followed by 30 cycles of 94^{o}C for 30 sec, 55^{o}C for 30 sec, 72^{o}C for 30 sec, and terminated at 72^{o}C for 5 mins (Note – extension time was determined based on a rate of extension of 1 kb/min).

All PCR products were run on TAE agarose gels to check the product sizes. Briefly, 1 - 2% (w/v) agarose, depending on fragment sizes, was dissolved in 1x TAE and left to cool before adding 0.1 µg/ml ethidium bromide (Sigma-Aldritch). Samples were mixed with 1/6 volume of 6 x loading dye containing bromophenol blue and then loaded into the wells of the gel along with a 1 Kb ladder (New England Biolabs) or a 100 bp ladder (New England Biolabs). Gels were run at 120 V for ½ to 1 hour and then visualised for bands on a UV light box.

2.8 Mutagenesis using PCR to generate mutant UTR

Mutant UTR was synthesized by 3 steps PCR. This was performed using 4 primers, i.e. 2 mutagenic primers and 2 non-mutagenic primers (see below, * represents mutant site). Initially, 2 mutagenic primers (b + c) were designed with multiple point mutations in the seed target site. First PCR was performed using primer “a” and mutagenic primer “b”, which will synthesis the first front half of the DNA with mutation at the seed target site. Then primer “d” and mutagenic primer “c” were used in the second PCR and generated the second half of the DNA. The first and second

PCR products generated the first and second halves of the UTR that have mutant seed target site but it was not a complete sequence. Finally, a third PCR was performed using primers "a" and "d" on the mixture of two PCR products to generate the complete UTR with mutant seed target site (Figure 2.1).

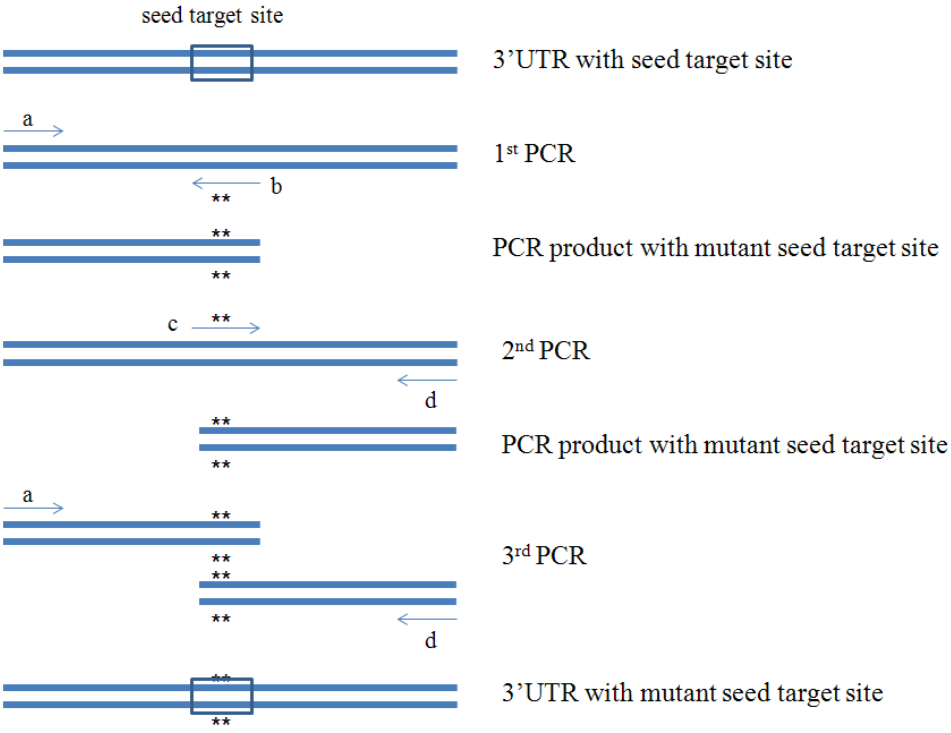


Figure 2.1 Mutagenesis by PCR

2.9 Lentivirus preparation

2.9.1 Production of lentivirus

One day prior to transfection, $1.5x10^6$ HEK293 cells were seeded in a 10 cm tissue culture plate containing 15mls of D10 (DMEM with glutamax and hepes (Invitrogen), P/S and 10% heat inactivated FBS). The next morning, replaced medium with 12 mls of fresh D10 and performed transfection in the afternoon. Performed a plasmid DNAs mix of 15 µg lentivector, 10µg pCMVΔ8.91 packaging construct and 5 µg VSV-g

envelope construct, in a ratio of 3:2:1 (Lentivector: Packaging: Envelope). Calcium phosphate transfection kit (Invitrogen) was used in the transfection.

Then, sterile water and calcium phosphate were added to a total volume of 500 µl. Followed by, 500 µl of 2x HBS added drop-wise into the plasmid/ calcium phosphate mix, while mixing them rigorously and incubated at room temperature for 30 minutes. 25µM chloroquine (Sigma) was added to the media just before the addition of plasmid/ calcium phosphate mix, gently drop-wise. The cells were incubated overnight at 37^{o}C with 5% CO_2. After 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.

2.9.2 Preparation for lentiviral infection

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

2.10 Flow cytometry analysis

Cells were detached by trypsinised and washed with PBS. Cells were analyzed in FACScan (BD Biosciences) using CELLQUEST software (BD Biosciences). Ten thousand cells were acquired for each sample. WinMDI software was used to plot the results.

2.11 Ligation of PCR product into pGEM-T easy vector

The PCR products were cloned into pGEM-T easy vector (Promega) following manufacturer protocol. In brief, 2-3 µl of the PCR sample was added to 1 µl of pGEM-T easy vector, 5 µl of 2 x rapid ligation buffer, 1 µl of T4 DNA ligase, and made up to 10 µl with water. The ligation reaction was incubated at room temperature for 2 hrs. It was then added to 50 µl of DH5α competent cells and incubated on ice for 20 minutes. The cells were subjected to heat shock for 30 sec at 42^{o}C, and then placed on ice for 2 minutes. LB plate supplemented with 100 µg/ml ampicillin was pre-warmed at 37^{o}C. 200 µl of 2x YT media was added to the cells and spread evenly onto the pre-warm LB plate and incubated at 37^{o}C overnight. Colonies were picked the next day. Plasmids were purified, digested and sequenced to validate cloning products (Figure S2.1).

2.12 Construction of pGL3 and pMIR reporter vectors

pGEM-T easy vector (Promega, A1360) was used to clone all PCR products (Table 2.1 listed all the primers used in the cloning of reporter vectors), prior to insert into a reporter vector. pGL3-control (Promega, E1741) and pMIR-REPORT™ miRNA expression reporter (Invitrogen, AM5795) vectors were used to generate the reporter constructs below. All 3' UTRs were cloned into XbaI and FseI sites at positions 1934 and 1953 respectively in pGL3-control vector (Figure S2.2). In pMIR-REPORT vector, UTRs were inserted into SpeI and SacI sites at positions 525 and 519 (Figure S2.3).

1) pGL3 – BTG1 UTR (520bp)
2) pGL3 – BTG2 UTR (698bp)
3) pGL3 – CDH1 UTR (680bp)
4) pGL3 – DNMT3B UTR (470bp)

5) pGL3 – Lefty1 UTR (401bp)
6) pGL3 – PTEN UTR (417bp)
7) pGL3 – Rock1 UTR (305bp)
8) pGL3 – SP3 UTR (795bp)
9) pMIR – NCAM2 UTR (307bp)
10) pMIR – HMGA2 UTR (433bp)
11) pMIR – ZNF148 UTR (429bp)
12) pGL3 – Mutant BTG1 UTR
13) pGL3 – Mutant CDH1 UTR
14) pGL3 – Mutant DNMT3B UTR
15) pGL3 – Mutant PTEN UTR
16) pMIR – Mutant NCAM2 UTR
17) pMIR – Mutant HMGA2 UTR

	Primers		Sequences
1	BTG1 UTR (miR-302a)	Fwd	atgctagctgccatagtttggacagtac
		Rev	atggccggccaatgtacagagagctggctg
		Size	520bp
2	BTG1 Mutant UTR	Normal	gacttttacctagcacttaaatatgtat
		Fwd	gacttttacctcgtatctgaatatgtat
		Rev	atacatattcagatacgaggtaaaagtc
		Size	329bp an 207bp
3	BTG2 UTR (isomiR-367)	Fwd	atgctagcttggaaccacatgaaagtct
		Rev	atggccggccggtggccatcctggccaaat
		Size	698bp
4	CDH1 UTR (miR-9)	Fwd	ATGCTAGCCTCACTCCTGAATTCAGTTG
		Rev	ATGGCCGGCCGATCCAAATCAAGATCCTCA
		Size	680bp
5	CDH1 Mutant UTR	Normal	TGCTGCAGCCAAAGACAGAG
		Fwd	TGCTGCAGACGTATGCAGAG
		Rev	CTCTGCATACGTCTGCAGCA
		Size	400bp and 300bp
6	DNMT3B UTR	Fwd	ATGCTAGCGCAGAGCCACCTGACTCTTG

	(isomiR-9)	Rev	ATGGCCGGCCTAATAGGTCCCGTGCAGACT
		Size	470bp
7	DNMT3B	Normal	TGGCTAAGATACCAAAACCACAGT
	Mutant UTR	Fwd	TGGCCAAGATGCAACCACCACAGT
		Rev	ACTGTGGTGGTTGCATCTTGGC
		Size	300bp and 200bp
8	LEFTY1 UTR	Fwd	ATGCTAGCGTAGCCATCGAGGGACTTGA
	(miR-302a)	Rev	ATGGCCGGCCTGGATTGGGGATGCACAA
		Size	401bp
9	PTEN UTR	Fwd	ATGCTAGCGTAGGGTACAAGTTTAATGT
	(miR-367)	Rev	ATGGCCGGCCTAACAAATGGACATCTGATT
		Size	417bp
10	PTEN	Normal	AATTTTGTGCAATATGTTCATAACGAT
	Mutant UTR	Fwd	AATTTTACGCGTAATGTTCATAACGAT
		Rev	CACAGCCATCGTTATGAACATTACGCGTAAAAT
		Size	375bp and 65bp
11	Rock1 UTR	Fwd	atgctagcGTAGAAGGTTGCACCAACAT
	(isomiR-302a)	Rev	atggccggccACATATCCATCAGTGCGGCT
		Size	305bp
12	NCAM2 UTR	Fwd	AGTCTAGAACAATATTACAGGGGCTTGA
	(isomiR-9)	Rev	ATGGGCCCATAGAGCACTTTAGCCACAT
		Size	307bp
13	NCAM2	Normal	CCTATGACCAAAACTATTCCATTG
	Mutant UTR	Fwd	CCTATGGCTTAGGCTATTCCATTG
		Rev	CAATGGAATAGCCTAAGCCATAGG
		Size	254bp and 80bp
14	HMGA2 3'UTR	Fwd	AGTCTAGATAGTCAATCACTGCACTGCA
	(miR-9)	Rev	ATGGGCCCTGGCTCTGTAGGAAGTAGAT
		Size	433bp
15	HMGA2	Normal	GTTTAGAACACCAAAGATAAGGACTA
	Mutant UTR	Fwd	GTTTAGAACTGCACTGCTAAGGACTA
		Rev	TAGTCCTTAGCAGTGCAGTTCTAAAC
		size	302bp and 168bp
16	SP3 UTR	Fwd	ATGCTAGCACAAATCAAGTTTCCAAGCA

	(miR-302a*)	Rev	ATGGCCGGCCGCTCTTACAAGACCAGCAAT
		Size	795bp
17	ZNF148 UTR	Fwd	ATGCTAGCGAAGTGAGTACCAATGTGCT
	(miR-302a*)	Rev	ATGGCCGGCCCACTAAGTTTTGCGGTCTTC
		Size	429bp

Table 2.1 Primer sequences for reporter vector/ UTR cloning

2.13 Restriction endonuclease digestion

Digestion mix was prepared in a 1.5 mls DNase free eppendorf tube and incubated at 37°C for the required period of time and the digested product was analysed by agarose gel electrophoresis. The digestion mix consisted of 1 µg DNA, 2 µl 10x restriction enzyme buffer (1 – 4 depending on which enzyme was used), 0.5 µl of each restriction enzyme (New England Biolabs) and water upto a total volume of 20 µl.

2.14 Northern hybridisation

2.14.1 Total RNA separation in denaturing gel, semi-dry blot and UV crosslinking

20 – 40 µg of total RNA was separated on a 15% polyacrylamide denaturing gel (7M Urea) in 0.5 x TBE buffer at 250 V. Six pieces of Whatman filter paper, and 1 piece of Hybond N+ nylon membrane (GE healthcare Amersham) was cut to the same size as the gel and soaked in 0.5 x TBE. Three 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 also added to the gel sandwich (Figure 2.2). A long pipette was used to squeeze out the bubbles within the gel sandwich by rolling 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 sandwich for 35 minutes (~5 V). The membrane was washed in 0.5 x TBE for 5 minutes, and then placed on top of a piece of filter paper or plastic saran wrap and UV cross-linked at 1200 μJoules, twice.

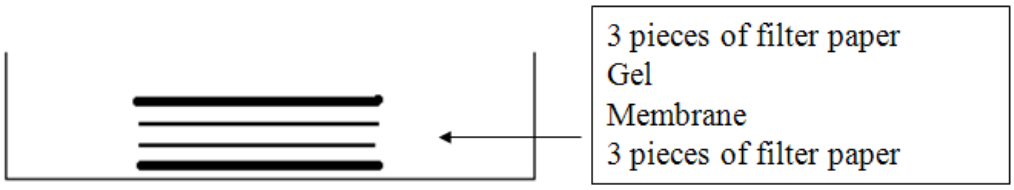


Figure 2.2
Arrangement of gel sandwich

2.14.2 Labelling of oligonucleotide probe by ^{32}P γATP

Oligonucleotide probes complementary to the targets miR-21, miR-9* and miR-302a* (Table 2.2) were labelled with ^{32}P γATP. Procedure was performed with necessary radioactive precaution. In brief, 0.5 μl of the oligonucleotide probe (50 μM) was added to 1 μl of polynucleotide kinase (PNK) buffer, 1 μl of PNK enzyme, 1 μl of ^{32}P γATP and toped up to 10 μl with water. The sample mix was incubated at 37^{o}C for 1 hour. To remove any excess ^{32}P γATP, an ethanol precipitation step was performed 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. Finally, 250 μl of -20^{o}C 100% ethanol was added to the sample and left on ice for 1 hour. The sample was then centrifuged at 1300 rpm at 4^{o}C for 20 minutes. The supernatant was removed and resuspended in 100 μl TE. The radioactivity of the precipitate and the removed supernatant were checked with a Geiger counter and the precipitate was stored at -80^{o}C. A reading of 2:1 ratio between precipitate and supernatant was considered as a successful labelling.

Name		Sequence
miR302a*	Mature	acuuaaacguggauguacuugcu
	probe	agcaagtacatccacgtttaagt
miR-9*	Mature	auaaagcuagauaaccgaaagu
	Probe	actttcggttatctagctttat
miR-21	Mature	uagcuuaucagacugauguuga
	Probe	tcaacatcagtctgataagcta

Table 2.2
Probe sequences for northern hybridisation

2.14.3 Hybridisation

The membrane was washed with 2 x SSC and 0.1% SDS with gentle agitation for 5 minutes at room temperature. 15 mls of hybridisation buffer consisting of 7.5 ml of 20 x SSC, 1.5 ml of 50 x Denhardt's solution, 0.375 ml of 20% SDS, and water was pre-warmed to 42^{o}C. The membrane was placed inside a 50 ml falcon tube with the transferred RNA side facing upwards, and pre-hybridised with 5 mls of the hybridisation buffer with constant rotation at 42^{o}C for 30 minutes. The solution was removed and fresh hybridisation buffer was added to the tube along with 50 -100 µl of the ^{32}P γATP labelled oligonucleotide probe or digoxigenin (DIG) labelled locked nucleic acid (LNA) probe (Exiqon). The probe was hybridised to the membrane overnight at 42^{o}C.

For membrane hybridised with ^{32}P γATP labelled probe, it was washed twice in 2 x SSC + 0.1% SDS at room temperature and exposed to an x-ray film in a cassette with intensifying screen at -80^{o}C overnight or up to 5 days. For membrane hybridised with

DIG-labelled LNA probe, it was washed twice with 2 x SSC and 0.1% SDS at 42^{o}C for 15 minutes, followed by washed twice with 0.1 x SSC and 0.1% SDS 42^{o}C for 5 minutes. Then, the membrane was briefly rinsed with 1 x SSC at 42^{o}C for 10 minutes. Subsequently, the membrane was incubated in blocking buffer (Roche) or 1x Maleic acid with 1% BSA for 3 hours at room temperature. The solution was replaced by fresh blocking buffer added with anti-DIG antibody (1:10,000) (Roche) and incubated at room temperature for 30 minutes. The membrane was then washed in DIG wash buffer (Roche) 4 times for 15 minutes each. The membrane was incubated in development buffer (Roche) for 5 minutes. Then, Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro)tricycle [3.3.1.1^{3,7}]decan}-4-yl)phenyl phosphate (CSPD) (1:100) substrate solution was added to the development buffer and applied to the surface of the membrane and incubated for 5 minutes. After that, the membrane was placed in a heat-sealable plastic bag and any extra buffer was squeezed out, and incubated at 37^{o}C for 15 minutes in the dark. Finally, the membrane was exposed to X-ray films in a cassette at room temperature for a suitable amount of time depending on the signal intensity.

2.15 Western blotting

2.15.1 Cell lysis and protein extraction by RIPA buffer

Cells were washed 1x in PBS at 4^{o}C and lysed in 50-200 µl of RIPA buffer directly in the wells (the amount of RIPA buffer depends on the size of tissue culture plate and the cell confluency). Cells were scraped using a cell scraper and the lysates were transferred to a 1.5 ml eppendorf. It was kept at 4^{o}C for 5 minutes and followed by centrifuge at 13,000 x *g* for 10 minutes at 4^{o}C to pellet any insoluble debris. The sample/ supernatant was transferred to a new eppendorf and stored at -20^{o}C.

2.15.2 SDS-PAGE electrophoresis

Cell lysates were run on a 12% (w/v) resolving with 4% (w/v) stacking acrylamide gel. The bottom and top atrium of the tank (Invitrogen) were filled with 1 x SDS running buffer. 1:4 Laemmli was added to each sample protein. A bromophenol dye was added to each well followed by the samples. Samples were loaded at equal amount and the gel was run at 120 volts for 2 to 3 hours. 10µl colorplus prestained protein ladder (New England Biolabs) was also loaded for subsequent determination of the size of the band.

2.15.3 Nitrocellulose wet transfer

Gels were transferred using a semi-dry blotting system (Biorad) onto a nitrocellulose membrane (0.2 µm pore size). Gels were carefully removed from the plastic cassette (1 mm) (Invitrogen) and layered into the permeable folding apparatus in the following order; a sponge layer, 2 sheets of Whatman chromatography paper, the nitrocellulose sheet (Hybond-ECL, GE healthcare Amersham), the gel, 2 further sheets of Whatman blotting paper (Whatman), and finally another sponge layer. The gel sandwich was soaked in a tray filled with transfer buffer and any air bubbles were removed by rolling a tube across the sandwich. The gel sandwich was placed on the platform of the transfer apparatus and transfer buffer was poured onto it until it was soaked. The transfer apparatud lid was then attached and protein was transferred for 1 hour and 45 minutes at 300 mA (<15 volt). The nitrocellulose membrane was stained with Ponceau S (Sigma) for general proteins as a loading control and indication of successful transfer. The membrane was scanned or photographed at this point to obtain a permanent record.

2.15.4 Antibody hybridisation

The membrane was washed with 1 x TBS-T buffer for 5 - 10 minutes, rolling at room temperature in a 50 ml falcon tube (Appleton Woods) to remove any residual ponceau staining. The wash buffer was then removed and replaced with 4 ml of 5% milk in 1 x TBS-T, blocked for 30 minutes at room temperature. Then, 5% milk in 1 x TBS-T was added 1:500 dilution of the DNMT3B antibody (rabbit polyclonal IgG, Santa Cruz, sc-20704) or NCAM2 (mouse monoclonal IgG, Santa Cruz, sc-136328) and incubated at 4^{o}C overnight rolling. The membrane was then washed 3 times with 1 x TBS-T buffer for 10 minutes, before incubated for 2 hours rolling at room temperature with 4 ml of 2% milk in 1 x TBS-T added with 1:10,000 dilution of anti-rabbit or anti-mouse IgG horse radish peroxidase (HRP) secondary antibody (Sigma). The membrane was then washed 3 times with 1 x TBS-T buffer for 10 minutes. Membranes were developed using the Immobilon chemiluminescent HRP substrate (1 ml of solution A and 1 ml of solution B) (Millipore). The membrane was laid flat on cling film and the HRP substrate was left on the membrane for 5 minutes. Finally, the membrane was taken to the dark room and light emission was detected by GRI biomax film and developed with a developing machine (Kodak). Whenever necessary, the membrane blot was stripped by rolling with western blot stripping buffer (Thermo Scientific) twice for 30 minutes at room temperature. It was then washed extensively with 1 x TBS-T and re-probed.

2.16 Argonaute immunoprecipitation

Cells from 8 wells of 6-well plate were washed with PBS and lysed with 10 ml of NP-40 lysis buffer. The cells were spun at 3,000 x g for 30 minutes at 4^{o}C, and the supernatant was added to a fresh 15 ml falcon tube. 2 mls of Argonaute 1 (Ago1) and

Argonaute 2 (Ago2) hybridoma supernatants were added, these were supplied by Gunter Meister, Max-Planck Institute, Germany. The tubes were rolled at 4^{o}C overnight and then 80 µl of protein-G beads (Santa Cruz) was added to the lysate and rolled at 4^{o}C for 2 hrs. The beads were then spun down at 3,000 x g for 5 minutes and washed with 3 x 10 mls of NP-40 wash buffer and 1 x 10 mls of PBS, before resuspended in 200 µl of TE buffer. Next the beads solution was transferred to a 1.5 ml eppendorf tube and an equal volume of phenol pH 8 was added. The tube was vortexed for 1 minute and then spun at maximum speed for 2 minutes at room temperature. The resultant aqueous phase was pipetted into a fresh 1.5 ml tube and ethanol precipitated with 1 µl of glycoblue (Ambion), 1/10 volume of 3M sodium acetate and 3 x volumes of 100% ethanol, overnight at -20^{o}C or on ice for 2 hrs. The precipitated RNA was resuspended in DEPC-treated water and stored at -80^{o}C or directly loaded onto a 15% denaturing PAGE gel for northern hybridisation.

2.17 Construction of sponge (reporter and expression vectors)

2.17.1 Generation of pMIR reporter sponge constructs with 6 multiple miRNA binding sites and 2 multiple miRNA binding sites

The design of sponges was described in detail in chapter 4. MiR-9 (CDH1) sponge (Eurogentec) was excised from pUC57 by XbaI and HindIII at position 425 and 471 respectively and ligated into pMIR report between SpeI and HindIII at position 525 and 463 in a multiple cloning site downstream to luciferase sequence. This generated pMIR-miR9 sponge reporter vector (Figure S2.4).

IsomiR-9 (DNMT3B) sponge (Eurogentec) was excised from pUC57 by SalI and HindIII at position 448 and 471 respectively and then ligated into pMIR report between XhoI and HindIII at position 545 and 463 in a multiple cloning site downstream to luciferase sequence. This generated pMIR-isomiR9 sponge reporter vector (Figure S2.4).

pMIR sponges with 2 binding sites was created by excising a segment of the sponge (133bp) containing 4 binding sites using SpeI restriction enzyme. Thus, produced pMIR-miRNA sponges with 2 multiple binding sites. Clone 3 of pMIR- miR-9 (CDH1) sponge and all clones of pMIR-isomiR9 (DNMT3B) sponge were successfully generated (Figure S2.5).

2.17.2 Generation of pcDNA3.1(+) miR-9 and isomiR-9 sponges expression vectors

MiR9 (CDH1) and isomiR9 (DNMT3B) sponges were ligated into pcDNA3.1(+) (Invitrogen) at EcoRI (Position 952)/ApaI (Position 1001) and HindIII (Position 911)/XbaI (Position 991) respectively (Figure S2.6). These expression sponges have 6 multiple binding sites and their expression are driven by CMV promoter.

2.18 Generation of DNMT3B coding region along with its full length 3'UTR

The coding region of DNMT3B was amplified from a plasmid (Plasmid 35522: pcDNA3/Myc-DNMT3B1) obtained from addgene. This plasmid contained an insert of the full length (2562bp) of the coding region of DNMT3B. The 3'UTR of DNMT3B (1560bp) was then amplified from human genomic DNA (Promega, long

PCR kit). PCRs of the coding region of DNMT3B (DC1) and 3'UTR of DNMT3B (DC2) were performed (94-2mins, 94-30sec, 60-30sec, 65-2mins (31 cycles) and 72-10mins; GoTaq® Long PCR Master Mix (M4021), Promega) (Figure S2.7). Finally, PCR was performed to generate the DNMT3B with the full length 3'UTR (DC12) (94-2mins, 94-30sec, 60-30sec, 65-4mins (30 cycles) and 72-10mins) (Figure S2.7). 1.25µl of mixed 50ng/µl DC1 and DC2 was used as template in the PCR. Primers used in the PCR were listed in Table 2.3. DNMT3B along with its full length 3'UTR was cloned into pcDNA^{TM}3.1(+) (Invitrogen) between BamHI and XbaI at position 929 and 991, respectively (Figure S2.8).

No	Name		Sequence
1	DC1F(DNMT3B coding region)	Fwd	ATGGATCCATGAAGGGAGACACCAGGCATCTCA
	DC1R(DNMT3B coding region)	Rev	GTCTGTGTAGTGCACAGGAAAGCCA
			Expected size: 2451bp
2	DC2F DNMT3B 3'UTR	Fwd	TGGCTTTCCTGTGCACTACACAGAC
	DC2R DNMT3B 3'UTR	Rev	ACTCTAGAAGGTAAACTCTAGGCATCCGTCATCT
			Expected size: 1560bp

Table 2.3 Primer sequences for DNMT3B expression vector cloning

2.19 Construction of miRNA expressing pTRIPZ lentivector

Human genomic DNA (RP11-148B6; chromosome 4) comprising miR-302 cluster, accompanied by 120 bp upstream and 150 bp downstream to the cluster was amplified by PCR (Primers listed in Table 2.4). The amplified product is 975 bp in length. The amplified fragment was ligated into pGEM-T easy vector and verified by sequencing (Figure S2.9). Subsequently, it was excised and ligated into XhoI and MluI restriction

sites, at position 3806 and 4064 respectively of pTRIPz inducible lentiviral vector (a gift from Dr Laki Buluwela, Imperial College London). This cluster consists of 5 precursor miRNAs in the following sequence, miR-302b, miR-302c, miR-302a, miR-302d and miR-367.

No	Primers		Sequence
1	miR-302cl	Fwd	TACTCGAGATCTTTGGGAACTAGTTCAG
		Rev	TCACGCGTGGATACTGGAGATCTAAAAG

Table 2.4 Primers for amplification of miR-302 cluster from human genomic DNA.

Table 2.5 lists the primers used in the detection of pluripotency and neural related gene expression and Table 2.6 lists the primers used in the sequencing of pGEM-T easy vector.

	Primers		Sequence	Product size
1	GAPDH	Fwd	tgcaccaccaactgcttagc	80bp
		Rev	ggcatggactgtggtcatgag	
2	Oct3/4	Fwd	cttgctgcagaagtgggtggaggaa	167bp
		Rev	ctgcagtgtgggtttcgggca	
3	Sox2	Fwd	cccccggcggcaatagca	448bp
		Rev	tcggcgccggggagatacat	
4	Nanog	Fwd	agcctctactcttcctaccacc	278bp
		Rev	tccaaagcagcctccaagtc	
5	Lin28A	Fwd	ggggaatcaccctacaacct	82bp
		Rev	acttccctatccaggccact	
6	Nestin	Fwd	CAGCTGGCGCACCTCAAGATG	209bp
		Rev	AGGGAAGTTGGGCTCAGGACTGG	

7	PAX6	Fwd	AACAGACACAGCCCTCACAAAC	275bp
		Rev	CGGGAACTTGAACTGGAACTGAC	

Table 2.5 Primer sequences used in the detection gene expression

No	Primers		Sequence
1	pUC/M13	Fwd	CGCCAGGGTTTTCCCAGTCACGAC
		Rev	TCACACAGGAAACAGCTATGAC

Table 2.6 pUC/M13 sequencing primers for pGEM-T easy vector

2.20 Reagents and constructs

2.20.1 Northern hybridisation 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

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

2.20.2 Western blotting reagents

Antibodies

Primary
DNMT3B rabbit polyclonal IgG (Santa Cruz, sc-20704)
NCAM2 mouse monoclonal IgG (Santa Cruz, sc-136328)

Secondary
Peroxidase conjugated Goat anti-rabbit IgG (Sigma Aldrich)
Peroxidase conjugated Goat anti-mouse IgG (Sigma Aldrich)

Polyacrylamide 12% gel
10.15 ml Deionised water, 20 ml 30% Acrylamide/Bis solution (Biorad), 18.75 ml 1 M Tris (pH 8.8), 0.5 ml 10 % SDS, 0.5 ml 10% ammonium persulphate, 30 µl TEMED

Polyacrylamide stacking gel
6.8 ml Deionised water, 1.7 ml 30% Acrylamide /Bis solution (Biorad), 1.25 ml 1 M Tris (pH 6.5), 0.1 ml 10 % SDS, 0.1 ml 10 % ammonium persulphate, 10 µl TEMED

RIPA/SDS lysis buffer
1% Nonidet P-40, 1% Triton X-100, 1% Sodium Deoxycholate, 0.1% SDS, 150 mM NaCl, 10 mM Tris pH 8.0, 2 nM NaF (RIPA/SDS was stored at 4°C and 10µl per ml of protease inhibitor mix and aprotinin was added just before used)

Protease inhibitor mix

20 mg/ml phenyl methyl sulfonyl fluoride (PMSF), 20 mg/ml 1-10 phenanthroline, 20 mg/ml Benxamine. Dissolved in ethanol and stored at -20°C.

Laemmli lysis buffer

20% Glycerol, 2% SDS, 0.1 M Tris pH 6.8, 10% ß-Mercaptoethanol, 7 M Urea

Laemmli loading dye

20% Glycerol, 2% SDS, 0.1 M Tris pH 6.8, 7 M Urea, 10% w/v bromophenol blue

10x Running buffer: 121g Tris base, 578g Glycine, 40g SDS, water to 4 litres

Transfer buffer

25 mM Tris base, 0.2 M Glycine, 20% Methanol

Ponceau S

0.2 % Ponceau Red, 5% Acetic acid

Blocking buffer

5% non fat milk/TBS plus 0.15% TWEEN 20

2.20.3 Immunoprecipitation reagents

NP-40 lysis buffer

25 mM Tris HCl pH 7.4, 150 mM KCl, 0.5% NP-40, 2 mM EDTA

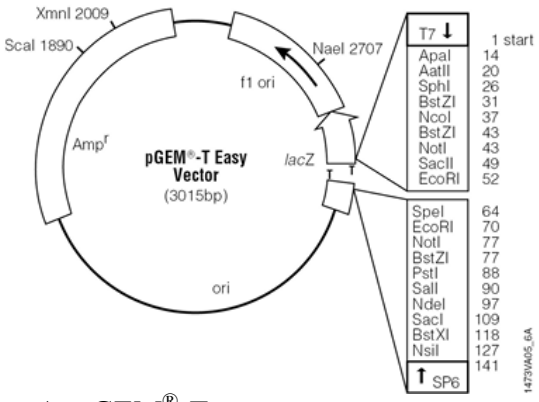
Add fresh 1mM NaF, 0.5 mM DTT, 1% proteinase inhibitors and 10 U/ml RNase out

NP-40 wash buffer

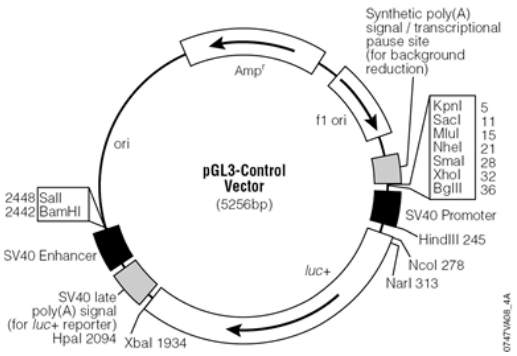
300 mM KCl, 50 mM Tris-HCl pH 7.4, 1 mM MgCl2, 0.1 % NP-40

Figure 2.3, 2.4 and 2.5 show map of vectors used in the cloning of reporter vectors and expression vectors.

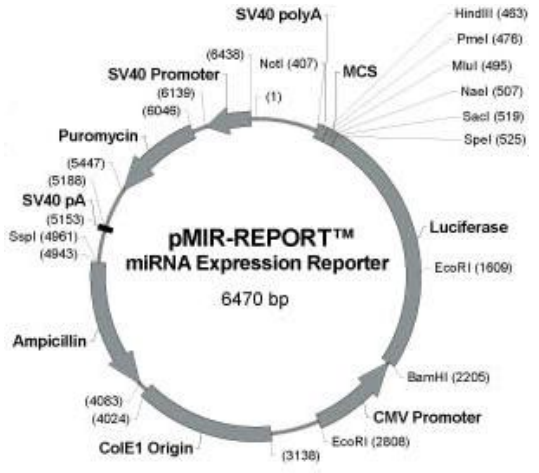
2.21 Vectors used in reporter assay cloning



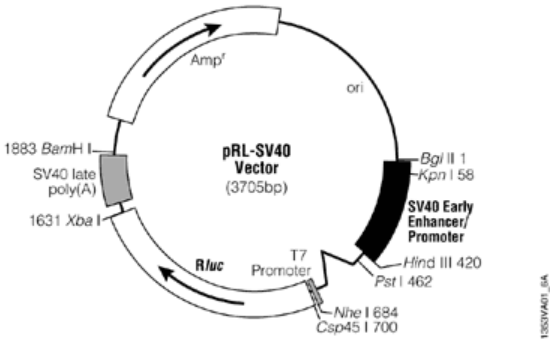
A. pGEM®-T easy vector



B. pGL3-Control Vector



C. pMIR-REPORT™ Reporter Vector



D. pRL-SV40 Vector

Figure 2.3 Vectors used in luciferase assays
pGEM-T easy vector (A), pGL3-Control vector (B) and pMIR-REPORT™ Reporter vector (C) were used in the cloning of reporter constructs. pRL-SV40 was used as normalisation control. pGEM-T easy, pGL3-Control and pRL-SV40 vectors map reproduced from Promega technical manual. pMIR-REPORT™ Reporter vector map reproduced from Invitrogen technical manual.

2.22 Vector used in cloning for sponge and DNMT3B expressions

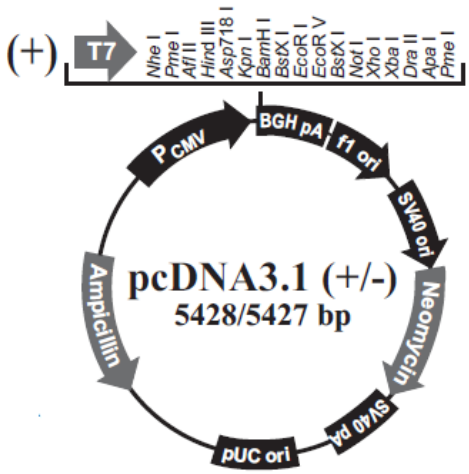


Figure 2.4
pcDNA3.1(+) vector in the cloning of sponge and DNMT3B expression vector. pcDNA3.1(+) was used in the cloning for construction of sponge and DNMT3B expression vector. Vector map reproduced from Invitrogen technical manual.

2.23 Vector used in cloning for miR-302 cluster expression

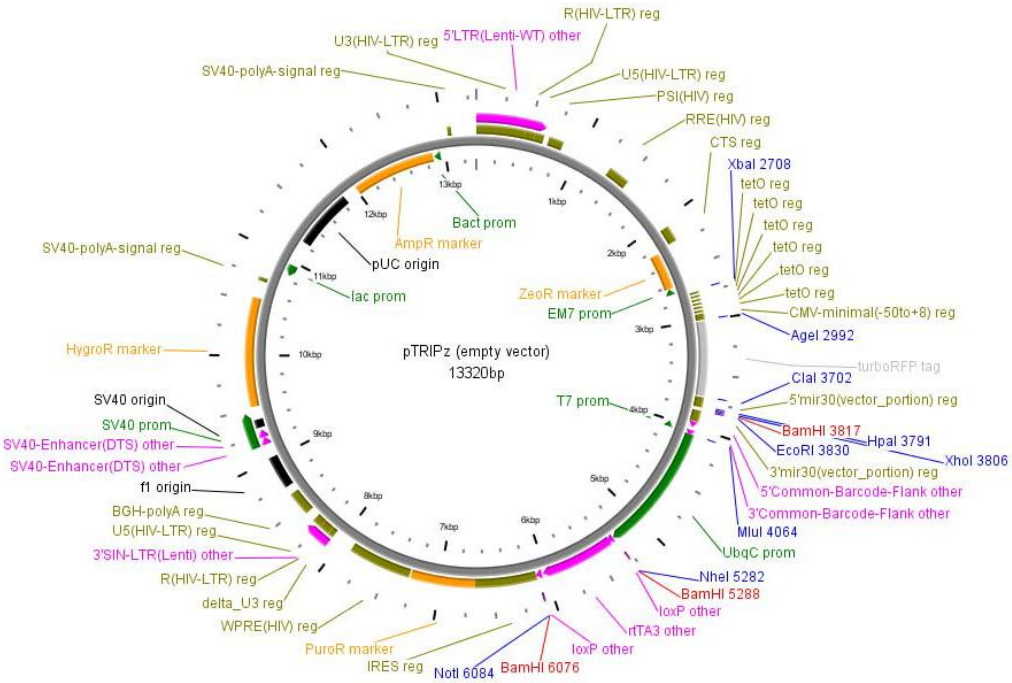


Figure 2.5
pTRIPz inducible lentiviral vector in the cloning of miR-302 cluster expression vector. pTRIPz inducible lentiviral vector was used in the cloning for construction of miR-302 cluster expression vector. Vector map reproduced from Thermo Scientific Open Biosystems technical manual.

2.24 List of cell lines used in my thesis

No	Cell lines		Characteristics	Origins
1	HEK293	Human embryonic kidney cells	These cells are easy to transfect.	Nick Dibb's lab.
2	MRC5	Human lung fibroblasts	These cells were derived from normal lung tissue of a 14-week-old male fetus and are capable of 42 to 46 population doublings before the onset of senescence.	Nick Dibb's lab.
3	H1	Human embryonic stem cells	These cells were derived from blastocyst and are capable of continuous self-renewal and differentiate into cells of any of the 3 germ layers.	Wei Cui's lab.
4	H7	Human embryonic stem cells	These cells were derived from blastocyst and are capable of continuous self-renewal and differentiate into cells of any of the 3 germ layers.	Wei Cui's lab.
5	NSC	Neural stem cells	These cells were differentiated from human embryonic stem cells.	Wei Cui's lab.
6	HeLa	Cervical cancer cells	These cells are remarkably durable and easy to transfect.	Nick Dibb's lab.
7	HepG2	Liver cancer cells	These cells were derived from human hepatoma.	Nick Dibb's lab.
8	MCF7	Breast cancer cells	These cells were derived from metastatic site (pleural effusion).	Nick Dibb's lab.
9	PC3	Prostate cancer cells	These cells do not response to androgen.	Kindly donated by Alwyn Dart from Charlotte Bevan's lab.
10	DU145	Prostate cancer cells	These cells were derived from brain metastasis and are not hormone sensitive.	Kindly donated by Alwyn Dart from Charlotte Bevan's lab.
11	LNCaP	Prostate cancer cells	These cells were derived from left supraclavicular lymph node metastasis and are androgen sensitive.	Kindly donated by Alwyn Dart from Charlotte Bevan's lab.

2.25 Bioinformatics programs

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

TargetScan - http://www.targetscan.org/

PicTar - http://pictar.mdc-berlin.de/

miRGen - http://www.diana.pcbi.upenn.edu/cgi-bin/miRGen/v3/Targets.cgi

Diana Lab TarBase - http://diana.cslab.ece.ntua.gr/tarbase/

MicroRNAdb -

http://bioinfo.au.tsinghua.edu.cn/micrornadb/browse_seq.php?ID=hsa-mir-23a

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

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

NCBI Blast - http://blast.ncbi.nlm.nih.gov/Blast.cgi

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

Venny - http://bioinfogp.cnb.csic.es/tools/venny/index.html

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

Hoodlab – Institute of System Biology - http://hood.systemsbiology.net/

OligoAnalyzer 3.1 -

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

MiRanda/ EMBL miRNA target prediction - http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/

Chapter 3 Characterisation and evaluation of IsomiRs

3.1 Introduction

During her PhD, Elcie Chan generated miRNA libraries from human ESCs, NSCs and MSCs using Solexa or 454 technologies in collaboration with David Baulcombe and Attila Molnar. Human ESCs were derived from the inner cell mass of a blastocyst (Thomson et al., 1998; see Introduction). Neuronal stem cells (NSCs) were derived from hESCs by blocking the bone morphogenetic protein signalling using noggin (Gerrard et al., 2005) and human MSCs were derived from first trimester fetal bone marrow (Guillot et al., 2007). In general, MSCs have a fibroblast-like morphology and can differentiate into cells of the mesenchymal lineage, namely bone, cartilage and fat cells. In addition to their multipotent ability, MSCs have immunosuppressive properties and the ability to support the growth of other cell types (Uccelli et al., 2008). MSCs can be isolated from bone marrow (Friedenstein et al., 1970), amniotic fluid (In't Anker et al., 2003a), placenta (Parolini et al., 2008), fetal tissues (In't Anker et al., 2003b) and umbilical cord blood (Bieback et al., 2004).

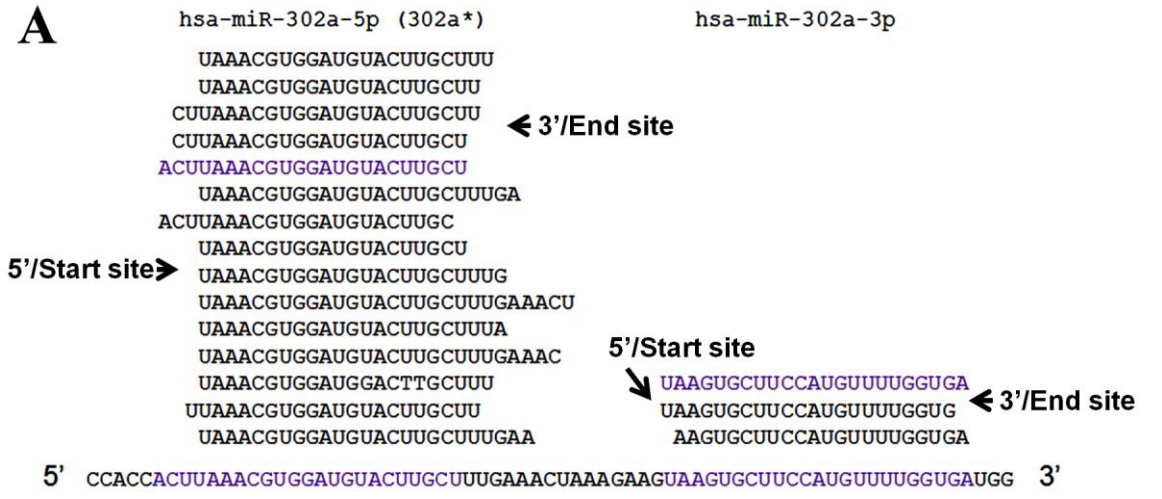
It was observed from the sequencing data that the vast majority of miRNAs in all 3 stem cell types are expressed as isomers (isomiRs) (see Introduction; Figure 3.1). Many other deep sequencing studies have discovered that mature miRNA consists of a group of isomiRs that differ in length (Morin et al., 2008; Lee et al., 2010; Cloonan et al., 2011). In principle, 5' isomiRs have different seed regions to their canonical miRNA and therefore could have a different subset of target genes. Here it is tested whether isomiRs are functional and more importantly whether 5' isomiRs can repress new mRNA subsets.

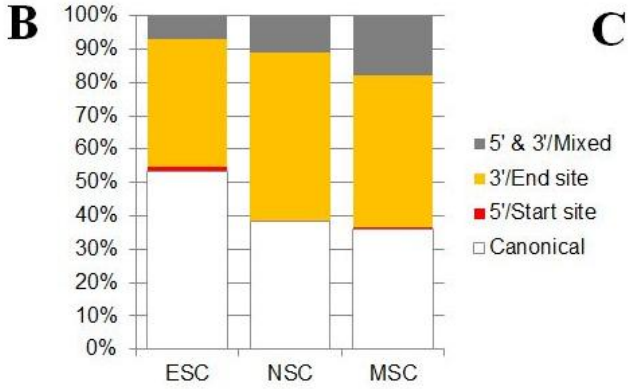
3.2 Results

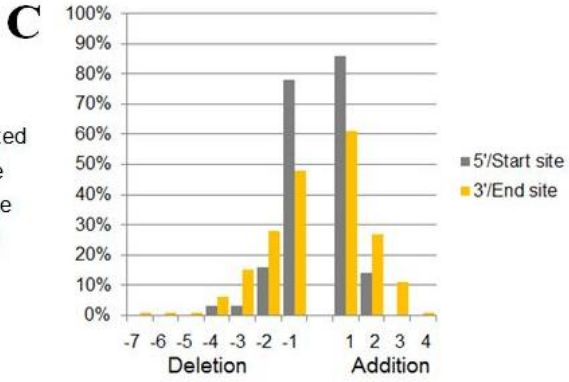
3.2.1The distribution of different categories of isomiRs in embryonic stem cells (ES), neural stem cells (NS) and mesenchymal stem cells (MS)

Figure 3.1 is an analysis of a sequencing study by Elcie Chan (see above) which indicates that miRNA isomers are widely expressed by three different stem cell types (Table S3.1). Eight percent of the miRNAs in hESCs have 5' isomiRs, with 9.6% in NSCs and 20% in hMSCs. Meanwhile, 50% of miRNAs in hESCs have 3' isomiRs, with 72% in NSCs and 71% in hMSCs. The relatively small percentage of miRNA that have 5' isomiRs suggests that processing at the 5' end of the miRNA could be tightly regulated, perhaps because the 5' end harbours the seed sequence which is an important site for target recognition.

The number of isomiRs with differences at the 5' end is small, representing about 10% of all miRNAs in hESCs and hNSCs and about 20% in hMSCs. However, the number of isomiRs with differences at the 3' end is huge, constituting about 50 to 60% of the miRNA (Figure 3.1). About 80% of the 5' isomiRs have additions or deletions of only 1 nucleotide (Figure 3.1D). In contrast the variation in the size of deletions or additions is bigger at the 3' end (Figure 3.1E).







D

5' isomirs	deletion					addition	
	-4	-3	-2	-1		+1	+2
ES	0	1	0	4	-	1	0
NS	1	0	2	4	-	1	1
MS	0	0	3	17	-	10	1
	3%	3%	16%	78%		86%	14%

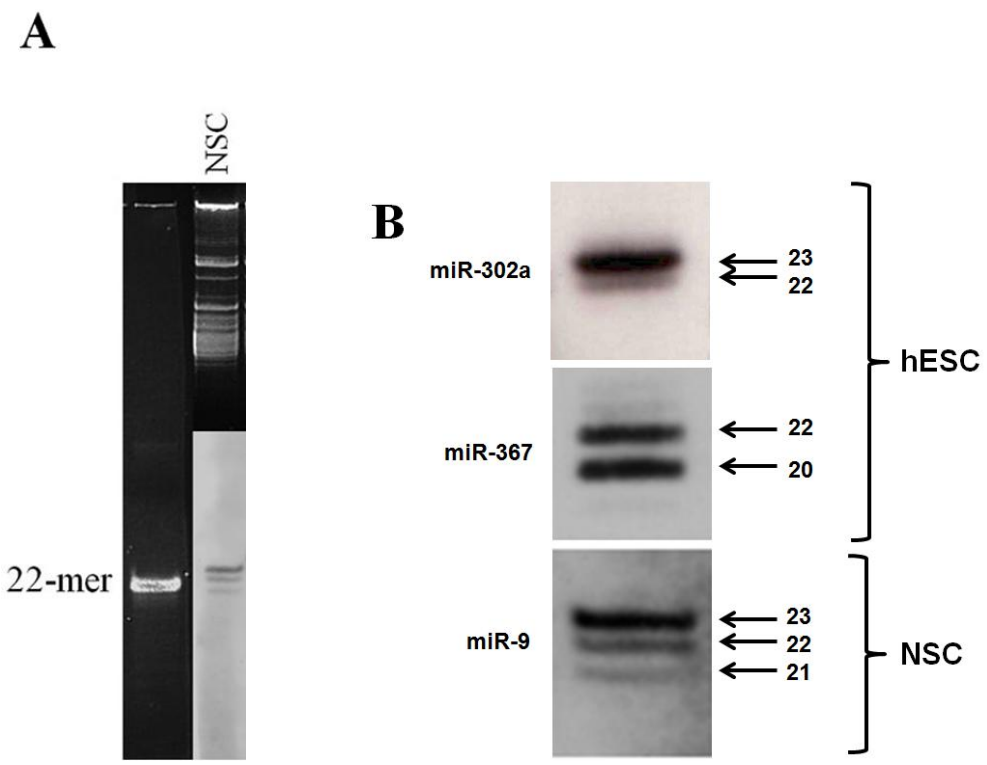
E

3' isomirs	deletion								addition			
	-7	-6	-5	-4	-3	-2	-1		+1	+2	+3	+4
ES	0	0	0	0	0	7	8	-	13	6	3	0
NS	0	0	0	2	1	9	17	-	15	10	5	1
MS	1	1	1	5	16	17	30	-	29	9	2	0
	1%	1%	1%	6%	15%	28%	48%		61%	27%	11%	1%

Figure 3.1
The distribution of 5' and 3' isomiRs in embryonic stem cells (hESCs), neural stem cells (NSCs) and mesenchymal stem cells (MSCs). A) miR-302a as an example of a miRNA expressed by hESCs with variation at 5' and 3' ends giving rise to isomiRs, denoted as 5' or start site isomiRs and 3' or end site isomiRs. Purple texts represent the canonical miRNAs. B) A bar graph illustrating the percentage of each category of isomiRs in hESC, hNSC and hMSC. C, D, E) Bar graph and tables show the number of additions and deletions of bases at 5' and 3' isomiR ends.

3.2.2 IsomiRs are not sequencing artefacts

IsomiRs are not sequencing artefacts because they were consistently detected by northern blotting (Figure 3.2). The intensity of the bands in the northern blots roughly corresponds to the sequencing numbers of miR/isomiRs (Figure 3.2). For miR-302a, the most highly sequenced miR length was 23 nts (sequencing number: 54), follow by 22 nts (sequencing number: 34), which correspond to 2 bands in the northern blot, a darker band above and a lighter band below. In miR-367, the lengths of miR/isomiR are 20 nts (sequencing number: 11) and 22 nts (sequencing number: 34), which corresponds to the two bands. However, the intensity of both bands was almost the same - based on the sequencing number, the top band should be 3 times darker than the band below. For miR-9, there is a good overall correspondence between the sequencing and northern blot results, although we did not detect a band of 24 nucleotides that was observed by sequencing (Figure 3.2).



C

miR	Sequence	Length	Seq.no
302a	TAAGTGCTTCCATGTTTTGGTGA	23	54
	TAAGTGCTTCCATGTTTTGGTG	22	18
	AAGTGCTTCCATGTTTTGGTGA	22	16

Length of miR/isomiR	Total seq.no.
23	54
22	34

miR	Sequence	Length	Seq.no
367	AATTGCACTTTAGCAATGGTGA	22	34
	AATTGCACTTTAGCAATGGT	20	11

Length of miR/isomiR	Total seq.no.
22	34
20	11

miR	Sequence	Length	Seq.no
9	TCTTTGGTTATCTAGCTGTATGA	23	587
	CTTTGGTTATCTAGCTGTATGA	22	116
	TCTTTGGTTATCTAGCTGTATG	22	49
	TCTTTGGTTATCTAGCTGTATGAA	24	37
	CTTTGGTTATCTAGCTGTATGAA	23	15
	TCTTTGGTTATCTAGCTGTAT	21	13
	TCTTTGGTTATCTAGC	16	13
	TTTGGTTATCTAGCTGTATGA	21	10

Length of miR/isomiR	Total seq. no.
24	37
23	602
22	165
21	23
16	12

Figure 3.2 IsomiRs are not sequencing artefacts
IsomiRs observed in sequencing results were also detected by northern blotting. Figure shows comparison of northern blots and sequencing results of miR-302a, miR-367 and miR-9. A) Northern blotting result of total RNA of NSC. 22-mer oligonucleotides were stained with ethidium bromide acted as ladder. B) Northern blotting of total RNA of either hESC or NSC hybridised with miR302a, miR-367 and miR-9 probes (with the predicted length of miR/isomiR) and C) sequencing results of corresponding miRNAs with the total sequencing number of each isomiRs based on their length.

3.2.3 Expression of miR/isomiRs varies in different human cell lines and mouse tissues

IsomiRs of let-7a, miR-151-5p and miR-221 were readily detected in a variety of cell and tissue types confirming that isomiRs are commonly expressed *in vivo* (Figure 3.3). Intriguingly, different cell lines and tissues express different ratios of isomiRs, as indicated by red arrows (Figure 3.3). For example, MRC5 cells and lung tissue have relatively more of the smallest isomer of miR-151-5p. In contrast, the middle band isomer of MCF7 is darkest while liver has the darkest uppermost band. This differing band intensity between cell types was also seen for miR-221 (Figure 3.3).

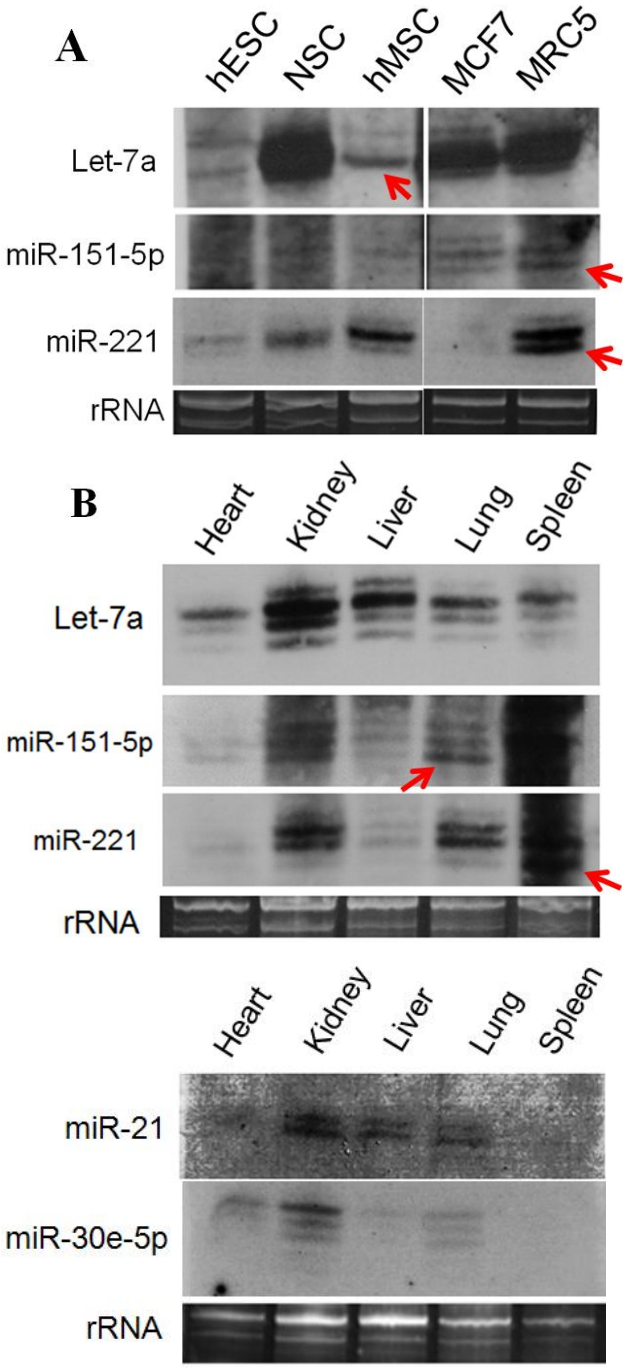


Figure 3.3
Expression of miR/isomiRs in different human cell lines and mouse tissues. Northern blots for the indicated miRNAs of total RNA prepared from A) Human cell lines: embryonic stem cells (hECS), neural stem cells (NSC), breast cancer cells (MCF7), lung fibroblasts (MRC5) and mesenchymals stem cells (hMSC) B) Mouse tissues: heart, kidney, liver, lung and spleen. Total RNA containing rRNA was stained with ethidium bromide as a loading control.

3.2.4 Detections of isomiRs by northern blotting in immunoprecipitated Ago1 and Ago2

In order to determine whether isomiRs in general are likely to be functional, we tested whether they associated with Argonaute (Ago) proteins *in vivo* by northern blot analysis of miRNAs that were first immunoprecipitated with antibodies against Ago1 or Ago2 (Figure 3.4). Ago1 and Ago2 antibodies were kindly provided by Gunter Meister from University of Regensburg, Germany. The Ago immunoprecipitation (IP) results of hESC (Figure 3.4A) indicate that miR-302a and miR-367 and their isomiRs were immunoprecipitated with Ago1 and Ago2. Interestingly, the star strand of miR-302a* and its isomiRs were also detected in Ago immunoprecipitations. MiR-9 isomiRs that were associated with Ago were detected in NSCs (Figure 3.4B). As a control we show that miRNAs were not immunoprecipitated with antibodies against a target other than Ago (Figure 3.4C) and we have also shown that the mRNAs that are precipitated under these conditions are very distinctive and not simply reflective of the total mRNA (Chan, unpublished).

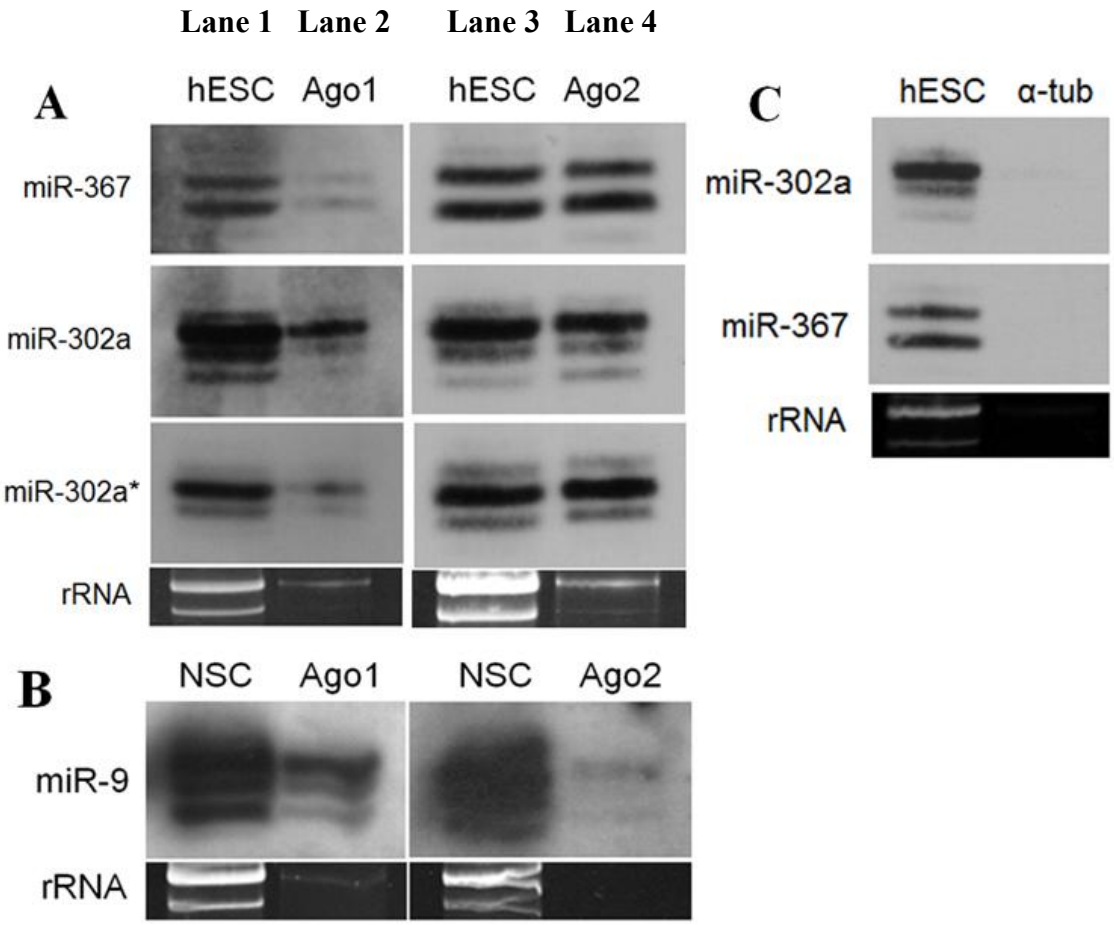


Figure 3.4
Detection of isomiRs by northern blotting following immunoprecipitation of Argonaute 1 (Ago1) or Argonaute 2 (Ago2) proteins of human embryonic stem cells (hESC) and neural stem cells (NSC). A-B) Lanes 1 and 3 are northern blots of total RNA for the indicated miRNAs prior to immunoprecipitation (IP) with Ago 1 (lane 2) or Ago2 (lane 4). C) Control showing that miRNAs were not precipitated with antibody against α-tubulin. Ribosomal RNA present in total RNA was stained with ethidium bromide as a loading control.

3.2.5 Changes of miRNA expression during hESC to NSC differentiation

We next wanted to re-establish hESC to NSC differentiation that was previously used to obtain our miRNA sequencing libraries (Chan et al., unpublished). Using a protocol developed in Dr Wei Cui's lab, hESCs were differentiated to NSCs by blocking the bone morphogenetic protein pathway (Gerrard et al., 2005) using noggin. At passage 4 (approximately 4 weeks after differentiation), the cells started to disperse into single cell morphology, as expected (Figure 3.5A, Hook et al., 2011). Cells were collected at 4 different stages of differention, i.e. hESCs (P0), a week after neural induction (P1), 4 weeks after neural induction (P4) and NSC at passage 50 (NS50) and passage 60 (NS60). Total RNAs were extracted for analysis by RT-PCR and northern blotting (Figure 3.5B,C).

The RT-PCR and northern blot analysis presented here validated our previous sequencing and microarray results (Elcie Chan, unpublished, Figure 3.5). As expected, pluripotency markers such as Oct4, Sox2, Nanog and lin28A were present in hESCs at the early stages of differentiation, while Nestin and Pax6 were seen after differentiation and continued to express into NSCs. Notably, CDH1 and DNMT3B were expressed in hESCs but were downregulated upon differentiation (Figure 3.5B). It should be noted that lane P1 of Figure 3.5B is overloaded. MiR-302a and miR-367 were confined to hESCs and disappeared after differentiation, and *vice-versa* for miR-9 (Figure 3.5C). Although isomiRs of these two miRNAs were observed in cells undergoing differentiation, there was no clear change in their ratios during this process.

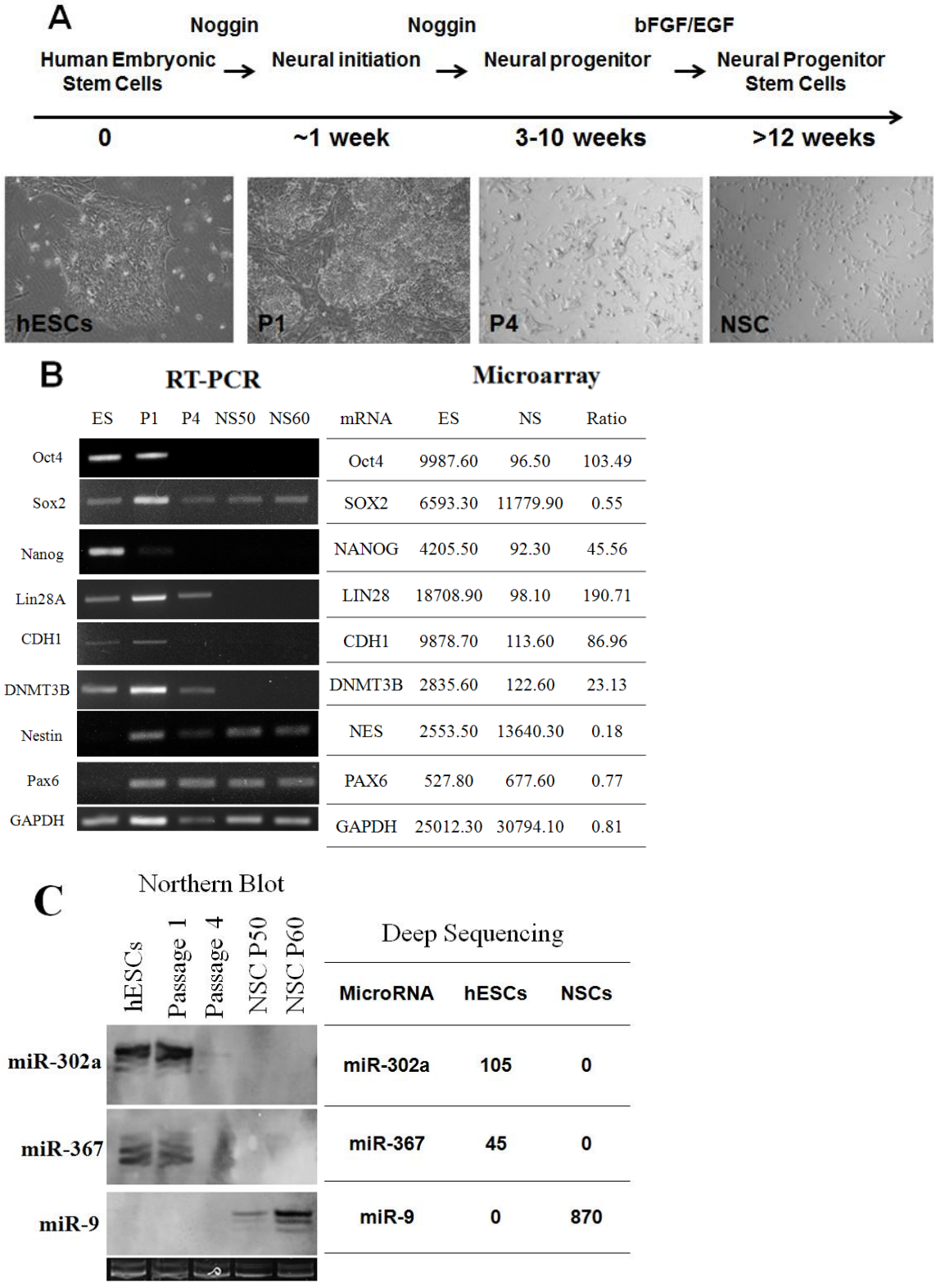


Figure 3.5 Changes of mRNA and microRNA expression during neural differentiation of human embryonic stem cells. A) Morphological changes during neural differentiation from human embryonic stem cells. B) RT-PCR and microarray analysis of mRNA expression of pluripotent and neural markers at different stages of neural differentiation. C) Northern blots for the indicated miRNAs during neural differentiation of hESCs. Total RNA containing the rRNA was stained with ethidium bromide as loading control.

3.2.6 5' isomiRs have different seed region from the canonical miRNA

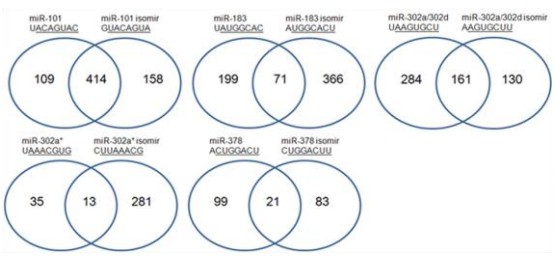
The 5' isomiRs that were identified in our miRNA sequencing databases are of particular interest as their seed sequences differ from the canonical or annotated miRNA. Table 3.1 shows some examples from miRNAs in hESCs of how additions or deletions of the 5' end of the microRNA alter its seed sequence. Table S3.2 shows this list in full.

hESC	5' difference	5' end of miRNA	Canonical seed	IsomiR seed
101	1 addition	GUACAGUACU	ACAGUAC	UACAGUA
183	1 deletion	UAUGGCACUG	AUGGCAC	UGGCACU
302a	1 deletion	UAAGUGCUUC	AAGUGCU	AGUGCUU

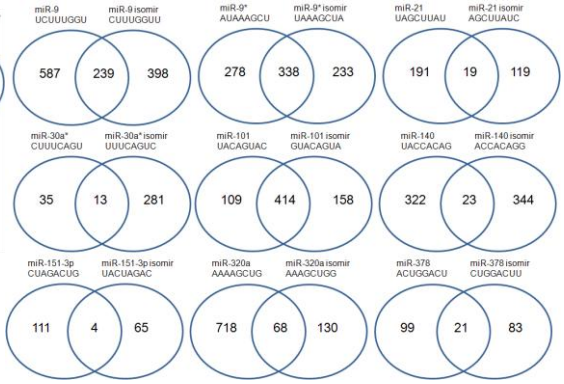
Table 3.1 Seed sequences of canonical miRNAs and isomiRs
Examples of how deletions or additions to the 5' end of miRNAs alter their seed sequence.

These differences in seed sequence should potentially alter their target selection or efficiency of target repression. I therefore investigated whether 5' isomiRs have different predicted targets to their canonical counterpart by using target prediction tools TargetScanHuman and TargetScan custom. I then cross-referenced the targets of canonical miRNA with the targets of 5' isomiR to determine which targets are in common and which are specific. Table S3.3A and B list the predicted targets of mir-9, miR-302a and their most common isomiRs that we sequenced. Bioinformatics analysis of all the miRNAs and isomiRs listed in Table S3.3 predicts that there are many specific targets of isomiRs and that the percentage of common targets is surprisingly low with an average value of about 22% (Table S3.4). This is illustrated in Figure 3.6.

A. Human embryonic stem cells



B. Neural progenitor stem cells



C. Mesenchymal stem cells

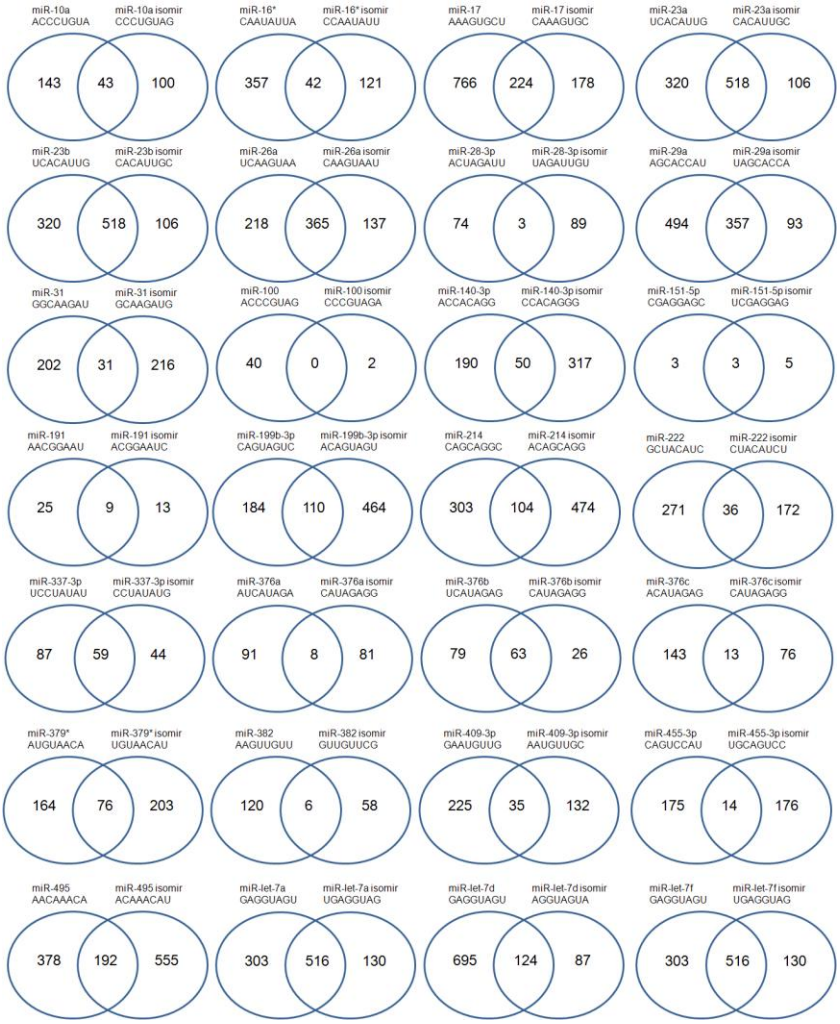


Figure 3.6 Venn diagram: TargetScanHuman and TargetScan custom prediction of canonical microRNAs and their most common isomiRs (A) Human embryonic stem cells (ES), (B) neural progenitor stem cells (NS) and (C) mesenchymal stem cells (MS). These isomiRs have a subset of predicted targets that are not predicted targets of their canonical microRNAs, as well as targets that are similar. For example, miR-101 and isomiR-101 isomer have 109 and 158 specific targets respectively and 414 common targets. Venn diagrams were generated by VENNY (Oliveros, 2007).

3.2.7 Predicting and testing targets of isomiRs

Table 3.2 shows some predicted targets of miR-9, miR-302 and miR-367 and their isomiRs that were chosen on the basis of their possible biological interest. We chose these particular miRNA genes to study because they are amongst the most abundantly expressed in hESCs and NSCs and because they express a sizeable percentage of isomiRs (see Figures 3.7 and 3.8). Table S3.3 lists the full range of predicted targets for these miRNAs. Lefty1, PTEN and BTG2 are predicted targets of both canonical miRs and isomiRs but the remaining mRNAs are specific targets (Table 3.2). The predicted target sites are well conserved between species (see Figure 1.4), which is reflective of the prediction tools that were used (see Table S3.5).

No	mRNA	MiRNA	Prediction	Luc assay	Notes
1	CDH1	miR-9	√	√	Confirmation
		isomiR-9	X	X	(Ma et al., 2010)
2	DNMT3B	miR-9	X	X	
		isomiR-9	√	√	New target
3	NCAM2	miR-9	X	X	
		isomiR-9	√	√	New target
4	HMGA2	miR-9	√	√	New target
		isomiR-9	X	√	
5	Lefty1	miR-302a	√	√	Confirmation
		isomiR-302a	√	√	(Rosa et al., 2011)
6	PTEN	miR-367	√	√	New target
		isomiR-367	√	√	
7	BTG1	miR-302a	√	√	New target
		isomiR-302a	X	√	
8	BTG2	miR-367	√	X	
		isomiR-367	√	X	
9	Rock1	miR-302a	X	nt	
		isomiR-302a	√	X	

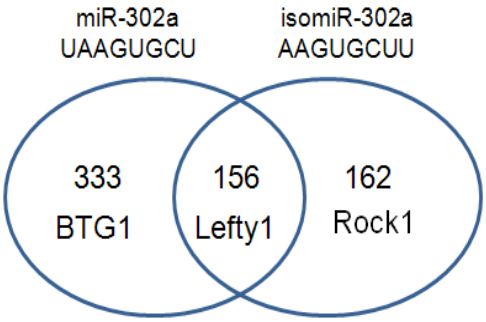
Table 3.2
Summary of luciferase assay tests of mRNAs that are predicted to be targeted by the indicated miRNAs. The shaded boxes highlight experimental results that do not agree with the predictions. √: Inhibition; X: no inhibition; nt: not tested; Luc: Luciferase.

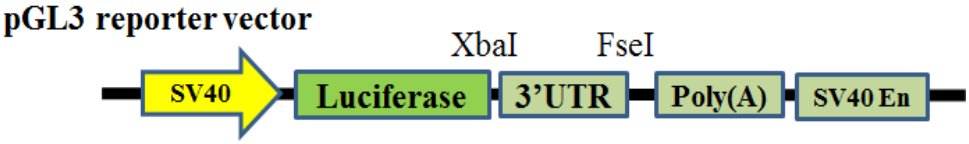
3.2.8 IsomiRs with 5' or 3' end differences are capable of targeting mRNAs *in vitro*

In order to find out whether isomiRs are functional, we constructed reporter vectors for the targets of miR-9, miR-302a, miR-367 and their isomiRs listed in Table 3.2., which also summarises the results presented in Figures 3.7-3.9. We first looked at isomiRs 302a and 367 as these have representative single nucleotide changes at the 5' and 3' ends respectively compared to their canonical miRNAs (Figure 3.7). Using targetscan and targetscan custom prediction databases, left-right determination factor (Lefty1) was predicted as a target for both miR-302a and isomiR-302a (Table 3.2), while phosphatase and tensin homolog (PTEN) is a target for both miR-367 and isomiR-367. Luciferase assays showed that both 5' and 3' isomiRs were able to target Lefty1 and PTEN 3' UTRs and therefore knockdown the expression of a luciferase reporter vector in HEK293 cells (Figure 3.7A and B). As expected, inhibition of luciferase expression was not seen if HEK293 cells were transfected with the control miRNA let-7d or with luciferase vectors with mutant seed target sites (Figure 3.7B).

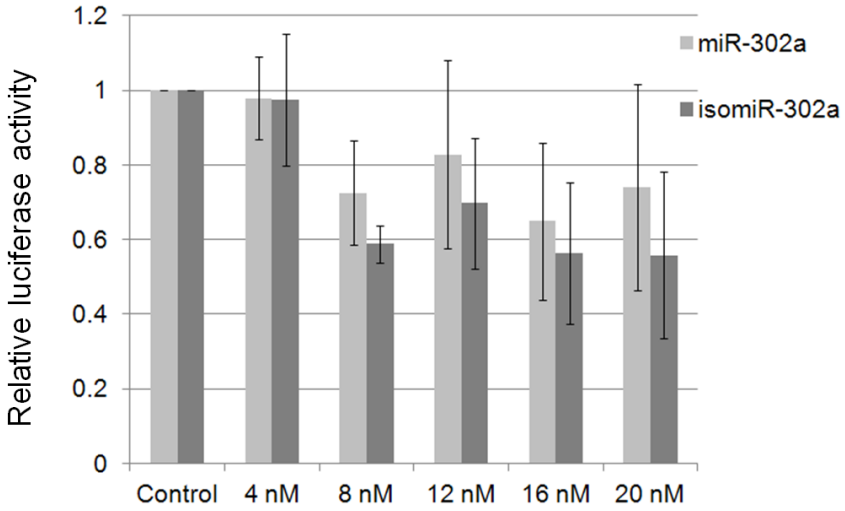
A) 5' isomiR-302a is able to repress left-right determination factor 1 (Lefty1)

No	miRNA	Sequence	Length	Seq.No.
1	miR-302a	5′ UAAGUGCUUCCAUGUUUUGGUGA 3′	23	54
2	isomiR-302a	5′ AAGUGCUUCCAUGUUUUGGUGA 3′	22	34



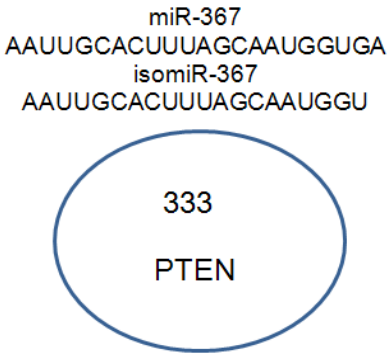


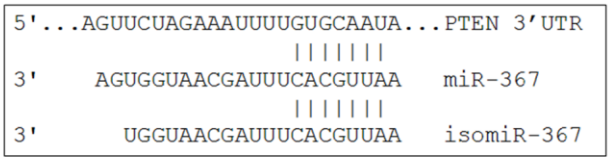

```
5'...CUGCACUAUAUUCUAAGCACUUA...Lefty1 3'UTR
                    |||||||
3'   AGUGGUUUUGUACCUUCGUGAAU   miR-302a
                   |||||||
3'   AGUGGUUUUGUACCUUCGUGAA    isomiR-302a
```

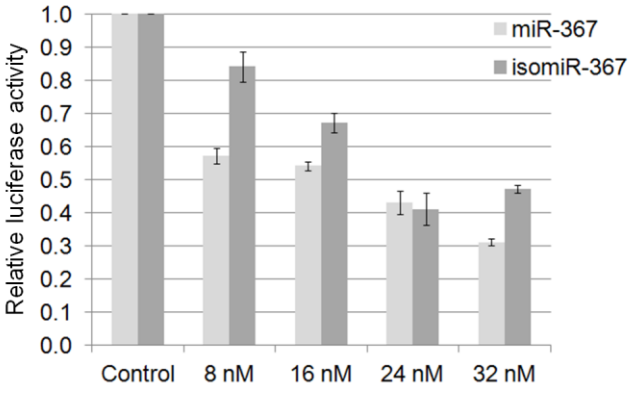


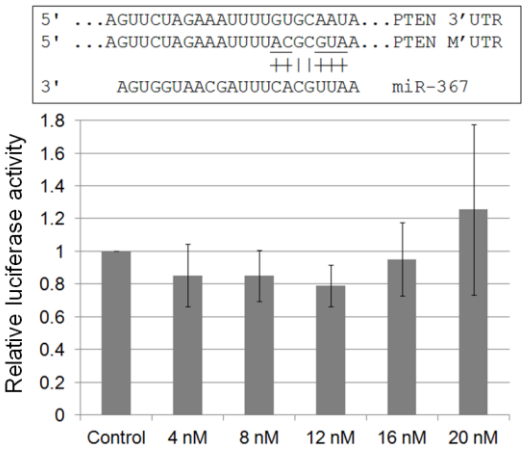
B) 3' isomiR-367 is able to repress phosphatase and tensin homolog (PTEN)

No	miRNA	Sequence	Length	Seq.No.
1	miR-367	5′ AAUUGCACUUUAGCAAUGGUGA 3′	22	34
2	isomiR-367	5′ AAUUGCACUUUAGCAAUGGU 3′	20	11









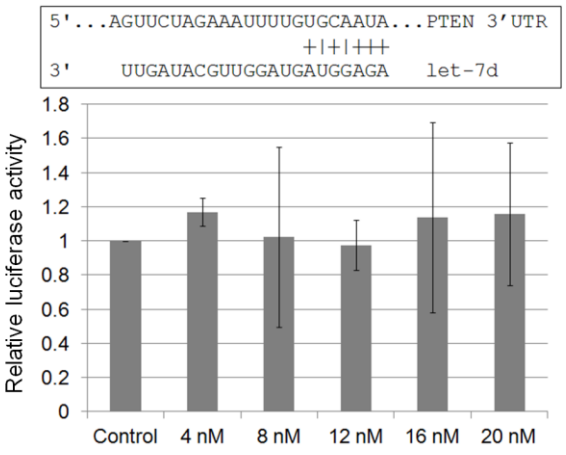


Figure 3.7

Both 5' and 3' isomiRs are functional. The isomiR-302a has a one nucleotide deletion at the 5' end, while isomiR-367 has a 2 nucleotide deletion at the 3' end. A) Both miR-302a and isomiR-302a (5' isomiR) were able to knockdown luciferase activity of Lefty1 reporter (pGL3-Lefty1), which has a 401 bp 3' UTR with a single target site. B) Both miR-367 and isomiR-367 (3' isomiR) were able to knockdown luciferase activity of a PTEN reporter (pGL3-PTEN, 417 bp 3' UTR, single target site) (see Materials and Methods). Error bars represent the standard deviation obtained from three independent experiments (n=3). Renilla luciferase was used as internal control to standardise against all firefly luciferase activities.

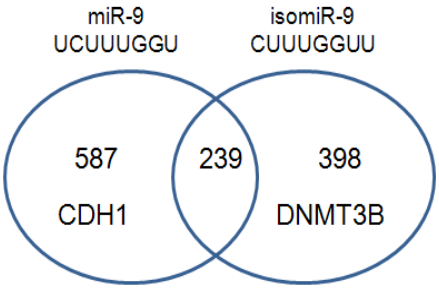
3.2.9 IsomiRs target different subsets of mRNA from their canonical/ annotated miRNAs

E-Cadherin (CDH1) and DNA methyltransferase 3 beta (DNMT3B) are predicted targets of miR-9 and isomiR-9 respectively (Table S3.3). Furthermore, these 2 genes are expressed in hESCs and downregulated upon differentiation, which corresponds with the appearance of miR-9 and isomiR-9 (Figure 3.5).

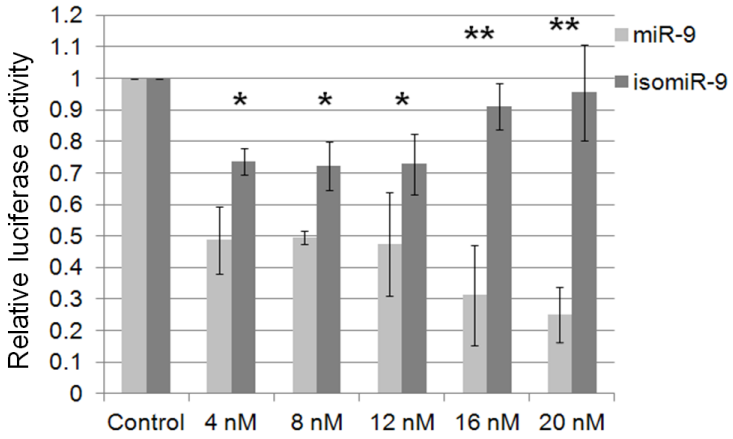
Two reporter vectors, CDH1 and DNMT3B were constructed for miR-9 and isomiR-9, respectively. Luciferase assays confirmed that the 680 bp UTR of CDH1 is a target of miR-9 but isomiR-9 was not able to knockdown CDH1 as efficiently (Figure 3.8A), whereas the 470bp UTR of DNMT3B is a target of isomiR-9 but not miR-9 (Figure 3.8B). Moreover, there was no repression, if these miRNAs were replaced with let-7d. Surprisingly, luciferase activity was increased when transfected with let-7d in CDH1 assay (Figure 3.8A). In order to confirm that the seed sequence is important, 4 out of 6 of the seed sequence were mutated to generate reporter vectors with mutated target sites within the 3' UTRs for CDH1 and DNMT3B. This markedly reduced the ability of miR-9 and isomiR-9 to repress their targets (Figure 3.8).

A) CDH1 is a target of miR-9 but not isomiR-9

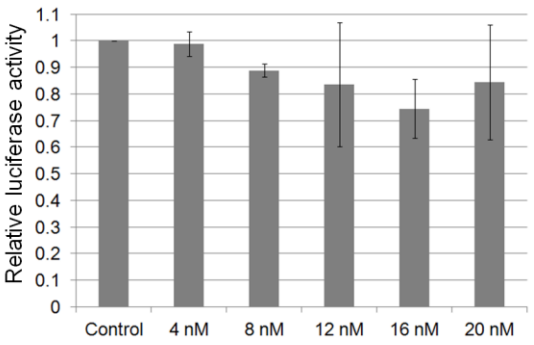
No	miRNA	Sequence	Length	Seq.No.
1	miR-9	5′ UCUUUGGUUAUCUAGCUGUAUGA 3′	23	602
2	isomiR-9	5′ CUUUGGUUAUCUAGCUGUAUGA 3′	22	165



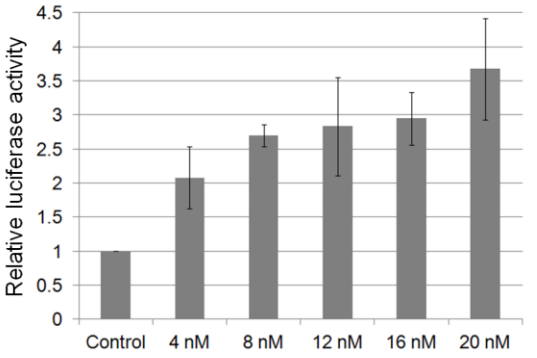

```
5'...CUAAAGUGCUGCAG--CCAAAGAC...CDH1 3'UTR
             |||       ||||||
3'   AGUAUGUCGAUCUAUUGGUUUCU    miR-9
             |||     +|||||
3'   AGUAUGUCGAUCUAUUGGUUUC     isomiR-9
```




```
5' ..CCUAAAGUGCUGCAG--CCAAAGAC...CDH1 3'UTR
5' ..CCUAAAGUGCUGCAG--ACGUAUGC...CDH1 M'UTR
             |||       +|++|+
3'   AGUAUGUCGAUCUAUUGGUUUCU    miR-9
```




```
5' ..CCUAAAGUGCUGCAG--CCAAAGAC...CDH1 3'UTR
             ||+      ++++++
3'    UUGAUACGUUGGAUGAUGGAGA    let-7d
```



B) DNMT3B is a target of isomiR-9 but not miR-9

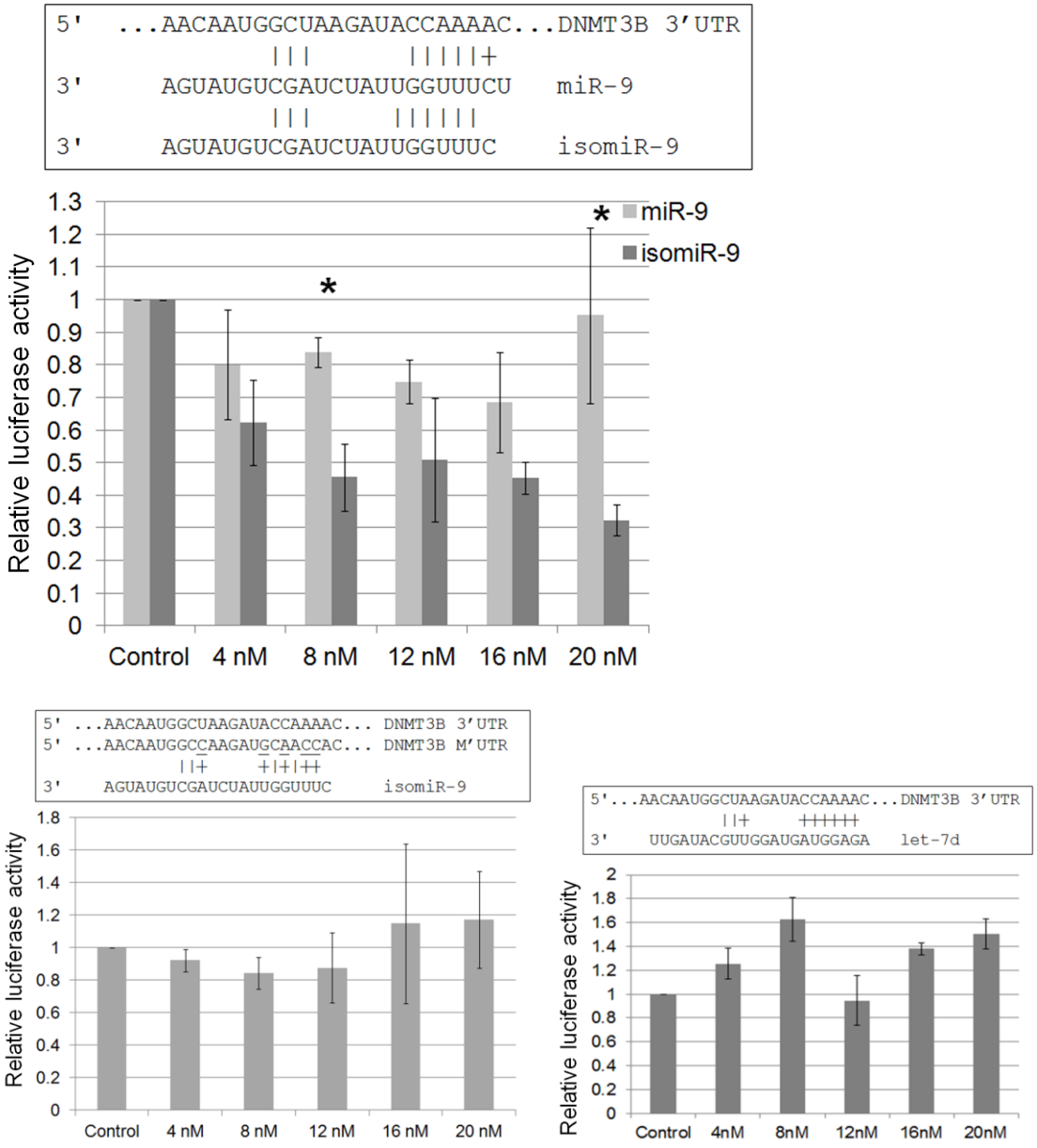


Figure 3.8
Reporter assay: CDH1 is a target of miR-9 and DNMT3B is a target isomiR-9
E-Cadherin (CDH1) and DNA methyltransferase 3 beta (DNMT3B) are predicted targets of miR-9 and isomiR-9, respectively. Relative activity of the firefly luciferase was plotted against increasing concentrations of miR-9 and isomiR-9 for A) CDH1 and B) DNMT3B reporters. miRNA repression efficiency was attenuated in mutant reporter vectors, and the control miRNA let7d as expected. Error bars represent the standard deviation obtained from three independent experiments (n=3). * and ** represent statistical significance at the levels of $p<0.05$ and $p<0.0001$ respectively (statistical difference is between miR-9 and isomiR-9). Renilla luciferase was used as internal control to standardise against all firefly luciferase activities.

3.2.10 NCAM2 is another target of isomiR-9 but not miR-9

Another predicted target of isomiR-9 but not miR-9, is the mRNA encoded by the gene for neural cell adhesion molecule 2 (NCAM2, Table S3.3, Figure 3.6). In luciferase assays, isomiR-9 was able to repress significantly the 307 bp 3' UTR of NCAM2 at 12 nM concentration only. However, miR-9 showed no repression at any concentration (Figure 3.9).

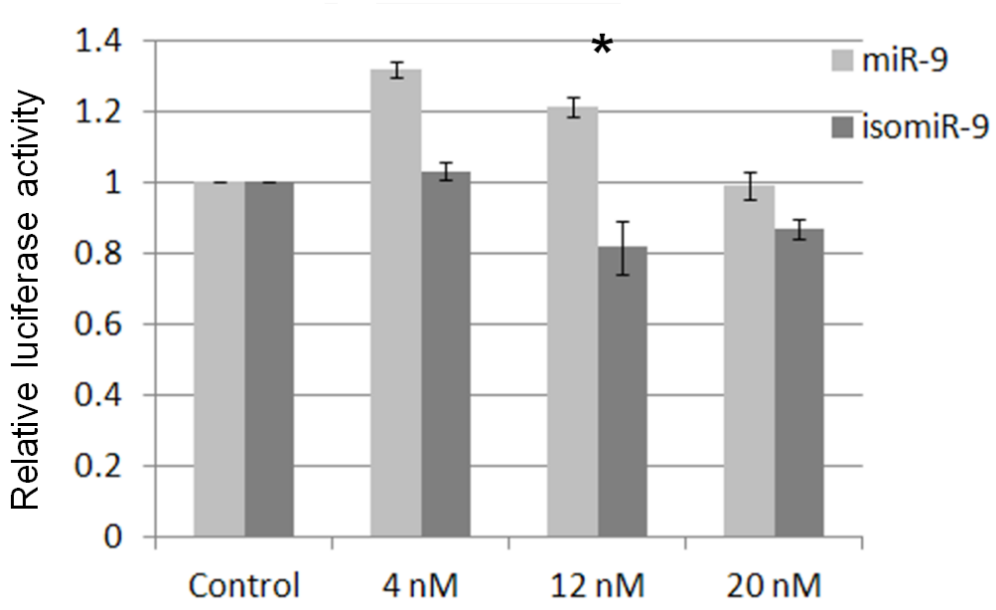


Figure 3.9
pMIR-NCAM2-3'UTR was co-transfected with either miR-9 or isomiR-9 miRNA mimic into HEK293 cells. Relative activity of the firefly luciferase for NCAM2 reporter was plotted against increasing concentrations of miR-9 and isomiR-9. Error bars represent the standard deviation obtained from three independent experiments (n=3). All results were normalised by renilla luciferase. * represent statistical significance at the level of $p<0.05$ (statistical difference is between miR-9 and isomiR-9).

3.2.11 Confirmation that miR-9 and isomiR-9 miRNA mimics are of different lengths

Total RNA was extracted from HEK293 cells transfected with miR-9 and isomiR-9 miRNA mimics and probed with miR-9 locked nucleic acid (LNA) by northern blotting. Figure 3.10 shows that isomiR-9 was smaller by one nucleotide, as expected. The mimics appeared to run more slowly than miR-9 and isomiR-9 (Figure 3.10, lane 1), for reasons that are not yet clear.

Transfected miR-9 and isomiR-9 mimics were expressed at different length

No	miRNA	Sequence	Length
1	miR-9	5′ UCUUUGGUUAUCUAGCUGUAUGA 3′	23
2	isomiR-9	5′ CUUUGGUUAUCUAGCUGUAUGA 3′	22

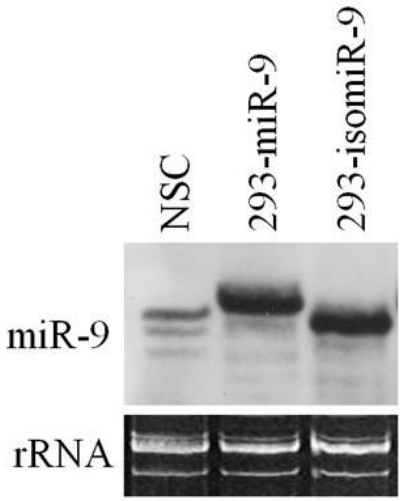


Figure 3.10
Testing the miR-9 and isomiR-9 mimics. The expected difference in length of transfected miR-9 (23 nts) and isomiR-9 (22 nts) mimics was confirmed by northern blotting. Total RNA containing the rRNA was stained with ethidium bromide as loading control.

3.2.12 Detection of miR/ isomiR-9 expression in different cell lines and tissues

Because miR-9 and isomiR-9 had generated our most interesting luciferase results, we decided to look in more detail at different cell types in order to characterise miR-9 expression and to see if miR-9 isomers might be differentially expressed. Northern blotting using a miR-9 probe was performed on a range of mouse organs as well as the indicated human cell lines (Figure 3.11A). NSC and mouse kidney were the only 2 cell/ tissue types that expressed miR-9, which largely confirms that miR-9 is a neural specific miRNA. Mouse cerebrum and cerebellum were collected subsequently and both expressed high levels of miR-9 compared to kidney tissue (Figure 3.11B). Interestingly, isomiR-9 expression was different between the mouse brain tissue and NSCs. NSCs showed darkest uppermost band and intensity reduces in the shorter isomiRs. In contrast, mouse cerebrum and cerebellum showed darker band at both the uppermost and lowermost bands than the middle band.

MiR-9/ isomiR-9 is differentially expressed in between NSC and mouse brain tissue

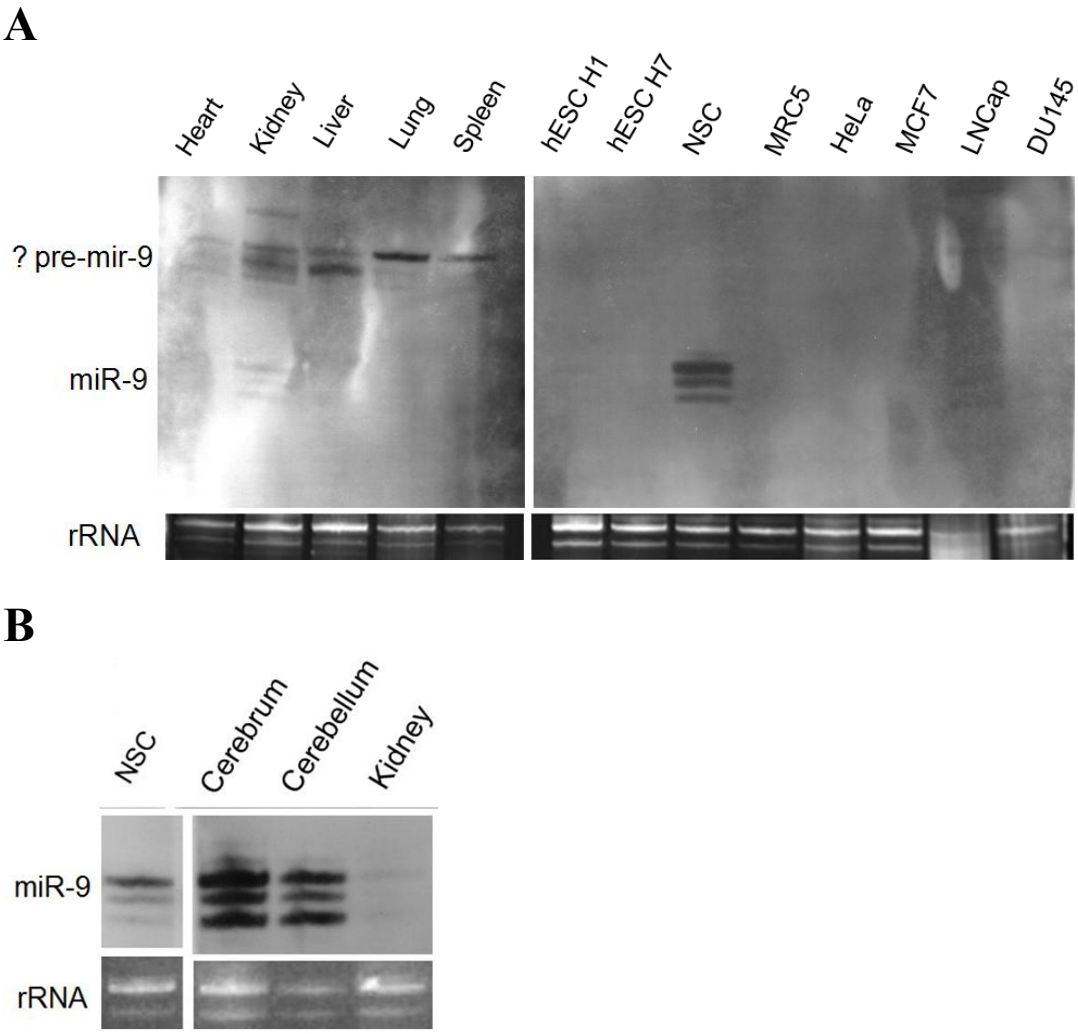


Figure 3.11
MiR-9/ isomiR-9 is differentially expressed in neural related cells and tissues. A) A miR-9 LNA probe performed on mouse tissues (heart, kidney, liver, lung and spleen) and human cell lines (H1 hESCs, H7 hESCs, NSC, MRC5, HeLa, MCF7, LNCaP and DU145). B) MiR-9 LNA probe performed on hNSC and mouse cerebrum, cerebellum and kidney. Total RNA containing the rRNA was stained with ethidium bromide as loading control.

3.2.13 False positive and false negative target predictions

Other predicted targets of miR-9, miR-302a and miR-367 that were tested included BTG1, BTG2, HMGA2 and Rock1 (Figure 3.12, Table 3.2). BTG1 is a predicted target of miR-302a but not isomiR-302a, but both were able to repress BTG1 (false negative target of isomiR-302a). Rock1 is a predicted target of isomiR-302a, however, it was not repressed (false positive target of isomiR-302a). BTG2 is a predicted target of miR-367 and isomiR-367 but neither were able to repress BTG2 (false positive target of miR-367 and isomiR-367). HMGA2 is a predicted target of miR-9 but not isomiR-9. However, both were able to repress it (false negative target of isomiR-9). Table 3.2 summarises the results of the luciferase tests of predicted targets of miR-302a, 367 and 9 and their isomiRs.

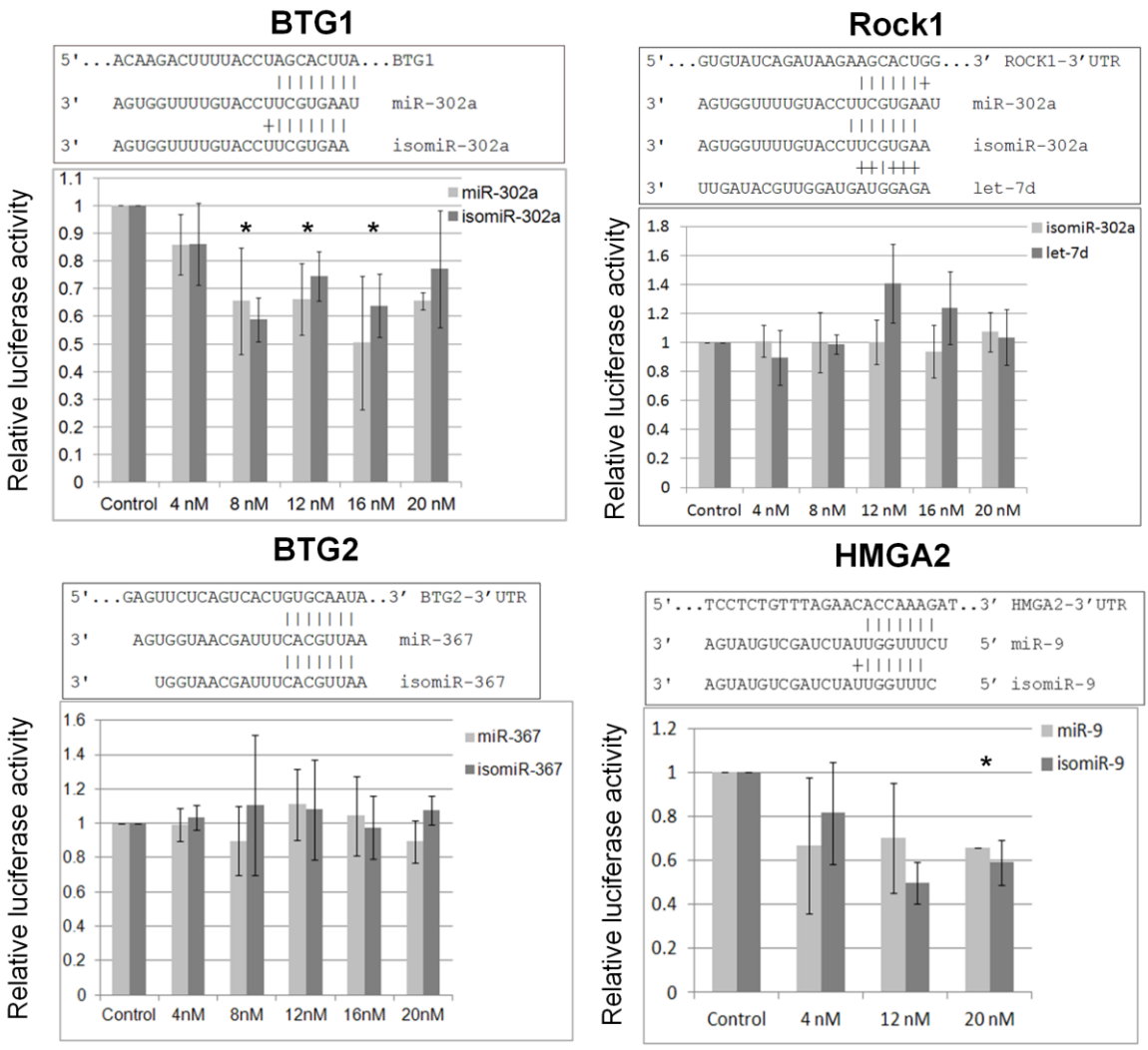


Figure 3.12

Other predicted targets that were tested. BTG1 and HMGA2 were false negative predicted targets of isomiR-302a and isomiR-9 respectively, while BTG2 was false positive predicted targets of miR-367 and isomiR-367, and Rock1 of isomiR-302a. Error bars represent the standard deviation obtained from three independent experiments (n=3). All results were normalised by renilla luciferase. * represent statistical significance at the level of $p<0.05$.

3.2.14 Validation of newly established seed target sites by seed mutation study

To validate that the repression of NCAM2, HMGA2 and BTG1 were dependent on the seed target sites in the 3' UTR, mutant reporter vectors were generated with mismatches to the miRNA seed regions. Repression was reduced or totally abolished in all mutant 3' UTRs (Figure 3.13).

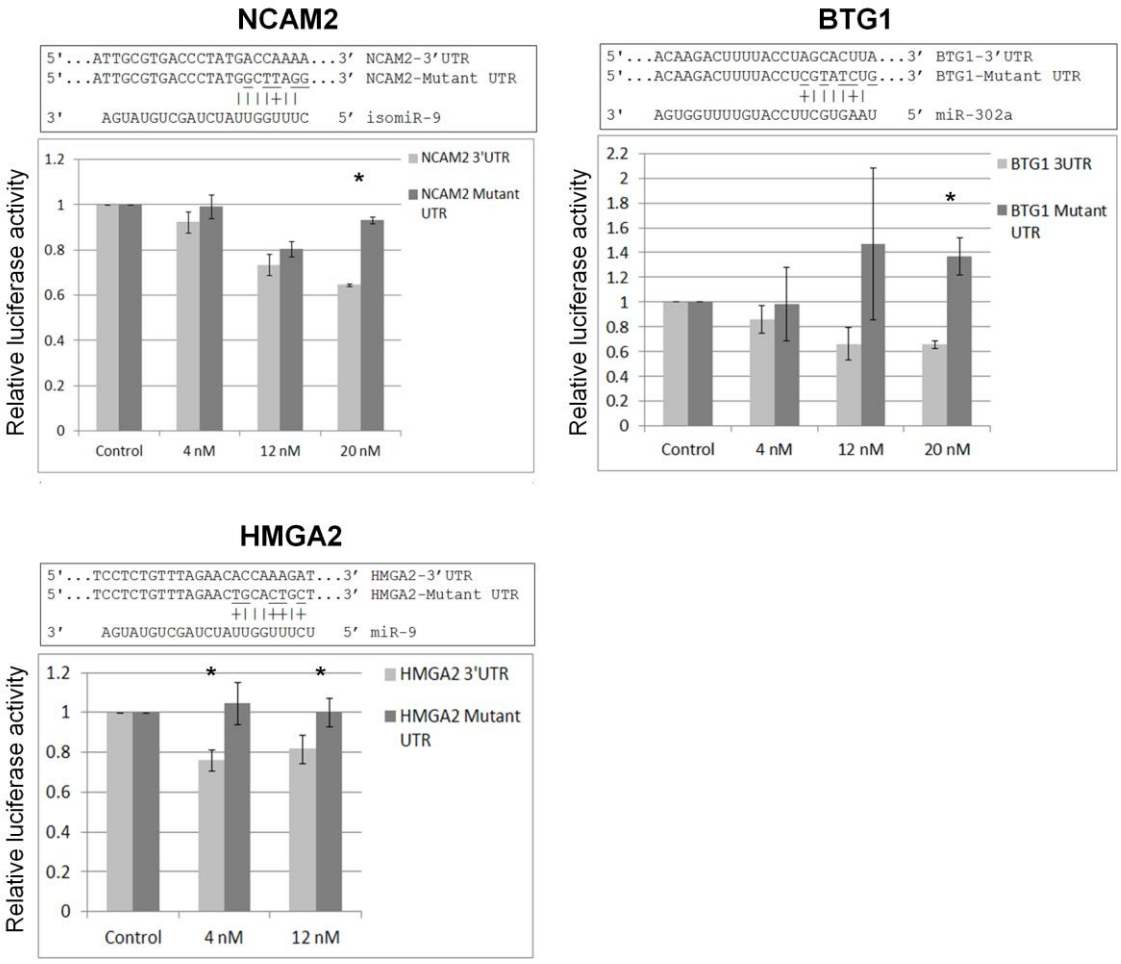


Figure 3.13
Novel miRNA target sites. Mutant NCAM2, HMGA2 and BTG1 seed target sites in the 3' UTR were made and tested by luciferase assays. In all experiments, miRNAs was transfected along with either reporter vectors with the original unmodified UTR or with a mutant UTR. Error bars represent the standard deviation obtained from three independent experiments (n=3). All results were normalised by renilla luciferase. * represent statistical significance at the level of $p<0.05$.

3.3 Discussion

Here we report that over half of the miRNAs from our three stem cell libraries are isomers (isomiRs) that have 5' or 3' differences compared to the dominant canonical sequence (Figure 3.1B). The variation we detected is unlikely to be an artefact as we observed similar variation in all cases that were tested by northern blotting (Figures 3.2 – 3.4). Previous miRNA sequencing projects have also reported the presence of isomiRs and similarly to our experiments demonstrated their association with Ago proteins (Morin et al., 2008; Lee et al., 2010; Cloonan et al., 2011; Figure 3.4).

IsomiRs with 3' deletions or additions occurred with a frequency of over 50% across the ESC, NSC and MSC miRNA sequencing libraries (Figure 3.1B). This finding is consistent with previous studies that 3' isomiR variants are more common than 5' variants in mouse, human and Drosophila samples (Burroughs et al., 2010; Lee et al., 2010; Cloonan et al., 2011; Wyman et al., 2011).

The 5' variants we sequenced occurred at a frequency of only 5 to 15% but we show that such variation would be expected to have a major impact upon mRNA targeting (Figure 3.6). We wanted to test these predictions and chose to analyse isomiRs, miR-9, miR-302a and miR-367 because they represent the most abundantly expressed miRNAs in NSCs and hESCs. IsomiR-367 was able to repress PTEN, just like its canonical miRNA, but this was not a surprise given that they only differ by 2 nucleotides at the 3' end (Figure 3.7) and is in agreement with the current opinion that the classical target selection depends on the seed region which is located at the 5' end of the miRNA.

Subsequently, bioinformatics analysis of target prediction was shifted to focus on 5' isomiRs. First, a target (Lefty1) for both miR-302a and isomiR-302a was chosen to investigate whether a 5' isomiR could function as efficiently as its canonical miRNA. Then, a target (DNMT3B) was chosen that is predicted to be targeted by a 5' isomiR but not its canonical miRNA. Indeed, reporter assays indicated that 5' isomiRs are also functional and more importantly that they can have different targets to their canonical miRNA. Seed sequence mutation studies confirmed that the predicted seed target sites were crucial for the recognition of both miRs and isomiRs (Figures 3.8 and 3.13). Two mRNAs (DNMT3B and NCAM2) were identified as targets of isomiR-9 but not the canonical miR-9 (Table 3.2, Figures 3.8 - 3.9) and we also found that isomiR-9 had lost the ability to repress CDH1. Out of 17 new tests that we made of the bioinformatics predictions, only 5 (29.4%) were incorrect (Table 3.2) and in general our results support the bioinformatic prediction that single nucleotide changes at the 5' end of a miRNA are likely to generate new targets. Intriguingly, my experiment showed an upregulation of luciferase activity when let-7d was transfected with some of the reporter vectors (see figure 3.8A) for reasons that are not yet clear. One possibility is let-7d might saturate the RISC complex and interfere with the repression mechanism.

Fukunaga et al., [25] described an *in-vivo* study where Dicer partner proteins may bind to Dicer and generate different isomiRs of a miRNA. Loquacious-PA generates a 21-mer miR-307a and loquacious-PB generates a 23-mer miR-307a. Thus by altering the Dicer partner proteins changes the choice of the cleavage site, producing isomiRs with different target specificities. They went on to show glycerol kinase and taranis were targets of 23-mer miR-307a but not 21-mer miR-307a (Fukunaga et al., 2012).

Humphrey et al., (2012) have also presented preliminary evidence to indicate that miR-133a and an isomiR have different target specificities in murine cardiomyoctyes.

We did not notice any obvious difference in the association of miRNAs with Ago1 and Ago2 (Figure 3.4), although our analysis was not extensive. Dueck et al., (2012) reported that human miRNAs are not differentially associated with Ago proteins, when analysed by northern blotting as opposed to sequencing. However miRNAs that associate with Ago2 peak at 22 nt, whilst peaks of 23 or 24 nucleotides are observed for Ago 1 and 3 (Dueck et al., 2012). This may be the reason why shorter isomiRs have a slight preference for binding to Ago2 (Dueck et al., 2012: Burroughs et al., 2011).

A number of groups have reported that isomiR expression patterns differ between cell lines or tissue types and in some cases the changes are as much as ten-fold (Fernandez-Valverde et al., 2010; Burroughs et al., 2010). These studies were based upon sequencing data, but it seems likely that they are essentially correct because our northern blotting results generally agreed with our sequencing data (Figure 3.2 -3.4). We also observed that the dominant isomiRs vary between cell and tissue types (Figure 3.3 and Table S3.1).

MiR-9 has been shown to be upregulated in breast cancer cells and to repress CDH1, which promotes cancer cells motility and invasiveness. MiR-9 mediated downregulation of CDH1 is also associated with the activation of vascular endothelial growth factor through the upregulation of beta catenin signalling, which increases tumour angiogenesis. Inhibition of miR-9 by miRNA sponge reduces metastasis

formation (Ma et al., 2010). Here, CDH1 was again validated as a target of miR-9. DNMT3B was also found to be overexpressed in a subset of hypermethylated breast cancer cells. qPCR analysis showed miRs-29c, 148a, 148b, 26a, 26b, and 203 in hypermethylator cell lines was reduced 60%–85% compared to non-hypermethylator cell lines (Sandhu et al., 2012). Further investigations are required to determine whether a downregulation of miR-9/isomiR-9 is associated with the subset of hypermethylation breast cancer and upregulation with the non-methylated breast cancer.

NCAM2 might be involved in neurological diseases such as Down's syndrome and autism. In humans, NCAM2 is located on chromosome 21. Trisomy 21 is the cause of Down's syndrome and excessive expression of NCAM2 has been suggested as a contributing factor to its development (Paoloni-Giacobino et al., 1997; Winther et al., 2012). The expression pattern of NCAM2 suggests it may have a role in the development of olfactory sensory neurones (Hamlin et al., 2004). MiRNA array revealed extensive regulation of miRNAs during the development of the brain, two of them i.e., miR-9 and miR-131 were also dysregulated in presenilin-1 null mice that exhibited severe brain developmental defects (Krichevsky et al., 2003).

It has been argued that isomiRs provide a new level of mRNA regulation (Neilsen et al., 2012) or that alternatively they are trivial variants produced by sloppy processing (Cummins et al., 2006a, b). One interesting proposal is that isomiR production might reduce the relative off target effect compared to a single miRNA (Cloonan et al., 2011). It seems unlikely that isomiRs are trivial because although an individual isomiR is by definition a minority species, our sequencing numbers for isomiR-9 and

isomir-302a were higher than for many canonical miRNAs (Table S3.6A and S3.6B). The question of whether isomiRs have important biological roles is perhaps not clear as yet. However, two of our observations indicate the possible therapeutic and/ or experimental value of isomiRs. First, we observed that isomiR-9 is an equally effective inhibitor of DNMT3B as miR-9 is of CDH1 *in vitro* (Figure 3.8). Second, some of the predicted mRNA targets of isomiRs are not predicted targets of any other miRNA (Figure S3.1 and Table S3.7).

Chapter 4 Evaluation of miR-9 and isomiR-9 targets by RNA sponges

4.1 Introduction

Most miRNAs are predicted to target hundreds to thousands of mRNAs, however, 30% of these predictions for human mRNAs are estimated to be false positives (Lewis et al., 2003). Experiments are therefore essential both to confirm targets and to explore the biological function of miRNAs and their isomiRs. Loss-of-function strategies are particularly informative and these include antisense oligonucleotide or antagomirs (Meister et al., 2004; Krützfeldt et al., 2005), miRNA sponges (Ebert et al., 2007) and genetic knockout animals (Miska et al., 2007; Park et al., 2010). Sponges are potentially more efficient than antisense oligonucleotide (Ebert et al., 2007) and also seem more likely to be able to inhibit specific isomiRs. Here we describe the use of miRNA sponges as a mean to confirm and extend our isomiR results of Chapter 3.

MicroRNA sponges, natural and synthetic were first described in 2007 (Ebert et al., 2007; Franco-Zorrilla et al., 2007). These sponges express non-coding RNAs that have multiple miRNA target binding sites, and their expression is usually driven by a CMV promoter. They act as decoy mRNAs that compete with endogenous mRNA for base pairing with miRNAs.

Sponges with binding sites containing a central bulge were reported to be more effective than sponges with perfect binding sites (Ebert et al., 2007). This may be due to degradation of the sponge transcripts by endonucleolytic cleavage activity of Ago2 upon perfect binding of the miRNA with sponge. Otaegi et al., (2011) have tested

various constructs for their ability to repress their target gene. Sponge constructs with short 6 nts separation between miRNA binding sites worked better than constructs with 29 and 42 nts separations. Constructs that started with a coding gene followed by sponge RNA worked better than constructs without the coding gene. Lastly, constructs with 6 or 12 multiple binding sites were better than constructs with 24 multiple binding sites (Otaegi et al., 2011). Ebert et al., (2007) has also described that sponges with >6 multiple binding sites repeats have only marginal increased efficiency, possibly due to saturation effects. Taking the above results as a guideline, the ingredients for a successful sponge design should be a sponge that has 6 multiple binding sites (MBS), a short separation between the MBS and that begins with a coding gene.

4.2 Results

4.2.1 Using as a different reporter vector (pMIR-Report) to validate the targets of miR-9 and isomiR-9.

We first wanted to confirm our results of Chapter 3 with a different vector. I therefore cloned the 3' UTRs of DNMT3B and CDH1 into the pMIR vector, which expresses luciferase from a CMV promoter. Figure 4.1 shows that isomiR-9 at a transfection concentration of 4 nM was able to repress luciferase activity (DNMT3B) but no inhibition was observed for miR-9. However, both miRNAs were unable to inhibit luciferase expression at 1 nM (Figure 4.1), probably because of the high luciferase expression driven by the CMV promoter.

Figure 4.2 shows that miR-9 was a better inhibitor than isomiR-9 of luciferase joined to the 3' UTR of CDH1 (described in Chapter 3) in the pMIR reporter vector, however this was only clear cut at higher miRNA concentrations. (Figure 4.2 A and B). Overall, these results confirmed that that miR-9 was a more effective inhibitor of CDH1 than isomiR-9 in luciferase reporter vector.

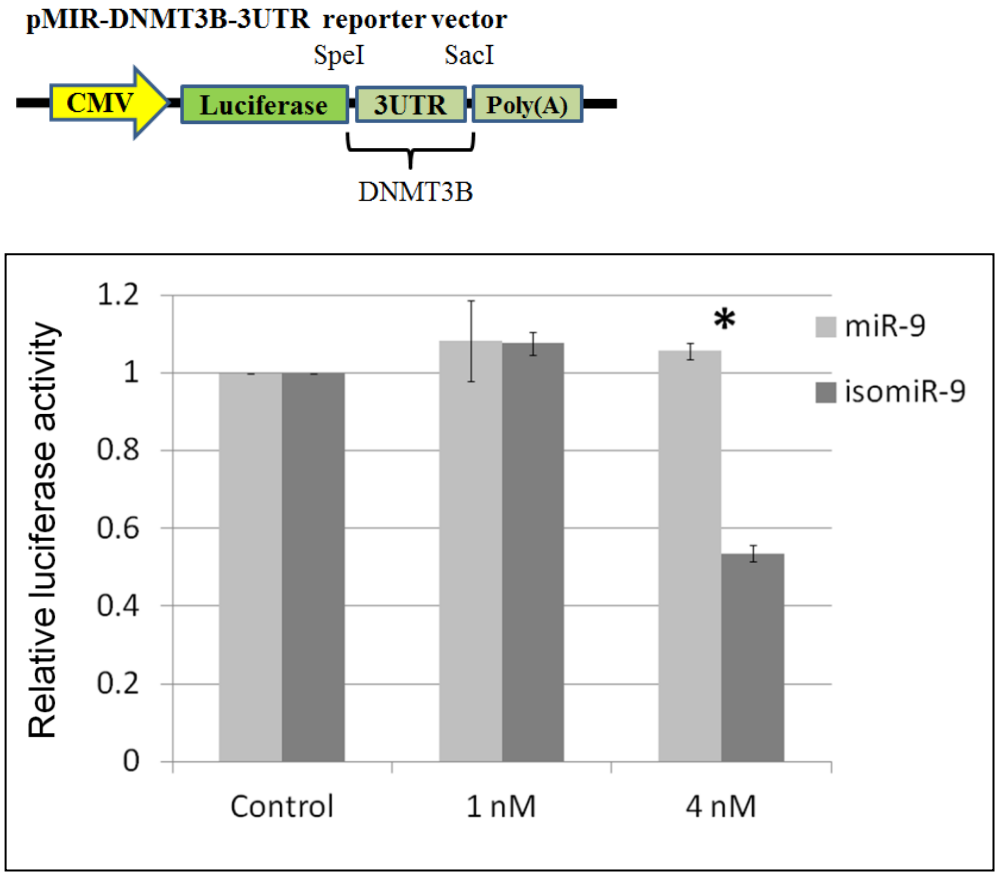
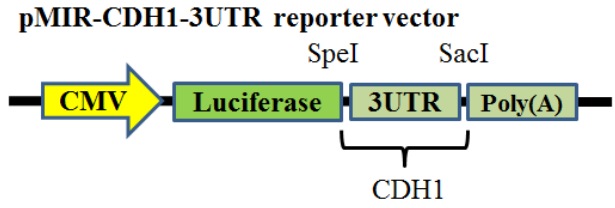
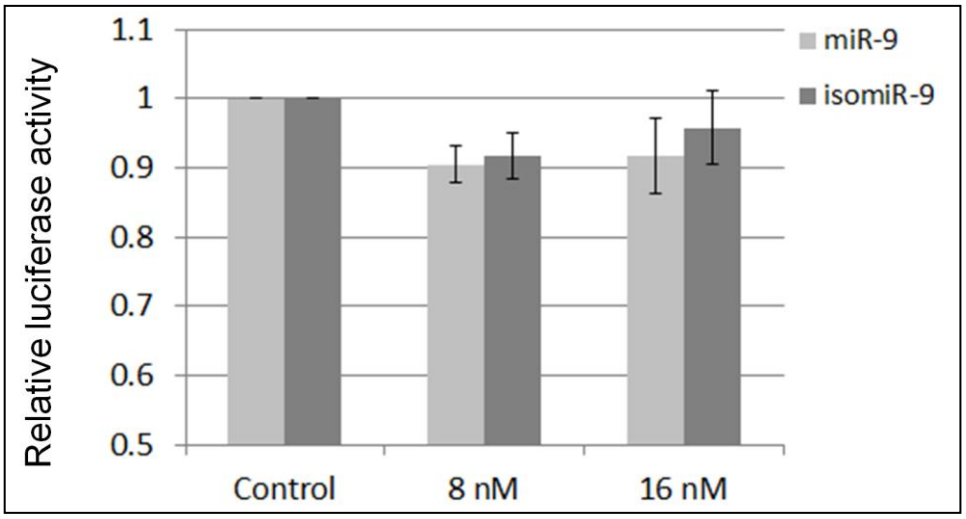


Figure 4.1
Luciferase assay of pMIR-DNMT3B-3'UTR transfected along with either miR-9 or isomiR-9. The luciferase expression in this vector is driven by a CMV promoter. A segment of DNMT3B 3' UTR was inserted between SpeI and SacI sites. pMIR-DNMT3B-3' UTR was co-transfected with either miR-9 or isomiR-9 at 1 nM and 4 nM into HEK293 cells. All results were normalised by Renilla. Error bars represent the standard deviation obtained from three independent experiments (n=3). * indicate p value <0.05 (statistical difference is between miR-9 and isomiR-9).



A



B

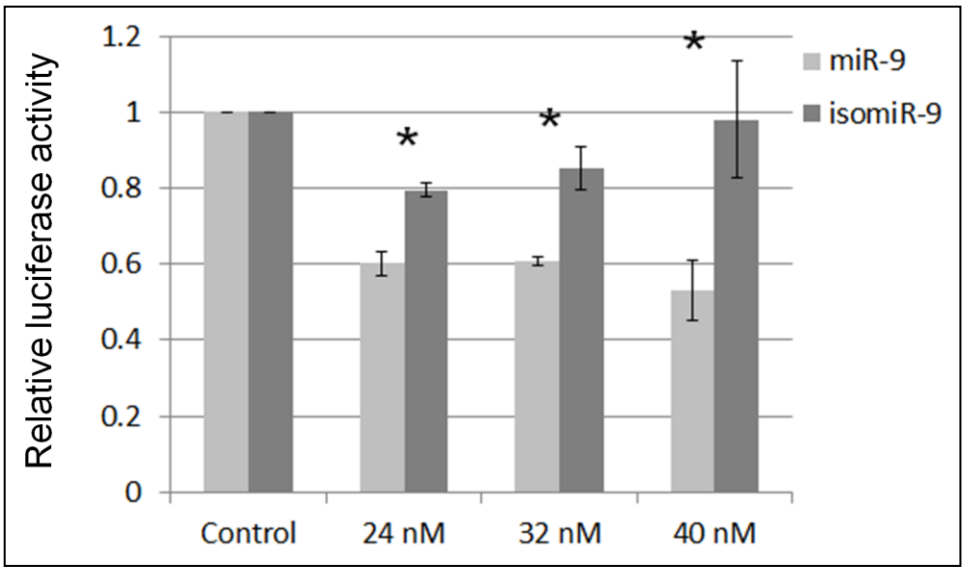


Figure 4.2
Luciferase assay of pMIR-CDH1-3'UTR, transfected along with either miR-9 or isomiR-9 at 8 and 16 nM (A). (B) 24, 32 and 40 nM.
pMIR-CDH1 3'UTR luciferase vector was constructed by cloning a segment of CDH1 3'UTR into pMIR-report vector. pMIR-CDH1-3'UTR was co-transfected with either miR-9 or isomiR-9 at 8 nM and 16 nM (A) into HEK293 cells. Then, repeated at 24 nM, 32 nM and 40 nM (B). All results were normalised by Renilla. Error bars represent the standard deviation obtained from three independent experiments (n=3). * indicate p value <0.05 (statistical difference is between miR-9 and isomiR-9).

4.2.2 Design of CDH1/ miR-9 and DNMT3B/ isomiR-9 sponges

After using a different reporter vector to validate the results of Chapter 3, the next question was whether it was possible to use sponges to specifically inhibit miR-9 and its isomiR in order to further strengthen my results. With the help of Leandro Castellano (Imperial College London), two sponges were designed and constructed with the intention to soak up either miR-9 or isomiR-9 separately. The 3' UTRs of CDH1 containing the target site of miR-9, and DNMT3B with the target site of isomiR-9 were used as the templates for the construction of these sponges. The initial sponges that were created contained 6 multiple miRNA binding sites (Figure 4.3). DNA sequences were synthesised by Eurogentec which were blunt end ligated into pUC57 at EcoRV (Position 431) within a multiple cloning site, which were subsequently validated by sequencing.

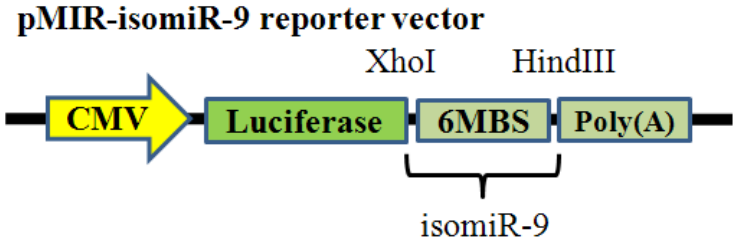
Selection of templates for the construction of sponges

Description	Sequences
miR-9	3′ AGTATGTCGATCTATTGGTTTCT 5′
CDH1/miR-9 mRNA target site in 3'UTR	5′..AGTGCCTAAAGTGCTGCAGCCAAAGA..3′
Template	AGTGCCTAAAGTGCTGCAGCCAAAGA
CDH1/ miR-9 sponge with 6 MBS	ACTAGTCGGAAGTGCCTAAAGTGCTGCAGCCAAAGACGATAGTGC CTAAAGTGCTGCAGCCAAAGAACGCGTACGGAAGTGCCTAAAGTG CTGCAGCCAAAGACGATAGTGCCTAAAGTGCTGCAGCCAAAGAAC TAGTACGGAAGTGCCTAAAGTGCTGCAGCCAAAGACGATAGTGCC TAAAGTGCTGCAGCCAAAGAAGCTT
isomiR-9	3′ AGTATGTCGATCTATTGGTTTC 5′
DNMT3B/isomiR-9 mRNA target site in 3'UTR	5′..AGAAACAATGGCTAAGATACCAAAA..3′
Template	AGAAACAATGGCTAAGATACCAAAA
DNMT3B/ isomiR-9 sponge with 6 MBS	ACTAGTCGGAAGAAACAATGGCTAAGATACCAAAACGATAGAAAC AATGGCTAAGATACCAAAAACGCGTACGGAAGAAACAATGGCTAA GATACCAAAACGATAGAAACAATGGCTAAGATACCAAAAACTAGT ACGGAAGAAACAATGGCTAAGATACCAAAACGATAGAAACAATGG CTAAGATACCAAAAAGCTT

Figure 4.3
Selection of templates and generation of miRNA sponges. The target sites of miR-9 in the 3' UTR of E-cadherin (CDH1) and target site of isomiR-9 in DNA methyltransferase 3b (DNMT3B) are shown (highlighted as green) and these were selected as templates for sponges. Sponges containing 6 templates or MBS for either miR-9 or isomiR-9 were constructed. The sequences of miR-9 and isomiR-9 are also shown. Green shading highlights the seed sequences of miR-9 and isomiR-9 and their seed targets within the sponges.

4.2.3 pMIR-isomiR-9 sponge with 6 multiple binding sites

In order to confirm that the sponge templates we designed were effective, they were first inserted as 3' UTRs downstream of the luciferase sequence in the pMIR-report vector. Figure 4.4 shows the results of experiments in which pMIR-isomiR-9 sponge with 6 MBS was co-transfected along with either miR-9, isomiR-9 or let-7d at increasing concentration from 4 to 16 nM. As expected the control miRNA let-7d was unable to inhibit luciferase activity but surprisingly both miR-9 and isomiR-9 were able to knockdown luciferase activity (Figure 4.4). The observation that both miR-9 and isomiR-9 could repress luciferase activity might be because the multiple binding sites had somehow enhanced the ability of miR-9 to recognise the binding site for isomiR-9. To test this possibility the multiple binding sites were reduced to 2. This was performed by a simple digestion with SpeI which removed 4 of the 6 multiple binding sites. Using a pMIR-isomiR-9 sponge with 2 MBS, miR-9 and isomiR-9 were still able to knockdown luciferase activity. However, isomiR-9 appeared to be more effective at lower concentrations of 4 and 12 nM (Figure 4.5 A). The experiment was repeated at a lower miRNA concentration (1 – 4 nM) and at miRNA concentrations of 1 nM, 2 nM, 4 nM and 12 nM and this confirmed that the differences in repression between miR-9 and isomiR-9 were significant at lower concentrations of miRNAs (Figure 4.5 A and B).




```
DNMT3B/isomiR-9 sponge template –
       5′..AGAAACAATGGCTAAGATACCAAAA..3′
                        ||||    ||||||
isomiR-9    3′ AGTATGTCGATCTATTGGTTTC  5′
                        ||||    ||||||+
miR-9       3′ AGTATGTCGATCTATTGGTTTCT 5′
                   ||| |||        ++|+++
let-7d      3′ TTGATACGTTGGATGATGGAGA  5′
```

pMIR-isomiR-9 sponge (6MBS)

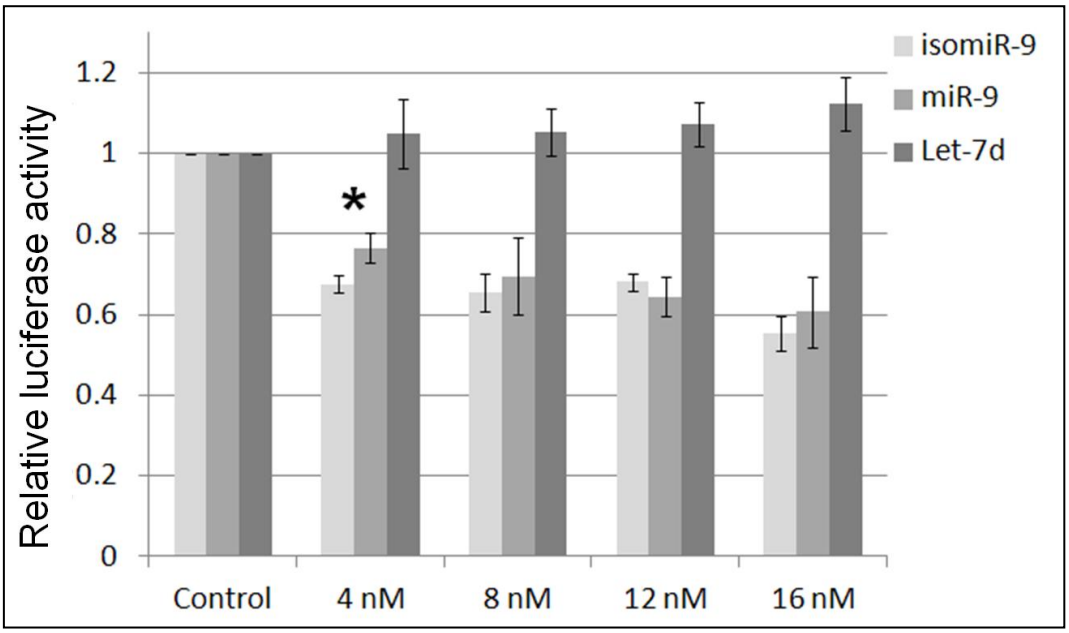
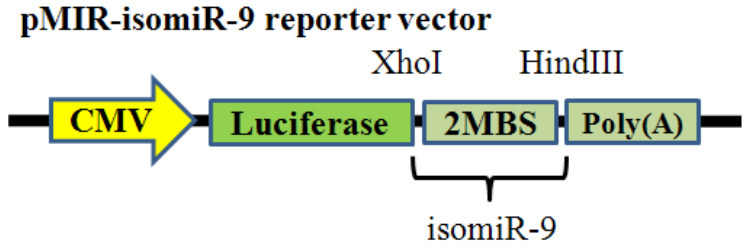
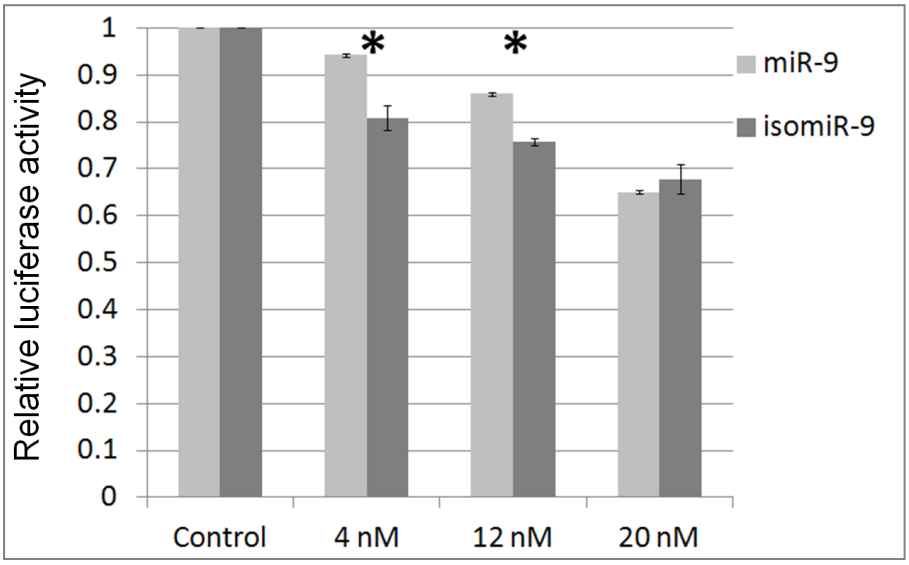


Figure 4.4
pMIR-isomiR-9 sponge with 6 MBS was co-transfected with either isomiR-9, miR-9 or let-7d. A fixed amount of 200 ng of this vector was transfected along with isomiR-9, miR-9 and let-7d at increasing miRNA concentration (4 nM, 8 nM, 12 nM and 16 nM). All results were normalised by renilla luciferase. Error bars represent the standard deviation obtained from three independent experiments (n=3). * indicate p value <0.05 (statistical difference is between miR-9 and isomiR-9).



pMIR-isomiR-9 sponge (2MBS)

A



B

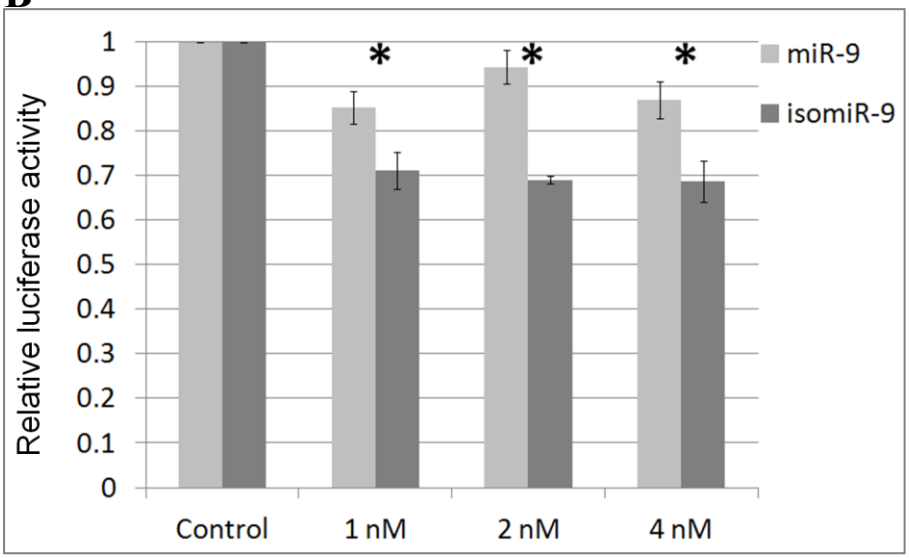
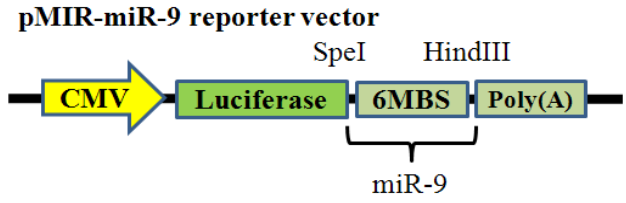


Figure 4.5
pMIR-isomiR-9 sponge with 2 MBS was co-transfected with either miR-9 or isomiR-9. A fixed amount of 200 ng of vector was transfected along with isomiR-9 and miR-9 at increasing concentrations (A) 4 nM, 12 nM and 20 nM and (B) 1 nM, 2 nM and 4 nM. All results were normalised by renilla luciferase. Error bars represent the standard deviation obtained from three independent experiments (n=3). * indicate p value <0.05 (statistical difference is between miR-9 and isomiR-9).

4.2.4 pMIR-miR-9 sponge with 6 multiple binding sites

Subsequently, pMIR-miR-9 sponge with 6 MBS was co-transfected with either miR-9, isomiR-9 or let-7d at increasing concentration from 4 nM to 20 nM (Figure 4.6). Overall miR-9 was better inhibitor of expression than isomiR-9, however, the results were not as convincing as previous results using vector with a single miR-9 binding site (Figure 3.8A).




```
CDH1/miR9 mRNA target site in 3'UTR -
      5' AGTGCCTAAAGTGCTGCAG--CCAAAGA 3'
                        |||       ||||||
miR-9       3' AGTATGTCGATCTATTGGTTTCT 5'
                        |||       |||||
isomiR-9    3' AGTATGTCGATCTATTGGTTTC  5'
                       |||        ++++++
let-7d      3'  TTGATACGTTGGATGATGGAGA 5'
```

pMIR-miR-9 sponge (6MBS)

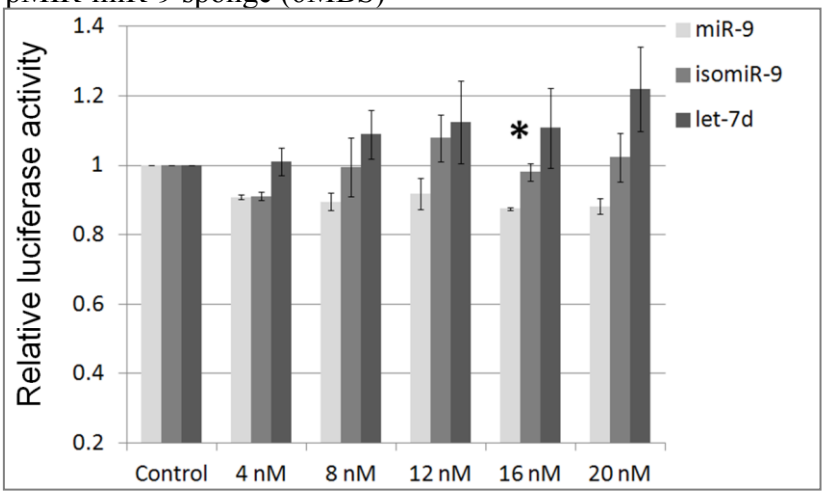


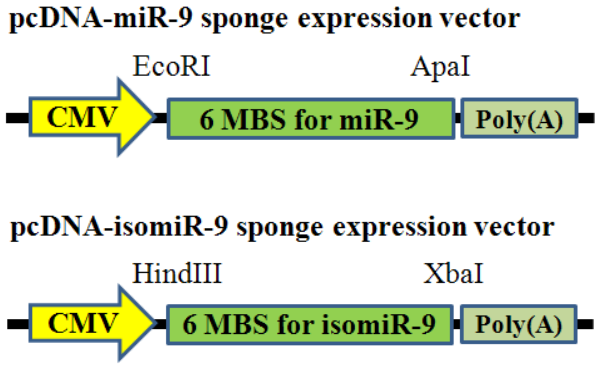
Figure 4.6
pMIR-miR-9 sponge with 6 MBS was co-transfected with either miR-9, isomiR-9 or let-7d. Increasing concentration of miRNA (4, 8, 12, 16 and 20 nM) was transfected along with a fixed concentration of the pMIR-miR-9 sponge reporter vector (200ng). All results were normalised by renilla luciferase. Error bars represent the standard deviation obtained from three independent experiments (n=3). * indicate p value <0.05 (statistical difference is between miR-9 and isomiR-9).

4.2.5 pcDNA3.1(+) -miR-9 and –isomiR-9 sponges selectively absorb miR-9 and isomiR-9 respectively

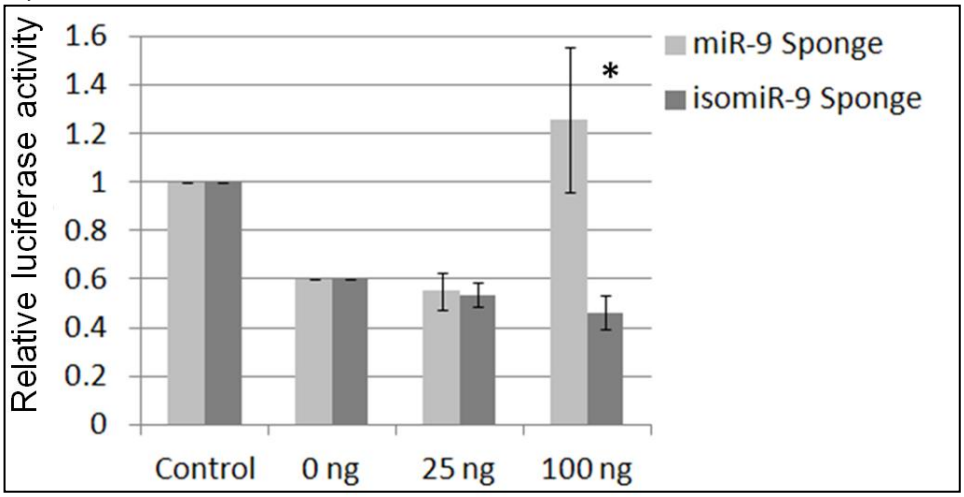
The multiple binding sites on the sponges described above may have compromised the ability of luciferase vectors to distinguish between miR-9 and isomiR-9 effects. Next, these RNA sponges were used for their intended purpose and were first introduced into expression vectors and then co-transfected along with a reporter vector and miRNA (Figure 4.7).

MiR-9 sponge and isomiR-9 sponge regions (Figure 4.3) were excised from pUC57 and ligated into pcDNA3.1(+) at EcoRI (Position 952)/ApaI (Position 1001) and HindIII (Position 911)/XbaI (Position 991) respectively. These pcDNA-miR-9 and pcDNA-isomiR-9 sponges expression vectors produce RNA sponges that have 6 multiple binding sites and their expression are driven by a CMV promoter (Figure 4.7A). Fixed amounts of pGL3-DNMT3B-3'UTR (400ng) and isomiR-9 (12 nM) were transfected together with either pcDNA-miR-9 sponge or pcDNA-isomiR-9 sponge at different concentrations into HEK293 cells (Figure 4.7B). The experiment was repeated using pGL3-CDH1-3'UTR and miR-9 (Figure 4.7C). The control columns report pGL3 transfections only. The 0 ng columns show the results for the inhibition of the pGL3 vector by miR-9 (Figure 4.7B) or by isomiR-9 (Figure 4.7C). Figure 4.7B shows that the repression caused by miR-9 was alleviated by the introduction of 100 ng of miR-9 sponge but not by the isomiR-9 sponge. By contrast, isomiR-9 sponge partially alleviated the repression of isomiR-9 on DNMT3B-3’UTR but sponge miR-9 did not (Figure 4.7C). NCAM2 was also tested by these sponges (Figure 4.7D). Similarly, isomiR-9 sponge partially rescued the repression by isomiR-9 on NCAM2-3'UTR whereas miR-9 sponge did not.

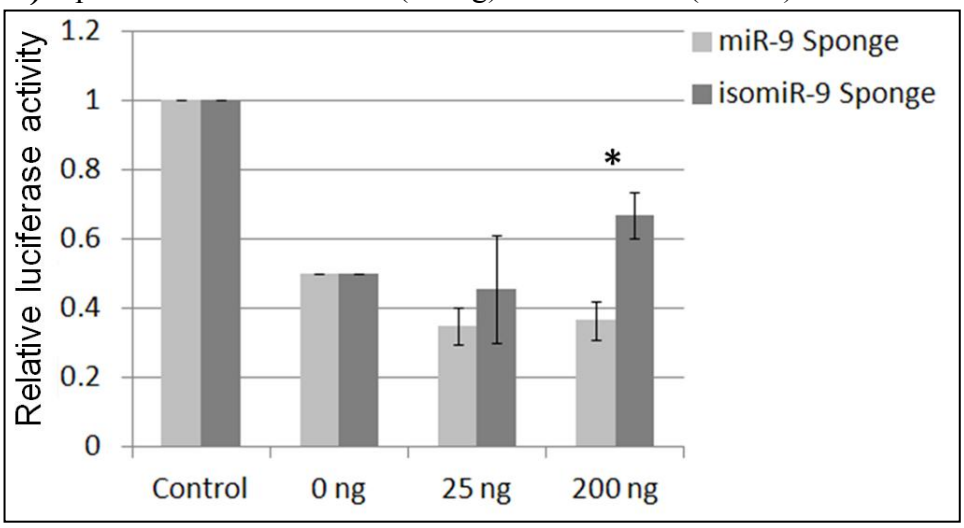
A)



B) pGL3-CDH1-3'UTR (400ng) and miR-9 (12 nM)



C) pGL3-DNMT3B-3'UTR (400ng) and isomiR-9 (12 nM)



D) pMIR-NCAM2-3'UTR (200ng) and isomiR-9 (12 nM)

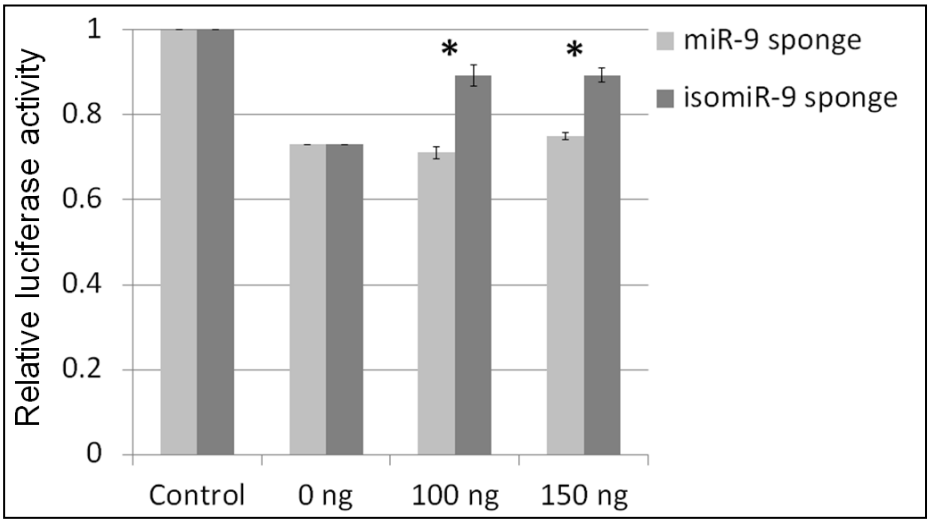


Figure 4.7
Sponge inhibitors of miR-9 and isomiR-9. A) Structure of sponge constructs pcDNA-miR-9 sponge and pcDNA-isomiR-9 sponge. HEK293 cells were transfected with the indicated concentrations of sponge vectors and either B) pGL3-CDH1-3'UTR (400ng) and miR-9, C) pGL3-DNMT3B-3'UTR (400ng) and isomiR-9 or D) pMIR-NCAM2-3'UTR (200ng) and isomiR-9. Control: Reporter vectors only. 0 ng: Indicates reporter vector and miR-9 or isomiR-9 (12 nM). 25 ng, 100 ng, 150 ng and 200 ng are the amount of sponge DNA that was introduced. D) All results were normalised by renilla luciferase. Error bars represent the standard deviation obtained from three independent experiments (n=3). * indicate p value <0.05 (statistical difference is between miR-9 sponge and isomiR-9 sponge).

4.2.6 In search of a cell line that expresses DNMT3B or NCAM2

To further test whether isomiR-9 could repress DNMT3B in an endogenous system, a cell line that expressed DNMT3B is required. Western blotting was performed to look for cell lines that express DNMT3B. Based on the limited number of cell lines that were tested, hESC was the only cell line that expresses DNMT3B protein (Figure 4.8). As hESC is a hard to transfect cells, I screened for the presence of an alternate target of isomiR-9 NCAM2 and recently found LNCaP cell line (a human prostate adenocarcinoma derived from metastatic supraclavicular lymph node)(a gift from Alwyn Dart, member of Charlotte Bevan group) expresses NCAM2 protein (Figure 4.8).

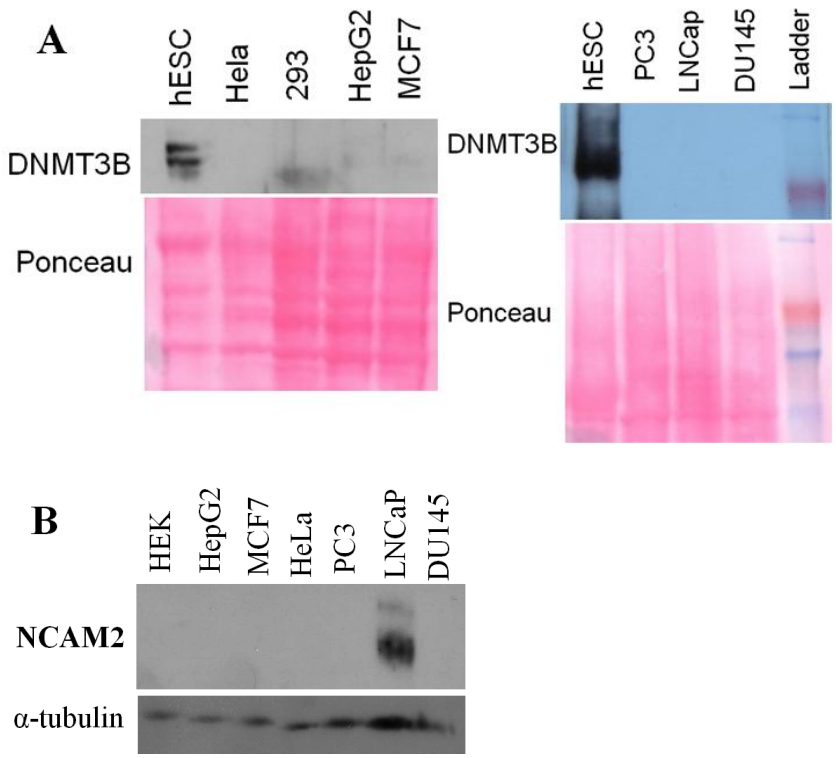


Figure 4.8
DNMT3B and NCAM2 protein expressions.
A) Western blotting results of DNMT3B antibody (Santa Cruz, 97kDa) on the indicated cell lines. Ponceau staining of the western blot membrane was used as loading control. Orange band of the ColorPlus prestained protein ladder (NEB UK) represents 80kDa. B) NCAM2 western blotting on different cell lines. α-tubulin was used as loading control.

4.2.7 Infection is the preferred method of introducing miRNA into hESCs

This experiment was designed to compare the efficiency of transfection and infection of hESCs. A red fluorescent tag miRNA mimic (Dharmacon) was used in the transfection and lentiviral infection was performed using PLVTHM, a lentiviral vector that expresses GFP. As expected, only an average of 5% of hESCs was transfected compared to 60% in HEK293 cells. In contrast, an average of 50% of hESCs was infected (Figure 4.9A) compared to 70% in HEK293 cells. There was therefore a 10-fold difference in efficiency between transfection and transduction of hESCs (Figure 4.9B).

To be certain that hESCs were the actual cells that were infected, rather than just stromal or differentiated cells, pRRL-cPPT-PGK-dsred lentiviral vector that expresses red fluorescent protein was used to infect T5-Oct4 GFP transgenic hESCs. The rationale was that if the green cells also express red fluorescent protein then infection has occurred in the hESCs. In this experiment, 42% of the cells were infected were both red and green fluorescent, indicating that 42% of the infected cells were hESCs. This constitutes about 2/3 of the stem cell population (Figure 4.9C). This was confirmed by direct visualisation of cells expressing green and red fluorescent proteins with a fluorescent microscope (Figure 4.9E). This result shows miRNA sponges can be effectively introduced into hESCs by lentiviruses.

It is important to note that lentivirus might not be a good way of introducing isomiR-9 ectopic expression as they will be expressed in their primary or precursor form and therefore be subjected to the usual processing that will generate all other isomiRs including the canonical miRNA.

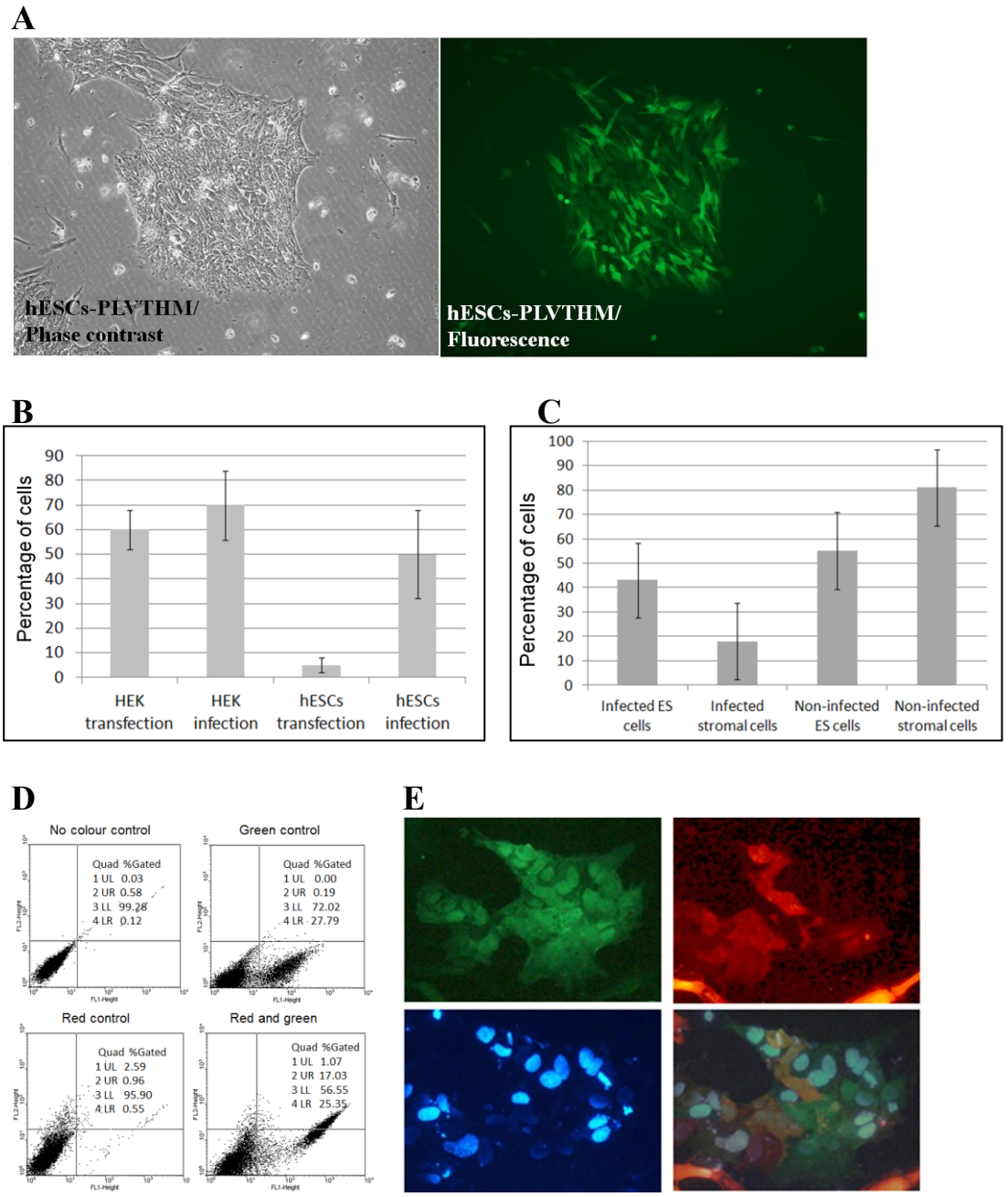


Figure 4.9

Transfection and viral infection efficiency in HEK293 and hESCs. A) PLVTHM-infected hESCs express GFP. B) Transfection was performed using red fluorescent tag miRNA mimic and infection using lentiviral vector that expresses GFP. Bar graph showed the percentage of cells that were transfected or infected. C) T5 Oct4-GFP transgenic hESCs were infected by PGK-RFP lentivirus. Bar graph showed the percentage of infected and non-infected hESCs and stromal cells. D) FACS analysis shows red and green compensation was performed before the evaluation of percentage of cells that have red and green colours. E) Red & green cells represent infected hESCs, red only cells represent infected stromal/ differentiated cells, green only cells represent uninfected hESCs and non-green/non-red cells represent uninfected stromal/ differentiated cells. Error bars represent the standard deviation obtained from two independent experiments (n=2).

4.2.8 Construction of a DNMT3B expressing vector

Returning to the DNMT3B study, as hESCs are hard to transfect, in order to evaluate whether isomiR-9 can knockdown DNMT3B, I constructed a vector that expresses DNMT3B along with its 3' UTR. A plasmid containing the full length coding region of DNMT3B (2562bp) was obtained from addgene (Plasmid 35522: pcDNA3/Myc-DNMT3B1) (Chen et al., 2005). The 3' UTR of DNMT3B was amplified from human genomic DNA (1560bp) by PCR and a full length DNMT3B with its 3' UTR was then constructed (see Materials and methods). This gene was first ligated into pGEM-T easy vector and sequenced. Then DNMT3B 3' UTR was cloned into pcDNA3.1(+) between BamHI and XbaI sites. DNMT3B expression was confirmed by transfection of pcDNA-DNMT3B into HEK293 cells using HiPerfect (Figure 4.10).

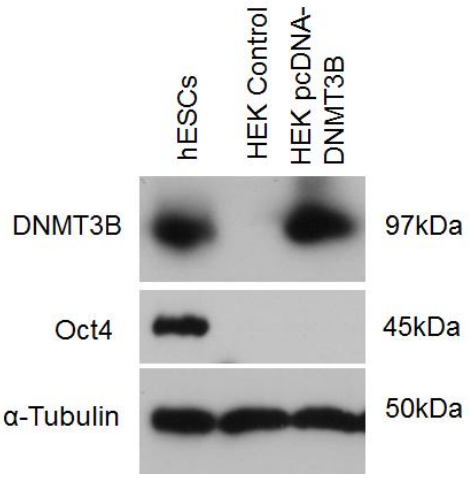


Figure 4.10
Ectopic expression of DNMT3B in HEK293 cells. 400 ng of pcDNA-DNMT3B was transfected into HEK293 cells (lane 3) and protein was extracted using RIPA buffer after 48 hours. hESCs control and HEK cells control were in lane 1 and 2 respectively. α-tubulin was used as loading control.

4.3 Discussion

Here we show that it is possible to make sponge vectors that can distinguish between miR-9 and isomiR-9 (Figure 4.7), which adds further assurance that isomiRs can recognise different targets to canonical/ annotated miRNAs. In principle, this advance will allow us to test the biological significance of isomiR production by selective sequestration of isomiRs in appropriate model systems.

As expected, miR-9 and isomiR-9 showed different targeting of the UTRs of CDH1 and DNMT3B when these were cloned into a different luciferase vector (Figures 4.1 and 4.2). However, it was more difficult to show selective targeting by the sponge regions with 6 target sites for CDH1 or DNMT3B (Figure 4.4 and 4.6), when these were cloned into a luciferase vector (Figures 4.4 to 4.6). Our analysis indicates that specificity can be lost by increases in the number of MBS and by the use of high concentrations of miRNAs (Figure 4.5).

The sponge regions worked better when they were used as decoy mRNAs that sequestered miRNAs from the co-transfected luciferase vectors (Figure 4.7). Nevertheless, the results of Figures 4.4 to 4.6 indicate that the effectiveness of sponges in general is likely to be dependent upon the relative concentration of endogenous miRNA and sponge expression level (Figure 4.11). For sponges that are trying to distinguish between very similar miRNAs and isomiRs, it would seem particularly important not to express an excessive amount of decoy mRNA relative to miRNA levels.

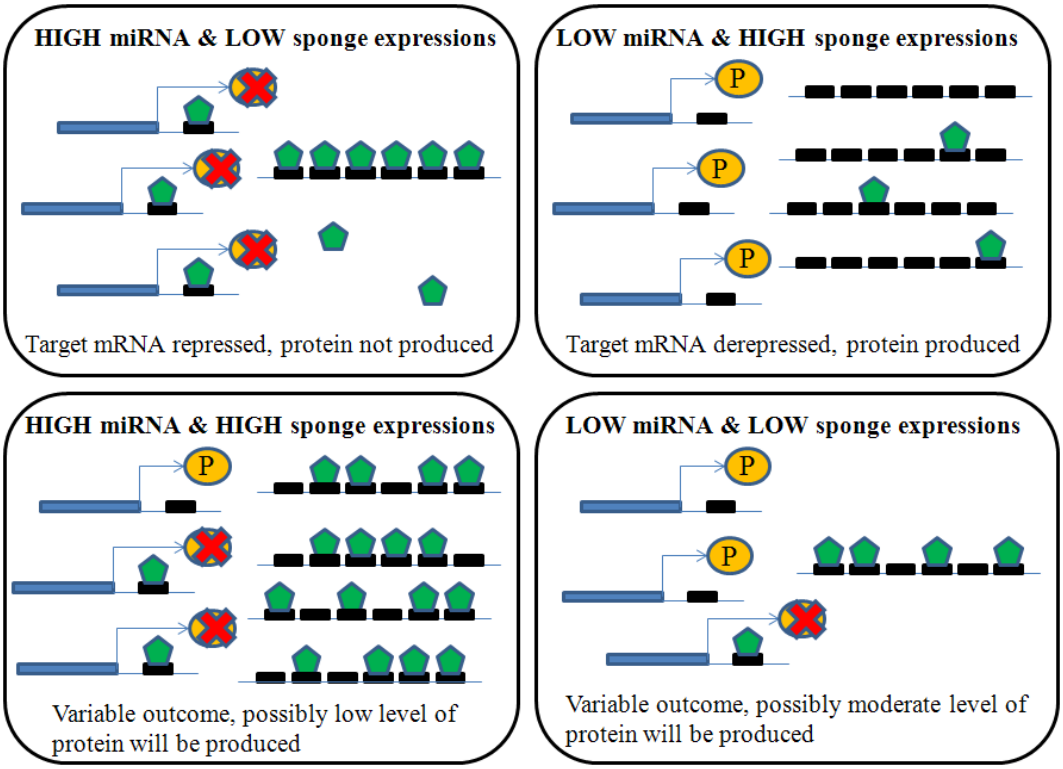


Figure 4.11
Sponges compete with target mRNA for binding with miRNA and the various outcomes as a result of the concentration differences between the miRNA and sponge. Long blue bar – target mRNA; short black bar – miRNA binding site; orange oval – protein; green pentAgon – endogenous miRNA; red cross – protein not produced; arrow – protein translation. In an environment where there is high level of miRNA concentration but low sponge expression, the most likely outcome is protein will not be produced. Conversely, if there is high sponge expression coupled with a low miRNA concentration, most invariably protein will be produced. The situation becomes unpredictable when there is either high level of both miRNA and sponge or low level of both miRNA and sponge.

A miR-9 sponge can repress genes such as CDH1 in breast cancer (Ma et al., 2010) and FoxP1 in chick spinal cord (Otaegi et al., 2011). Sponges therefore might have therapeutic potential for cancer treatment. MiR-21 and miR-221, which target PTEN, are also overexpressed in a variety of tumours (Garofalo et al., 2011).

Our results show that an “isomiR-9 sponge” could specifically sequester isomiR-9 at a better efficiency than the canonical miR-9 with one base difference at the 5' end, and *vice-versa*. This shows that miRNA sponge could potentially be used to knockdown specific isomiRs. It would be interesting to investigate whether isomiR-9 or its sponge could be of therapeutic important in breast cancer cells that overexpress DNMT3B (Sandhu et al., 2012).

In future, my plan is to investigate the biological effect of miR-9/isomiR-9 knockdown in neural stem cells. This could be achieved by either introducing miR/isomiR-9 sponges directly into NSC or by first establishing stable transgene expression of a sponge in hESCs by lentiviral infection (Figure 4.9) and then proceeding to neural differentiation. Our results show that hESCs are hard to transfect, indicating that a lentiviral system will be the best way for sponge delivery. A previous study also showed that a lentiviral system gave better gene delivery, in comparison with transfection. Nucleofection is another alternative but it was associated with low survival rate (Cao et al. 2010).

Chapter 5 MicroRNA 302 cluster and somatic cell reprogramming

5.1 Introduction

The polycistronic miR-302 cluster encodes five miRNA genes that have an important role in the regulation of embryonic stem function. These five miRNA genes include miR-302a, b, c and d, and miR-367. The cluster is located on chromosome 4 (Figure 5.1) within an intron of and in the opposite orientation to the gene LARP7, which unlike the miR-302 cluster is ubiquitously expressed. Most functional studies have been of miR-302a as this is generally consider as the functional guide strand and has a common seed region with other members of this cluster, namely miR-302b, c and d (Rosa et al., 2009; Rosa et al., 2011; Barroso-delJesus et al., 2011). Inhibition of miR-302 in stem cells resulted in the downregulation of pluripotency markers and *vice versa* (Rosa et al., 2009). Furthermore, global loss of miRNA in DGCR8 deficient stem cells resulted in defects in proliferation and differentiation (Gangaraju et al., 2009).

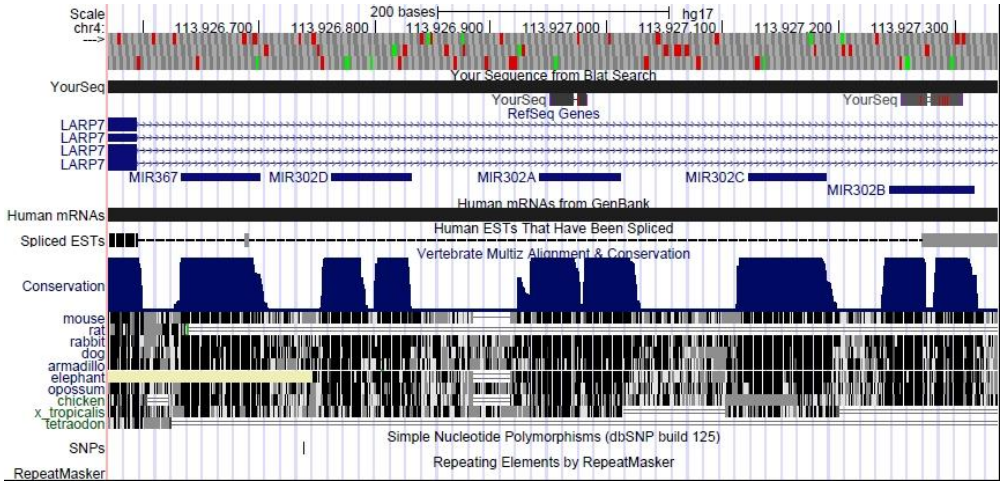


Figure 5.1
The polycistronic miR-302 cluster is conserved. Image was taken from UCSC Genome Browser website (http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=280340015).

Table 5.1 lists a collection of sequencing data of members of the miR-302 cluster from 5 different sources, including ours (Suh et al., 2004; Bar et al., 2008; Morin et al., 2008; Lipchina et al., 2011; Chan, unpublished). The differences in expression levels of miRNAs such as miR-302a to d and miR-367 that are transcribed from the same promoter is an interesting feature of many miRNA clusters (Table 5.1).

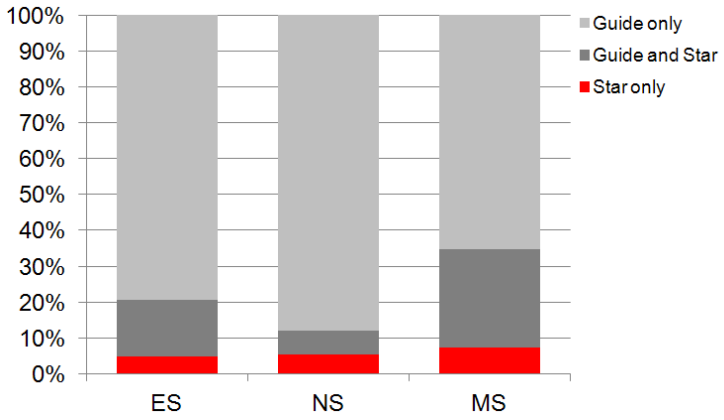
Source	**Suh et al., 2004**	**Bar et al., 2008**	**Morin et al., 2008**	**Lipchina et al., 2011**		**Chan unpublished**
Species/ Methods	Human	Human 454	Human Illumina	Human 454	Human Solexa	Human Solexa
302a	0.219	0.045	0.519	0.406	0.062	0.037
302a*	0.057	0	0.074	0	0.330	0.838
302b	0.429	0.402	0.214	0.015	0.255	0.041
302b*	0.019	0	0.005	0	0.007	0.002
302c	0.095	0.120	0.055	0.059	0.087	0.024
302c*	0.009	0	0.008	0	0.012	0
302d	0.124	0.433	0.121	0.244	0.080	0.027
302d*	0.048	0	0.004	0	0.002	0.016
367	0	0	0	0.276	0.164	0.015
367*	0	0	0	0	0.001	0

Table 5.1
Comparison of the fraction between members of miR-302 cluster (302a, 302a*, 302b, 302b*, 302c, 302c*, 302d, 302d*, 367 and 367*) in human embryonic stem cells of selected publications. Numbers that are highlighted in grey denote the most highly sequenced miRNA in the cluster. The numbers are expressed as fractions of the total sequencing reads. 454, Illumina and Solexa represent the sequencing platforms that were used in the deep sequencing.

MiRNA* is derived from the opposite arm to the guide strand in the precursor miRNA, and is usually detected at lower frequency than the guide stand miRNAs (Bartel, 2004; Lagos-Quintana et al., 2002; Aravin et al., 2003; Lim et al., 2003). Interestingly, our sequencing results revealed that miR-302a* was the most highly sequenced miRNA in human embryonic stem cells from the miR-302 cluster, some 20

fold more than the guide strand of miR-302, Lipchina et al., (2011) also reported that miR-302a* was the most highly sequence miRNA (Table 5.1).

Overall, we found that miRNA* strands were, as expected, expressed at much lower frequencies in all 3 stem cell lines (Figure 5.2). Across the three cell types, on average 5.8% of miRNA genes expressed only the star strand and 16.6% expressed both miRNA/miRNA* (Figure 5.2 and Table 5.2). Intriguingly, there were cell lines that expressed only the opposite strands, for example hESCs expressed miR-30e* only, while NSCs expressed miR-30e only (Table S5.1).



Cell types	Star only (%)	Guide and Star (%)	Guide only (%)
hESC	4.8	15.8	79.4
NSC	5.3	6.7	88
MSC	7.4	27.3	65.3
Average	5.8	16.6	77.6

Figure 5.2 and Table 5.2
Figure and table illustrate the percentage of miRNA genes that encode only guide, star or both miRNA strands in the deep sequencing results of combined human embryonic stem cells, neural stem cells and mesenchymal stem cells. The deep sequencing experiment was performed by Elcie Chan (unpublished). hESC – Human embryonic stem cells; NSC – Neural stem cells; MSC – Mesenchymal stem cells.

Recently, miRNA* was reported to be associated with Ago protein (Okamura et al., 2008), which is consistent with my northern blotting results where miR-302a* was detected in Ago 1 and 2 immunoprecipitations (Figure 3.4). Similarly to our deep sequencing data, Jagadeeswaran et al., (2010) observed that some miRNA* were expressed at a higher level than the corresponding guide strands (Table S5.1). Consequently, miRBase has replaced the star sign with either miR-5p or -3p. As expected, target prediction studies showed that the mRNA targets of opposite arms differ significantly (Griffiths-Jones et al., 2011). The star sequence of miR-367 has only been detected at a low level in all sequencing studies (Table 5.1), indicating that it is far less likely to have a biological function compared to miR-302a* or miR-302d*.

Forced expression of the miR-302 cluster can reprogram somatic cells to pluripotent stem cells or can enhance the production of stem cells by OSKM factors. Table 5.3 lists the various strategies that have been used to generate iPSCs through the use of miRNAs, and also shows the estimated efficiencies, where available. Some of the efficiency levels that are reported are orders of magnitude greater than the the OSKM method (Anokye-Danso et al., 2011). Induced pluripotent stem cells (IPSCs) could potentially be used as: disease models, for example, spinal muscular atrophy (Ebert et al., 2009) and LEOPARD syndrome (Carvajal-Vergara et al., 2010); drug testing and regenerative medicine (Wu et al., 2011). The mechanism by which stem cells can be reprogrammed from somatic cells is an area of great interest. For the miR-302 cluster it would appear that the expression of miR-302a and miR-367 is important (Anokye-Danso et al., 2011) but the potential contribution of the miR-302a* has not been addressed.

Recently, in addition to IPSCs, there were reports of miRNA-mediated conversion of fibroblasts to neurones (miR-9/9*, miR-124 and NeuroD2; Yoo et al., 2011) and cardiomyocytes (miR-1, miR-133, miR-208 and miR-499; Jayawardena et al., 2012). Here, a miR-302 cluster lentivirus was constructed in order to test the reproducibility of somatic cell reprogramming and then to use this technology to investigate the mechanism of reprogramming by the miR-302 cluster. We also wanted to find out whether the star/ passenger strand of miR-302a is important for somatic cell reprogramming

No	Method	Cells	Efficiency	Authors
1	Transduction; Retrovirus 302a,b,c,d	Human melanoma cells (Colo) Human prostate cancer cells (PC3)	2-5%	Lin et al., 2008
2	Transfection 291; 294; 295; 302d/ Transduction; Retrovirus OSK	Mouse embryonic fibroblasts	0.1-0.3% (294 and OSK)	Judson et al., 2009
3	Transduction 302a,b,c,d	Human hair follicle cells	-	Lin et al., 2011
4	Transduction; Retrovirus Cluster A: 200b, 200a, 429 Cluster B: 106a, 18b, 20b, 19b, 92a, 363 Cluster C: 302a,b,c,d, 367 OSK/OSKM	Mouse embryonic fibroblasts	Cluster B and C enhanced reprogramming by OSK/OSKM factors	Liao et al., 2011
5	Transfection 302b, 372/ Transduction; Retrovirus OSK/OSKM	Human foreskin (BJ) Lung fibroblasts (MRC5)	302b and/or 372 enhanced reprogramming by OSK/OSKM factors	Subramanyam et al., 2011
6	Transduction; Lentivirus 302a,b,c,d/ 302a,b,c,d,367	Mouse embryonic fibroblasts, Human dermal fibroblasts	10%	Anokye-Danso et al., 2011
7	Transfection 200c, 302a,b,c,d, 369-3p,-5p	Mouse adipose stromal cells, Human adipose stroma cells, Human dermal fibroblasts	0.0001% (mouse) 0.0002% (human)	Miyoshi et al., 2011

Table 5.3
List of publications in somatic cell reprogramming using miRNAs (Taken from Tan et al., 2012)

5.2 Results

5.2.1 Characteristics of miR-302 cluster

Table 5.4 shows that miR-302a to d have a common seed region, which is different to the conserved seed regions of the star miRNAs. MiR-367 has a distinctive seed region that is conserved with other species (Figure 5.1).

Mature miRNA sequence	MiRNA name
UAAGUGCUUCCAUGUUUUGGUGA	hsa-miR-302a-3p MIMAT0000684
UAAGUGCUUCCAUGUUUUAGUAG	hsa-miR-302b-3p MIMAT0000715
UAAGUGCUUCCAUGUUUCAGUGG	hsa-miR-302c-3p MIMAT0000717
UAAGUGCUUCCAUGUUUGAGUGU	hsa-miR-302d-3p MIMAT0000718
ACUUAAACGUGGAUGUACUUGCU	hsa-miR-302a-5p/ 302a* MIMAT0000683
ACUUUAACAUGGAAGUGCUUUC	hsa-miR-302b-5p/ 302b* MIMAT0000714
UUUAACAUGGGGGUACCUGCUG	hsa-miR-302c-5p/ 302c* MIMAT0000716
ACUUUAACAUGGAGGCACUUGC	hsa-miR-302d-5p/ 302d* MIMAT0004685
AAUUGCACUUUAGCAAUGGUGA	hsa-miR-367-3p MIMAT0000719
ACUGUUGCUAAUAUGCAACUCU	hsa-miR-367-5p/ 367* MIMAT0004686

Table 5.4
Table lists the members of miR-302 cluster with emphasis on the seed sequence difference between the guide and star strands. Highlighted sequences denote seed region. Yellow represents the common seeds for miR-302a/b/c/d. Light blue represents the common seed for miR-302a*/b*/c*/d* and miR367*. Green represents the seed region of miR-367. Grey areas represent variation of sequences between miRNAs.

Table 5.5 lists the numbers of predicted targets of each member of the miR-302 cluster and Figure 5.3 shows which of these targets are in common. The SP3 transcriptional factor is a predicted target of five miRNAs including miR-302a* and was selected for further investigation, largely because studies suggest that SP3 binds to the promoter region of Oct4 and Nanog genes and might regulate their expression

(Pesce et al., 1999; Wu et al., 2006). ZNF148 gene was of interest as there are miRNA binding sites in its UTR for miR-302a/b/c/d, miR-302a* and miR-367 (Figure 5.3 and Table S5.2). In figure 5.3, the Venn diagram shows target predictions of miR-302a/b/c/d (similar seed region – see table 5.4), miR-302a*, miR-302b*/d* and miR-367. MiR-302c* and miR-367 were not analysed because the deep sequencing reads for these two miRNAs were zero (see table 5.1). These miRNAs represent 3 of the most abundantly express miRNAs in the cluster. In addition, 3 different databases independently predicted that SP3 is a target of miR-302a* (Figure S5.1).

MiRNA	Number of targets
302a=302b=302c=302d	445
302b*=302d*	405
302a*	361
302c*	483
367	567
367*	125

Table 5.5
Table shows the total number of predicted targets of members of the miR-302 cluster. Target prediction was performed using Targetscan Custom Human 4.1.

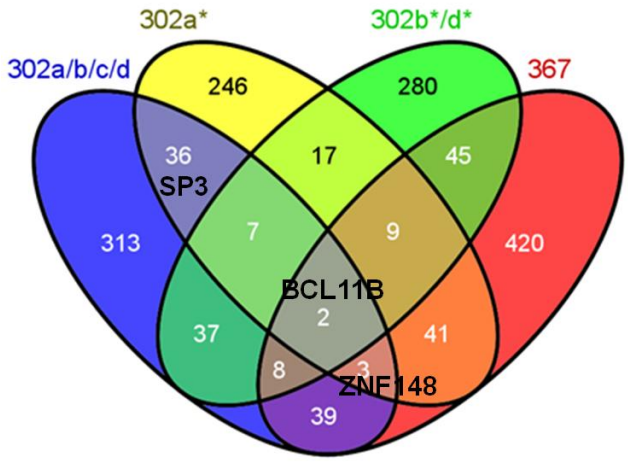


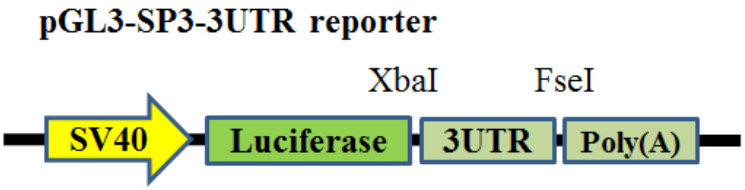
Figure 5.3
Venn diagram shows the number of predicted targets that are either shared by or unique to the members of the miR-302 cluster. SP3 is a target common to miR-302a/b/c/d and miR-302a*. ZNF148 is a target common to miR-302a/b/c/d, miR-302a* and miR-367. BCL11B is a target that is common to all 4 groups of miR-302 cluster. Targets that are in common are listed in Table S5.2.

5.2.2 Target evaluation of SP3 and ZNF148 reporters by luciferase assays

To test whether miR-302a* can repress its predicted targets, SP3 and ZNF148 luciferase reporters were constructed. A segment of SP3 and ZNF148 3'UTRs were amplified and ligated to pGEM-T easy vector and sequence verified. Finally, SP3-3'UTR (795bp) was cloned into Xba I and Fse I sites at positions 1934 and 1953 of pGL3 control vector (Figure 5.4A and Figure S2.2). While ZNF148-3'UTR (429bp) was cloned into Spe I and Sac I sites at positions 525 and 519 of pMIR report vector (Figure 5.5A and Figure S5.2).

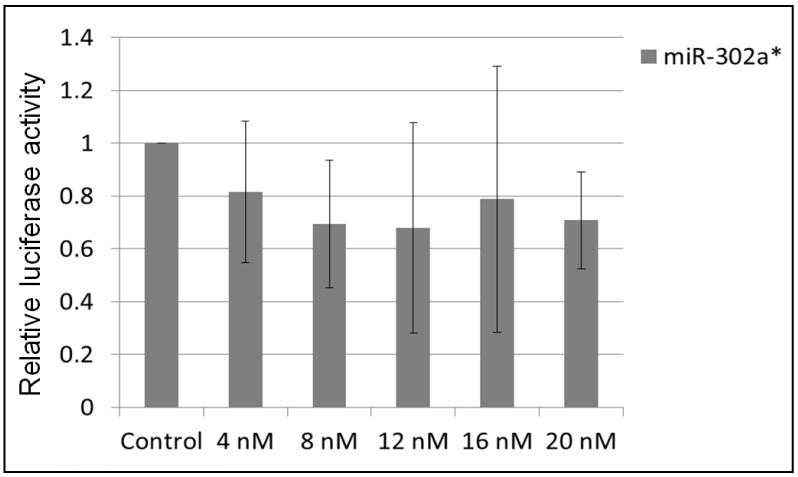
MiR-302a* was unable to consistently repress SP3, even at different miRNA concentrations (Figure 5.4B). Again, as seen earlier in figure 3.8A, the luciferase activity increased after transfection with let-7d (Figure 5.4). In ZNF148, at 4 nM the repression was slight >10% but the difference from let-7d was only marginally significant (p value = 0.049). There was no statistical difference between miR-302a* and let-7d at 12 nM and 20 nM (Figure 5.5B).

A



B

```
5′.CCUGUUGAAAAGCCAUGUUUAAGAU..3′  SP3-3′UTR
                 ||o|||||||
3′  UCGUUCAUGUAGGUGCAAAUUCA   5′  miR-302a* /302a-5p
                   +|+o+++
3′   UUGAUACGUUGGAUGAUGGAGA   5′  let-7d
```



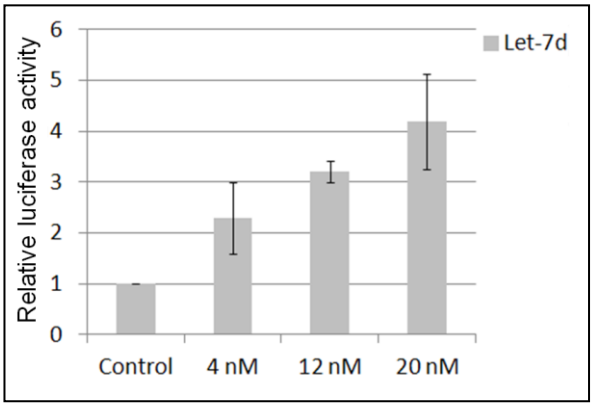
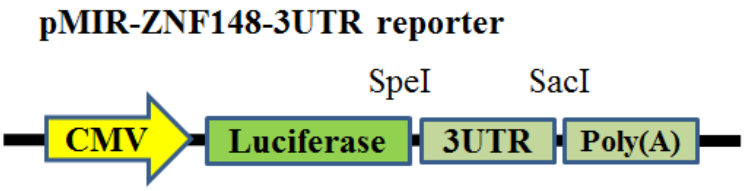


Figure 5.4 SP3 3'UTR reporter assay
A) In pGL3 vector, the luciferase expression is driven by SV40 promoter. SP3 3' UTR was inserted downstream to the luciferase sequence. B) pGL3-SP3-3'UTR was co-transfected with miR-302a* or let-7d into HEK293 cells. Relative activity of the firefly luciferase for SP3 reporter was plotted against increasing concentrations of miR-302a*. Control denotes transfection of reporter vector only. Error bars represent the standard deviation obtained from six independent experiments (n=6) for SP3 and three independent experiments (n=3) for let-7d. All results were normalised by renilla luciferase.

A



B

```
5′..GUUUUUACAGCUUGUUUUUAAGAAU..3′ ZNF148-3′UTR
                    ||||||
3′  UCGUUCAUGUAGGUGCAAAUUCA    5′ miR-302a* /302a-5p
                    |+o+++
3′   UUGAUACGUUGGAUGAUGGAGA    5′ let-7d
```

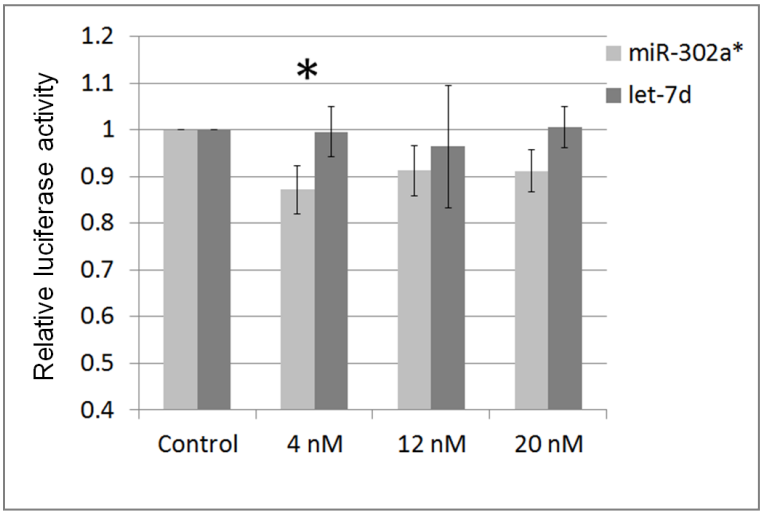


Figure 5.5 ZNF148 3'UTR reporter assay
A) In pMIR vector, the luciferase expression is driven by CMV promoter. ZNF148 3' UTR was inserted downstream to the luciferase sequence. B) pMIR-ZNF148-3'UTR was co-transfected with either miR-302a* and let-7d into HEK293 cells. Relative activity of the firefly luciferase for ZNF148 reporter was plotted against increasing concentration of miRNAs. Control denotes transfection of reporter vector only. Error bars represent the standard deviation obtained from three independent experiments (n=3). All results were normalised by renilla luciferase. * indicate p value is <0.05 (statistical difference is between miR-302a* and let-7d).

5.2.3 Construction of a lentiviral vector that expresses miR-302 cluster

MiR-302 cluster comprising of miR-302b, miR-302c, miR-302a, miR-302d and miR-367, accompanied by 120bp upstream and 150bp downstream of the cluster (975bp) was amplified by PCR from human genomic DNA (Figure S5.3). The amplified fragment was ligated into pGEM-T easy vector and verified by sequencing. Finally, it was cloned into XhoI and MluI sites at position 3806 and 4064 of pTRIPz inducible lentiviral vector (Figure 5.6 and Figure S5.4).

As this vector has a red fluorescent protein (RFP) marker, the pTRIPz-302 cluster lentivirus was first tested by infecting HEK293 (human embryonic kidney) and MRC5 (human lung fibroblasts) cells to observe for RFP. Doxycycline induced, lentiviral infected HEK293 and MRC5 cells expressed RFP (Figure 5.7A). Subsequently, northern blots of total RNAs collected from the infected HEK293 and MRC5 cells showed miRNA expressions from members of the miR-302 cluster, i.e., miR-302a, miR-302a* and miR-367 (Figure 5.7B).

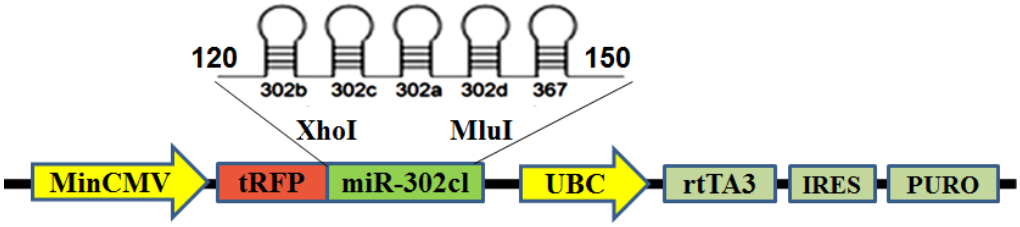


Figure 5.6
MiR-302 cluster in the pTRIPz-miR-302 cluster lentiviral vector is driven by minimal CMV with tetracycline response element. Its expression can be monitored by turbo red fluorescent protein expression. 120 and 150 represent the length of nts extended upstream and downstream from the miR-302 cluster gene. MinCMV – Minimal cytomegalovirus promoter; tRFP – turbo red fluorescent protein; miR-302cl – miR-302 cluster; UBC – Ubiquitin promoter; rtTA3 – Reverse transactivator; IRES – Internal ribosome entry site; PURO – Puromycin.

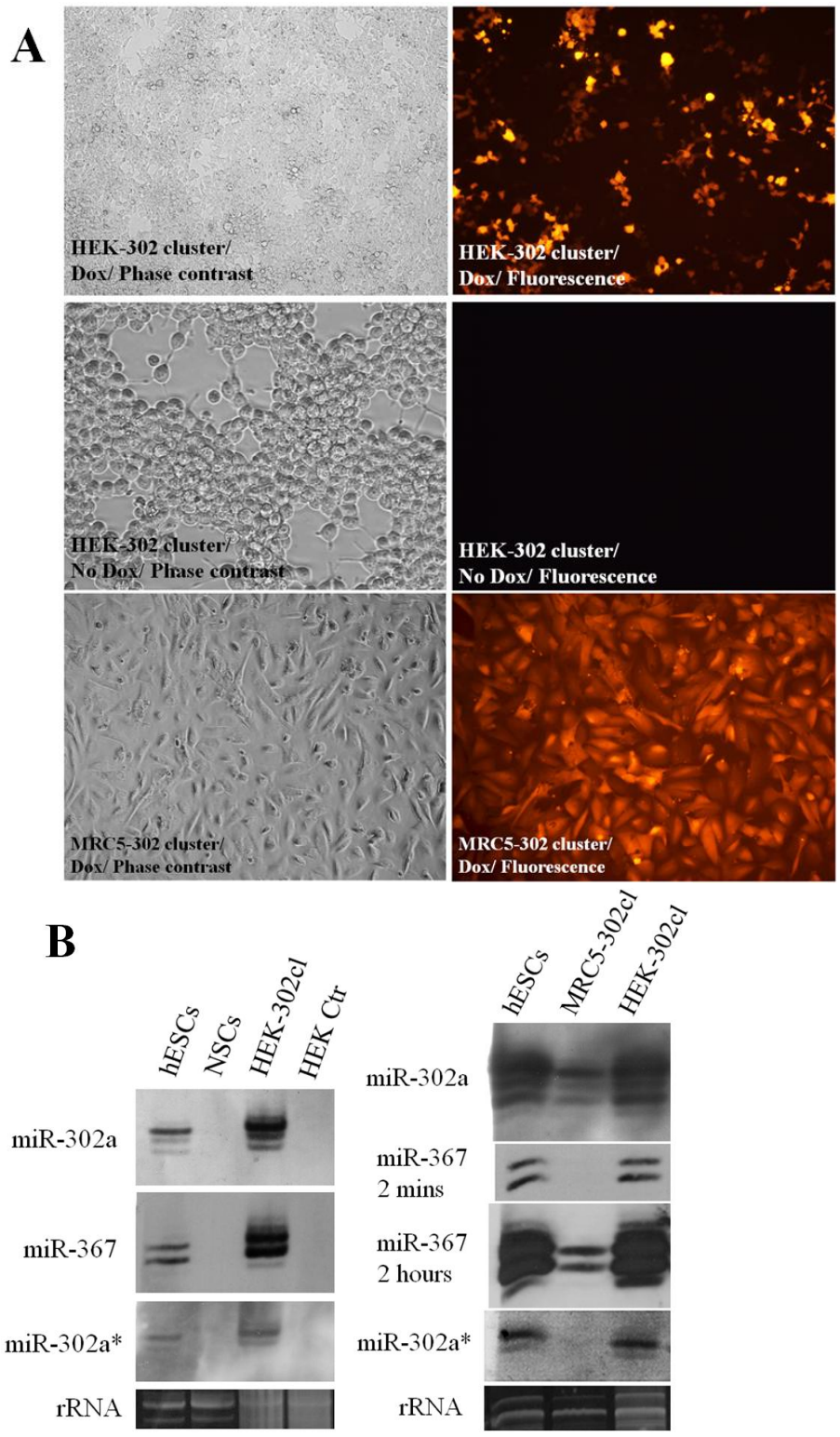


Figure 5.7
MiR-302 cluster expression in HEK and MRC5 cells
A) HEK and MRC5-infected cells expressed RFP after induction with doxycycline.
B) Northern blots of infected HEK and MRC5 cells were probed for miR-302a, miR-367 and miR-302a*. rRNA stained with ethidium bromide was used as loading control. hESCs- human embryonic stem cells; NSCs- neural stem cells; HEK- human embryonic kidney cells; 302cl- miR-302 cluster; Ctr-Control.

5.2.4 Evaluation of miR-302 cluster in the reprogramming of human lung fibroblasts

To test the potential of miR-302 cluster in somatic cell reprogramming, MRC5 cells were infected with the pTRIPz-302 cluster lentivirus and cultured in hESC conditions (matrigel coated plate and MEF-conditioned media). Cell colonies started to appear 6 to 8 weeks after infection (Figure 5.8A). This was seen in 2 out of 5 attempts.

These cells expressed RFP (Figure 5.8B), formed colonies (Figure 5.8C) and were faintly positive toward alkaline phosphatase (Figure 5.8D). In addition, RT-PCR showed these cells expressed low level of DNMT3B and Nanog, equivocal Oct4 and Lin28, but they did not express Sox2 (Figure 5.9), suggesting that they were not fully reprogrammed. This might be due to the low level of miR-302 cluster expression by pTRIPz-302 cluster in MRC5 cells.

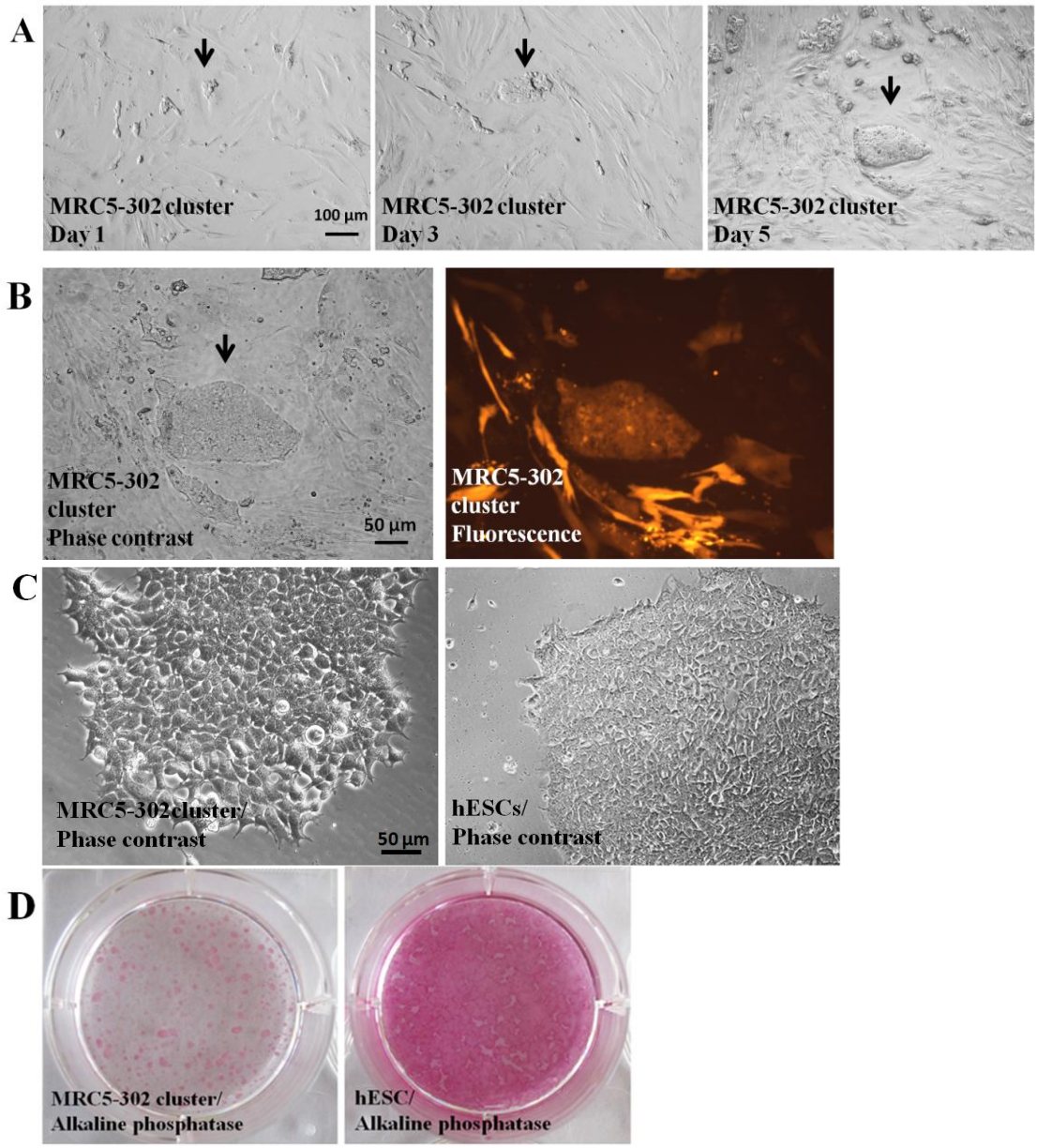


Figure 5.8
pTRIPz-302 custer lentivirus infection in MRC5 cells
MRC5 cells were infected with pTRIPz-302 cluster lentivirus, cultured in hESCs condition. Cell colonies appeared approximately 6-8 weeks after infection. A) Cell colonies at day 1, day 3 and day 5 from the day of appearance. Black arrows show cells forming colony. B) These cells expressed red fluorescent protein indicates that they were infected. C) Comparison of the morphology between the colony and hESCs colony. D) Alkaline phosphatase staining of the cell colonies and hESCs.

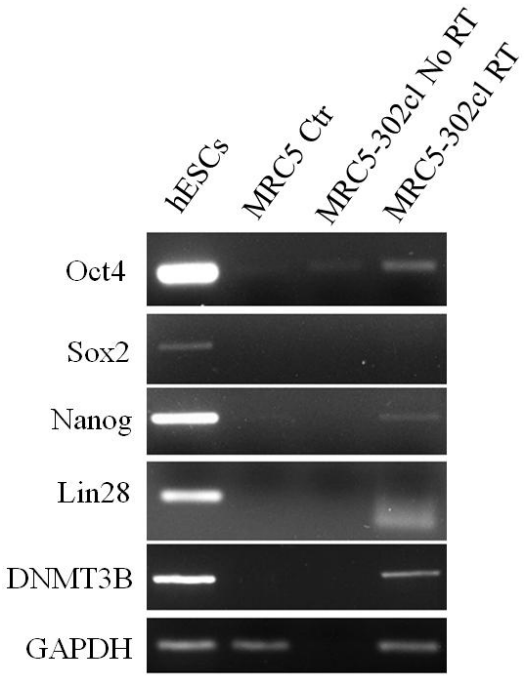


Figure 5.9
Comparison of pluripotency gene expressions between hESCs and infected MRC5
Total RNAs were extracted from wild type/ control and pTRIPz-302 cluster virus infected MRC5 cells with and without reverse transcription. The pluripotency genes were compared between the MRC5 cells and hESCs. Ctr- Control; 302cl- 302 cluster; RT- reverse transcription.

5.3 Discussion

For some miRNA genes, strand switching occurs during development so that the star strand becomes the dominant miRNA (Griffith-Jones et al., 2011), which is presumably reflective of our observation that about 6% of the miRNAs that we sequenced expressed the star strand rather than the guide strand listed in miRBase (Figure 5.2). We only sequenced miRNA from three cell types, which raises the possibility that the star strand of other miRNA genes might be more expressed in a different cell background. Our sequencing data also shows that some 15% of miRNA genes that are expressed in our libraries make substantial amounts of both the guide and star strand. Overall we suspect that star strand production is currently underestimated.

Several star strands have been established to be functional (Okamura et al., 2008; Qu et al, 2012), however, as yet there have been no publications as to the function of miR-302a*. The high level of expression of miR-302a* (Table 5.1) suggests that it is likely to have an important function in stem cells and it may also be important for the induction of stem from somatic cells, as our results indicate that miR-302a* is likely to have been produced by vectors previously used in iPSC studies (Figure 5.7B and Table 5.3).

I found that miR-302a* was able to repress ZNF148 at 4 nM, although the repression was not that convincing because the repression was inconsistent at higher miRNA concentrations (Figure 5.5). For this reason we did not undertake a mutational analysis in order to validate target specificity. It might be of interest to test whether the guide strands of miR-302a can repress the expression of ZNF148, alone or in

combination with miR-302a*, as there are target sites for miR-302a and 302a*. The mRNA encoded by SP3 was also not clearly inhibited by miR-302a*, despite predictions to the contrary (Figure 5.4).

To date, three groups have successfully reprogrammed somatic cells to a pluripotent state by using miRNA alone (Lin et al., 2008; Anokye-Danso et al., 2011; Miyoshi et al., 2012). Other groups have shown that miRNAs can enhance reprogramming by OSKM factors (Oct4, Sox2, Klf4 and Myc) (Liao et al., 2011; Subramanyam et al., 2011). Anokye-Danso et al., (2011) reported that they have achieved a reprogramming efficiency with miR-302 cluster alone that was 2 orders of magnitude higher than conventional reprogramming using OSKM factors. In addition, somatic cell reprogramming was successfully performed using mature miRNA by simple transfection (Myoshi et al., 2011).

Our data shows that pTRIPz-302 cluster did not fully reprogram the human lung fibroblasts to a pluripotent state, probably due to low miRNA expression. In the Lin et al., (2011) paper, the authors show that reprogramming of human hair follicle cells would only take place above a certain level of miRNA expression, approximately 1.5 fold of the miRNA expression in hESCs. Although we achieved a high level of expression of the miR-302 cluster in HEK293 cells, expression in MRC5 cells was relatively low (Figure 5.7). A high expression constitutive vector is perhaps a better choice to use in this study.

Relatively little is known about the regulation of pluripotency and differentiation of stem cells by the mir-302a cluster. Rosa et al., (2011) reported that NR2F2, an inhibitor of Oct4 was a target of miR-302a. They showed that both Oct4 and miR-

302a directly repress NR2F2, and regulate the maintenance of pluripotency and differentiation of ES cells. Notably, NR2F2 was a predicted target of miR-302a as well as miR-302a* by TargetScan Human (Table S5.2). SP3 is a possible inhibitor of Oct-4 and Nanog transcription factors (Pesce et al., 1999; Wu et al., 2006). Although SP3 is a predicted target of miR-302a* we were unable to confirm this.

Lin et al., (2011) attributed reprogramming by the miR-302a cluster to was global demethylation due to the repression of epigenetic regulators such as AOF1, AOF2, MECP1-p66 and MECP2. They found that expression of the miR-302 cluster repressed the above mentioned proteins, accompanied by the appearance of pluripotency associated proteins and global demethylation. Other targets of the miR-302 cluster that might be involved in the maintenance of pluripotency and differentiation include lefty1 and lefty2 (Rosa et al., 2009), CDKN1A; p21, Cyclin D1, BTG1, BTG2 and BTG3 (Wang et al., 2008; Card et al., 2008), RHOC and TGFb (Subramanyam et al., 2011), and PTEN (Sun et al., 1999; Lipchina et al., 2011).

Chapter 6 General Discussion

6.1 Star strands and isomiRs

The advent of deep sequencing has resulted in the discovery of large numbers of new miRNAs (Suh et al., 2004; Morin et al., 2008; Lipchina et al., 2011) with over 1600 entries for human miRNA genes in the most recent version of miRBase (Griffith-Jones., 2004). In parallel, the development of bioinformatics programs (Table 1.2) allows genome wide prediction of the mRNA targets of miRNAs (Lewis et al., 2005). Although the prediction programs are not entirely accurate (Ritchie et al., 2009), it is becoming increasingly feasible to test such predictions experimentally through the use of new technologies such as PAR-CLIP (Lipchina et al., 2011). These developments have opened up the ability to undertake genome scale investigation of miRNA regulation.

Star strands and isomiRs are emerging features of miRNAs that may in future become incorporated into the predictive and functional studies described above. Until recently the star strand has been considered to be a by-product that is degraded (Schwarz et al., 2003). However, it is now clear that miRNA* sequences are often highly expressed and associated with RISC proteins (Okamura et al., 2008). Indeed we found that miR-302a* and miR-9* were amongst the most highly expressed miRNAs in our database. Several papers have reported functional roles for miRNA star strands (Okamura et al., 2008; Yoo et al., 2011). The overall importance of star strands is strongly supported by the observation of arm-switching between cell types, where particular tissues switch to making more of the star strand than the guide miRNA (Ro et al., 2007; Ruby et al., 2007; de Wit et al., 2009; Chiang et al., 2009; Griffiths-Jones et al., 2011).

Arm-switching also occurs during evolution, where certain species switch to making the star strand of a miRNA (Griffith-Jones et al., 2011). These observations strongly indicate that the expression of some star strands have been selected during evolution, strongly indicating that they are functionally important.

I observed that about 5.8% of the miRNAs we sequenced were star strands rather than guide strands, which is comparable with other studies (Lagos-Quintana et al., 2002; Aravin et al., 2003; Lim et al., 2003; Bartel, 2004). However this is probably an underestimate of star strand importance because it does not take into account possible star strand production by different tissues or the evolutionary possibility that some guide strands were originally star strands.

6.2 IsomiRs and evolution

IsomiRs can also be highly expressed and are associated with the RISC complex (Cloonan et al., 2011; see Chapter 3). The publication by Fukunaga et al., (2012) is perhaps the only publication to experimentally establish the functional importance of two isomiRs of Drosphila, which show tissue specific expression. However, the identification of arm switching is an alternative approach that has contributed strongly to the evidence of the functional importance of star strands (see above). I suggest that the same approach can be adapted for isomiRs, essentially by looking for changes in isomiR production between closely related miRNA genes, tissue types or species. For example, Figure 6.1 illustrates that the dominant miRNA for hsa-miR-500a is AUGCACCUGGGCAAGGAUUCUG, whereas the dominant miRNA for hsa-miR-502 looks like a 5'/3' isomiR with the sequence AAUGCACCUGGGCAAGGAUUCA. These are the sequences for the two miRNA

genes that are listed in miRBase. Figure 6.1 illustrates that hsa-miR-500a and 502 are within the same cluster and encode very similar miRNAs, indicating that they most probably arose by duplication during evolution. It can be seen that both genes make many similar isomiRs but express different isomiRs at the highest level. The most straightforward explanation of Figure 6.1 is that all of the isomiRs are functional and that the two genes have evolved in order to express two isomiRs in particular, suggesting that both isomiRs are selectively advantageous and therefore biologically important. The example in Figure 6.1 was identified by using miRBase to look at the sequencing details of a list of 31 miRNA clusters (Yu et al., 2006). In addition I found two other clusters from the total of 31 that were screened that showed evidence of isomiR switching (Figure S6.1 and S6.2).

Similarly, I noted that the dominant isomiR of 9-1 that is expressed by megakaryoblasts has the sequence UCUUUGGUUAUCUAGCUGUAUGA, whereas the dominant isomiR expressed by a brain sample has the sequence UUGGUUAUCUAGCUGUAUGA (miRGator database, samples 1 and 5; Cho et al., 2012). It should be noted that I have not confirmed this sequencing data by an independent experimental approach. Nevertheless, the bioinformatics analyses I have done so far illustrate that this is a promising approach towards testing whether isomiRs are of biological and evolutionary importance.

hsa-miR-500a

hsa-miR-500a-5p	hsa-miR-500a-3p	Count	RPM
UAAUCCUUGCUACCUGGGUGAGAGU		41	1.97
UAAUCCUUGCUACCUGGGUGAGA		27	4.24
UAAUCCUUGCUACCUGGGUGAGAG		14	0.576
UAAUCCUUGCUACCUGGGUGAGAGUG		8	2.31
UAAUCCUUGCUACCUGGGUGAG		4	2.92
UAAUCCUUGCUACCUGGGUGAGAGUGC		2	0.0535
	AAUGCACCUGGGCAAGGAUU	86	7.61
	AAUGCACCUGGGCAAGGAUUC	72	13.5
	AAUGCACCUGGGCAAGGAUUCU	59	1.63
	AAUGCACCUGGGCAAGGAUUCUGA	27	0.76
	AAUGCACCUGGGCAAGGAUUCUG	18	0.42
	AAUGCACCUGGGCAAGGAU	13	0.283
	AUGCACCUGGGCAAGGAUUCU	824	22.5
	AUGCACCUGGGCAAGGAUUCUGA	608	15.4
	AUGCACCUGGGCAAGGAUUCUG	261	8.5
	AUGCACCUGGGCAAGGAUUC	194	4.67
	AUGCACCUGGGCAAGGAUU	190	5.09
	AUGCACCUGGGCAAGGAU	15	0.515
	AUGCACCUGGGCAAGGA	5	0.105

GCUCCCCCUCUCUAAUCCUUGCUACCUGGGUGAGAGUGCUGUCUGAAUGCA**AUGCACCUGGGCAAGGAUUCU**GAGAGCGAGAGC

hsa-miR-502

hsa-miR-502-5p	hsa-miR-502-3p	Count	RPM
UAAUCCUUGCUAUCUGGGUGC		2	0.0611
UAAUCCUUGCUAUCUGGGUGCUAGU		1	0.188
AAUCCUUGCUAUCUGGGUGCUAGUGC		9	0.159
AAUCCUUGCUAUCUGGGUGCUAGU		5	0.159
AAUCCUUGCUAUCUGGGUGCU		2	0.0676
AAUCCUUGCUAUCUGGGUGCUAGUG		2	0.0611
AAUCCUUGCUAUCUGGGUGCUA		2	0.0676
	AAUGCACCUGGGCAAGGAUUCA	592	16.3
	AAUGCACCUGGGCAAGGAUUCAG	555	14.7
	AAUGCACCUGGGCAAGGAUU	86	8.69
	AAUGCACCUGGGCAAGGAUUC	72	15.4
	AAUGCACCUGGGCAAGGAUUCAGA	22	0.567
	AAUGCACCUGGGCAAGGAU	13	0.324
	AUGCACCUGGGCAAGGAUUC	194	5.34
	AUGCACCUGGGCAAGGAUU	190	5.82
	AUGCACCUGGGCAAGGAUUCAGA	17	0.396
	AUGCACCUGGGCAAGGAUUCA	15	0.431
	AUGCACCUGGGCAAGGAU	15	0.588
	AUGCACCUGGGCAAGGA	5	0.12
	AUGCACCUGGGCAAGGAUUCAG	4	0.116

UGCUCCCCCUCUCUAAUCCUUGCUAUCUGGGUGCUAGUGCUGGCUCAAUGC**AAUGCACCUGGGCAAGGAUUCAG**AGAGGGGGAGCU

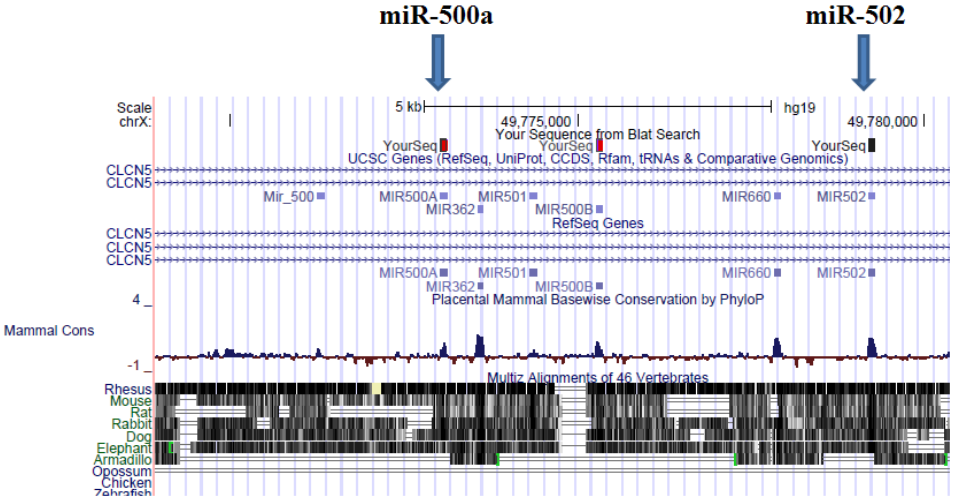


Figure 6.1
miR-500a and miR-502 are located in the same cluster and have mature sequence that are almost similar to each other with one or two bases different in the 5' and 3' ends. Deep sequencing results were taken from miRBase (August 2012, Griffith-Jones et al., 2004) and human genome map was taken from UCSC genome browser.

6.3 IsomiR expression

My northern blots findings were similar to the deep sequencing results of others in that the most dominant isomiRs varies between tissue types (Figure 3.3) (Fernandez-Valverde et al., 2010) and were able to associate with Argonaute proteins (Figure 3.4) (Burrough et al., 2011). My results that cross-referenced the predicted targets of isomiRs and canonical/ annotated miRNAs show only a small fraction of them have common targets (22%, see Figure 3.6 and Table S3.4). This reveals that large number of miRNA targets might have been missed as isomiRs were not included in the prediction of a miRNA, because only the annotated mature sequence is used in the prediction (TargetScan Human, Lewis et al., 2005).

My experimental studies did not establish whether the inhibition of DMNT3B or NCAM2 by isomiR-9 is of biological significance. They do however support the prediction that minor changes at the 5' end of miRNAs can have a major impact upon mRNA targeting. It remains to be seen whether natural selection is more likely to select conservative 5' isomiRs that target a similar set of mRNAs to the canonical miRNA, as suggested by Cloonan et al., 2011. Alternatively, tissue specific expression of isomiRs might allow a very different set of mRNAs to be targeted (Fukunaga et al., 2012).

In future, it might be interesting to observe the effect of isomiR-9 knockdown. As neural stem cells express high level of miR-9 and isomiR-9 (Table S3.1, Figure 3.5C), it is a good candidate to investigate the biological effect of isomiR-9 knockdown. This could be achieved by loss-of-function study using RNA sponge that we have developed (Figure 4.3). A lentiviral isomiR-9 sponge expression vector could be used

to introduce a sponge vector into NSCs or I could first establish stable transgene expression of a sponge in hESCs prior to neural differentiation. In addition, the effect of isomiR-9 knockdown could also be tested in a transgenic animal. Further studies are needed to test the function of miR-302a* in hESCs, where this miRNA is very abundant (Table S3.1).

6.4 NCAM2 and Prostate Cancer

NCAM2 is highly expressed in the brain and at low levels in various adult tissues including prostate, ovary, liver, kidney, pancreas and spleen (Paoloni-Giacobino et al., 1997). My finding of NCAM2 expression in androgen-dependent LNCaP but not in androgen-independent PC3 and DU145 prostate cancer cell lines (Figure 4.8) are consistent with a previous study (Takahashi et al., 2011). In future, it would be interesting to further investigate the interaction between isomiR-9 and NCAM2.

6.5 Conclusion

Overall my results support the bioinformatics prediction that single nucleotide changes at the 5' end of a miRNA are likely to generate functionally significant new targets, which was again supported by the RNA sponge experiment. Out of 17 miRNA target tests that I made based upon the bioinformatics predictions, 5 were incorrect, giving a 29.4% false positive or negative predictions, suggesting the target prediction is fairly reliable. This project supports the observation that isomiRs are functionally significant and more importantly may have very different targets to the canonical/ annotated miRNA. This suggests that the regulation of cellular processes by miRNA is far more complex than previously thought. This project also provides a good platform for future studies into the biological significant of isomiRs.

References

Agarwal S, Loh YH, McLoughlin EM, Huang J, Park IH, Miller JD, Huo H, Okuka M, Dos Reis RM, Loewer S, Ng HH, Keefe DL, Goldman FD, Klingelhutz AJ, Liu L, Daley GQ. Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. Nature. 2010 Mar 11;464(7286):292-6.

Alvarez-Buylla A, Seri B, Doetsch F. Identification of neural stem cells in the adult vertebrate brain. Brain Res Bull. 2002 Apr;57(6):751-8.

Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, Morrisey EE. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell. 2011 Apr 8;8(4):376-88.

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T. The small RNA profile during Drosophila melanogaster development. Dev Cell. 2003 Aug;5(2):337-50.

Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci U S A. 2008 Jun 10;105(23):7964-9.

Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 2008 Sep 4;455(7209):64-71.

Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, Kiledjian M. Differential regulation of microRNA stability. RNA. 2010 May;16(5):1032-9.

Bar M, Wyman SK, Fritz BR, Qi J, Garg KS, Parkin RK, Kroh EM, Bendoraite A, Mitchell PS, Nelson AM, Ruzzo WL, Ware C, Radich JP, Gentleman R, Ruohola-Baker H, Tewari M. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells. 2008 Oct;26(10):2496-505.

Barroso-delJesus A, Lucena-Aguilar G, Sanchez L, Ligero G, Gutierrez-Aranda I, Menendez P. The Nodal inhibitor Lefty is negatively modulated by the microRNA miR-302 in human embryonic stem cells. FASEB J. 2011 May;25(5):1497-508.

Barroso-delJesus A, Romero-López C, Lucena-Aguilar G, Melen GJ, Sanchez L, Ligero G, Berzal-Herranz A, Menendez P. Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. Mol Cell Biol. 2008 Nov;28(21):6609-19.

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004 Jan 23;116(2):281-97.

Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009 Jan 23;136(2):215-33.

Basyuk E, Suavet F, Doglio A, Bordonné R, Bertrand E. Human let-7 stem-loop precursors harbor features of RNase III cleavage products. Nucleic Acids Res. 2003 Nov 15;31(22):6593-7.

Beitzinger M, Peters L, Zhu JY, Kremmer E, Meister G. Identification of human microRNA targets from isolated Argonaute protein complexes. RNA Biol. 2007 Jun;4(2):76-84.

Bellin M, Marchetto MC, Gage FH, Mummery CL. Induced pluripotent stem cells: the new patient? Nat Rev Mol Cell Biol. 2012 Nov;13(11):713-26.

Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. Mol Cell. 2007 Oct 26;28(2):328-36.

Bieback K, Kern S, Klüter H, Eichler H. Critical parameters for the isolation of mesenchymal stem cells from umbilical cord blood. Stem Cells.2004;22(4):625-34.

Bonev B, Pisco A, Papalopulu N. MicroRNA-9 reveals regional diversity of neural progenitors along the anterior-posterior axis. Dev Cell. 2011 Jan 18;20(1):19-32.

Bonnamain V, Neveu I, Naveilhan P. Neural stem/progenitor cells as a promising candidate for regenerative therapy of the central nervous system. Front Cell Neurosci. 2012;6:17.

Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006 Dec;13(12):1097-101.

Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. PLoS Biol. 2005 Mar;3(3):e85.

Breving K, Esquela-Kerscher A. The complexities of microRNA regulation: mirandering around the rules. Int J Biochem Cell Biol. 2010 Aug;42(8):1316-29.

Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y, Daub CO. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res. 2010 Oct;20(10):1398-410.

Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Suzuki H, Hayashizaki Y, Daub CO. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol. 2011 Jan-Feb;8(1):158-77.

Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004 Dec;10(12):1957-66.

Camnasio S, Delli Carri A, Lombardo A, Grad I, Mariotti C, Castucci A, Rozell B, Riso PL, Castiglioni V, Zuccato C, Rochon C, Takashima Y, Diaferia G, Biunno I, Gellera C, Jaconi M, Smith A, Hovatta O, Naldini L, Di Donato S, Feki A, Cattaneo E.

The first reported generation of several induced pluripotent stem cell lines from homozygous and heterozygous Huntington's disease patients demonstrates mutation related enhanced lysosomal activity. Neurobiol Dis. 2012 Apr;46(1):41-51.

Cao F, Xie X, Gollan T, Zhao L, Narsinh K, Lee RJ, Wu JC. Comparison of gene-transfer efficiency in human embryonic stem cells. Mol Imaging Biol. 2010 Jan-Feb;12(1):15-24.

Card DA, Hebbar PB, Li L, Trotter KW, Komatsu Y, Mishina Y, Archer TK. Oct4/Sox2 regulated miR-302 targets cyclin D1 in human embryonic stem cells. Mol Cell Biol. 2008 Oct;28(20):6426-38.

Carvajal-Vergara X, Sevilla A, D'Souza SL, Ang YS, Schaniel C, Lee DF, Yang L, Kaplan AD, Adler ED, Rozov R, Ge Y, Cohen N, Edelmann LJ, Chang B, Waghray A, Su J, Pardo S, Lichtenbelt KD, Tartaglia M, Gelb BD, Lemischka IR. Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. Nature. 2010 Jun 10;465(7299):808-12.

Cenik ES, Zamore PD. Argonaute proteins. Curr Biol. 2011 Jun 21;21(12):R446-9.

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell. 2011 Oct 14;147(2):358-69.

Chan SP, Slack FJ. And now introducing mammalian mirtrons. Dev Cell. 2007 Nov;13(5):605-7.

Chang HT, Li SC, Ho MR, Pan HW, Ger LP, Hu LY, Yu SY, Li WH, Tsai KW. Comprehensive analysis of microRNAs in breast cancer. BMC Genomics. 2012;13 Suppl 7:S18.

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang DZ. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet. 2006 Feb;38(2):228-33.

Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chédin F. Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem. 2005 Aug 1;95(5):902-17.

Cheng WC, Chung IF, Huang TS, Chang ST, Sun HJ, Tsai CF, Liang ML, Wong TT, Wang HW. YM500: a small RNA sequencing (smRNA-seq) database for microRNA research. Nucleic Acids Res. 2013 Jan;41(Database issue):D285-94.

Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010 May 15;24(10):992-1009.

Cho S, Jang I, Jun Y, Yoon S, Ko M, Kwon Y, Choi I, Chang H, Ryu D, Lee B, Kim VN, Kim W, Lee S. MiRGator v3.0: a microRNA portal for deep sequencing,

expression profiling and mRNA targeting. Nucleic Acids Res. 2013 Jan;41(Database issue):D252-7.

Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E, Krishnan K, Gardiner B, Wang X, Nones K, Steen JA, Matigian NA, Wood DL, Kassahn KS, Waddell N, Shepherd J, Lee C, Ichikawa J, McKernan K, Bramlett K, Kuersten S, Grimmond SM. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. 2011 Dec 30;12(12):R126.

Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, Barad O, Bentwich Z, Szafranska AE, Labourier E, Raymond CK, Roberts BS, Juhl H, Kinzler KW, Vogelstein B, Velculescu VE. The colorectal microRNAome. Proc Natl Acad Sci U S A. 2006a Mar 7;103(10):3687-92.

Cummins JM, Velculescu VE. Implications of micro-RNA profiling for cancer diagnosis. Oncogene. 2006b Oct 9;25(46):6220-7.

de Wit E, Linsen SE, Cuppen E, Berezikov E. Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res. 2009 Nov;19(11):2064-74.

Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. Genes Dev. 2003 Feb 15;17(4):438-42.

Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G. microRNAs associated with the different human Argonaute proteins. Nucleic Acids Res. 2012 Oct;40(19):9850-62.

Easow G, Teleman AA, Cohen SM. Isolation of microRNA targets by miRNP immunopurification. RNA. 2007 Aug;13(8):1198-204.

Ebert AD, Yu J, Rose FF Jr, Mattis VB, Lorson CL, Thomson JA, Svendsen CN. Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature. 2009 Jan 15;457(7227):277-80.

Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat Methods. 2007 Sep;4(9):721-6.

Ebert MS, Sharp PA. Emerging roles for natural microRNA sponges. Curr Biol. 2010a Oct 12;20(19):R858-61.

Ebert MS, Sharp PA. MicroRNA sponges: progress and possibilities. RNA. 2010b Nov;16(11):2043-50.

Elkayam E, Kuhn CD, Tocilj A, Haase AD, Greene EM, Hannon GJ, Joshua-Tor L. The structure of human Argonaute-2 in complex with miR-20a. Cell. 2012 Jul 6;150(1):100-10.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2003;5(1):R1.

Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006 Apr;6(4):259-69.

Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science. 2005 Dec 16;310(5755):1817-21.

Fernandez-Valverde SL, Taft RJ, Mattick JS. Dynamic isomiR regulation in Drosophila development. RNA. 2010 Oct;16(10):1881-8.

Fire AZ. Gene silencing by double-stranded RNA (Nobel Lecture). Angew Chem Int Ed Engl. 2007;46(37):6966-84.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998 Feb 19;391(6669):806-11.

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet. 2007 Aug;39(8):1033-7.

Frank F, Sonenberg N, Nagar B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human Ago2. Nature. 2010 Jun 10;465(7299):818-22.

Friedenstein AJ, Chailakhjan RK, Lalykina KS. The development of fibroblast colonies in monolayer cultures of guinea-pig bone marrow and spleen cells. Cell Tissue Kinet. 1970 Oct;3(4):393-403.

Fukunaga R, Han BW, Hung JH, Xu J, Weng Z, Zamore PD. Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell. 2012 Oct 26;151(3):533-46.

Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinformatics. 2007 Mar 1;8:69.

Gangaraju VK, Lin H. MicroRNAs: key regulators of stem cells. Nat Rev Mol Cell Biol. 2009 Feb;10(2):116-25.

Garofalo M, Croce CM. microRNAs: Master regulators as potential therapeutics in cancer. Annu Rev Pharmacol Toxicol. 2011 Feb 10;51:25-43.

Gerrard L, Rodgers L, Cui W. Differentiation of human embryonic stem cells to neural lineages in adherent culture by blocking bone morphogenetic protein signaling. Stem Cells. 2005 Oct;23(9):1234-41.

Ghildiyal M, Xu J, Seitz H, Weng Z, Zamore PD. Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA. 2010 Jan;16(1):43-56.

Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. EMBO Rep. 2011 Feb;12(2):172-7.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008 Jan;36(Database issue):D154-8.

Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D109-11.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007 Jul 6;27(1):91-105.

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell. 2001 Jul 13;106(1):23-34.

Guillot PV, Gotherstrom C, Chan J, Kurata H, Fisk NM. Human first-trimester fetal MSC express pluripotency markers and grow faster and have longer telomeres than adult MSC. Stem Cells. 2007 Mar;25(3):646-54.

Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science. 1999 Oct 29;286(5441):950-2.

Hamlin JA, Fang H, Schwob JE. Differential expression of the mammalian homologue of fasciclin II during olfactory development in vivo and in vitro. J Comp Neurol. 2004 Jun 28;474(3):438-52.

Hammell CM, Lubin I, Boag PR, Blackwell TK, Ambros V. nhl-2 Modulates microRNA activity in Caenorhabditis elegans. Cell. 2009 Mar 6;136(5):926-38.

Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post transcriptional gene silencing in Drosophila cells. Nature. 2000 Mar 16;404(6775):293-6.

Han BW, Hung JH, Weng Z, Zamore PD, Ameres SL. The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. Curr Biol. 2011 Nov 22;21(22):1878-87.

Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006 Jun 2;125(5):887-901.

Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, Yang WY, Haussler D, Blelloch R, Kim VN. Posttranscriptional crossregulation between Drosha and DGCR8. Cell. 2009 Jan 9;136(1):75-84.

Haraguchi T, Ozaki Y, Iba H. Vectors expressing efficient RNA decoys achieve the long term suppression of specific microRNA activity in mammalian cells. Nucleic Acids Res. 2009 Apr;37(6):e43.

Heinrich EM, Dimmeler S. MicroRNAs and stem cells: control of pluripotency, reprogramming, and lineage commitment. Circ Res. 2012 Mar 30;110(7):1014-22.

Heo I, Ha M, Lim J, Yoon MJ, Park JE, Kwon SC, Chang H, Kim VN. Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell. 2012 Oct 26;151(3):521-32.

Höck J, Meister G. The Argonaute protein family. Genome Biol. 2008;9(2):210.

Hook L, Vives J, Fulton N, Leveridge M, Lingard S, Bootman MD, Falk A, Pollard SM, Allsopp TE, Dalma-Weiszhausz D, Tsukamoto A, Uchida N, Gorba T. Non-immortalized human neural stem (NS) cells as a scalable platform for cellular assays. Neurochem Int. 2011 Sep;59(3):432-44.

Hu S, Wilson KD, Ghosh Z, Han L, Wang Y, Lan F, Ransohoff KJ, Burridge P, Wu JC. MicroRNA-302 Increases Reprogramming Efficiency via Repression of NR2F2. Stem Cells. 2013 Feb;31(2):259-68.

Huang HP, Chen PH, Hwu WL, Chuang CY, Chien YH, Stone L, Chien CL, Li LT, Chiang SC, Chen HF, Ho HN, Chen CH, Kuo HC. Human Pompe disease-induced pluripotent stem cells for pathogenesis modeling, drug testing and disease marker identification. Hum Mol Genet. 2011 Dec 15;20(24):4851-64.

Humphreys DT, Hynes CJ, Patel HR, Wei GH, Cannon L, Fatkin D, Suter CM, Clancy JL, Preiss T. Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. PLoS One. 2012;7(2):e30933.

Hutvagner G. Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. FEBS Lett. 2005 Oct 31;579(26):5850-7.

In 't Anker PS, Noort WA, Scherjon SA, Kleijburg-van der Keur C, Kruisselbrink AB, van Bezooijen RL, Beekhuizen W, Willemze R, Kanhai HH, Fibbe WE. Mesenchymal stem cells in human second-trimester bone marrow, liver, lung, and spleen exhibit a similar immunophenotype but a heterogeneous multilineage differentiation potential. Haematologica. 2003a Aug;88(8):845-52.

In 't Anker PS, Scherjon SA, Kleijburg-van der Keur C, Noort WA, Claas FH, Willemze R, Fibbe WE, Kanhai HH. Amniotic fluid as a novel source of mesenchymal stem cells for therapeutic transplantation. Blood. 2003b Aug 15;102(4):1548-9.

Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, Hefferan MP, Van Gorp S, Nazor KL, Boscolo FS, Carson CT, Laurent LC, Marsala M, Gage FH, Remes AM, Koo EH, Goldstein LS. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature. 2012 Jan 25;482(7384):216-20.

Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H, Boulos M, Gepstein L. Modelling the long QT syndrome with induced pluripotent stem cells. Nature. 2011 Mar 10;471(7337):225-9.

Jagadeeswaran G, Zheng Y, Sumathipala N, Jiang HB, Arrese EL, et al. Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. BMC Genomics 2010;11: 52.

Jayawardena TM, Egemnazarov B, Finch EA, Zhang L, Payne JA, Pandya K, Zhang Z, Rosenberg P, Mirotsou M, Dzau VJ. MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes. Circ Res. 2012 May 25;110(11):1465-73.

Jeyaseelan K, Lim KY, Armugam A. MicroRNA expression in the blood and brain of rats subjected to transient focal ischemia by middle cerebral artery occlusion. Stroke 2008; 39:959-66

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS Biol. 2004 Nov;2(11):e363.

Judson RL, Babiarz JE, Venere M, Blelloch R. Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat Biotechnol 2009;27(5):459-61.

Kapsimali M, Kloosterman WP, de Bruijn E, Rosa F, Plasterk RH, Wilson SW. MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. Genome Biol. 2007;8(8):R173.

Karreth FA, Tay Y, Perna D, Ala U, Tan SM, Rust AG, DeNicola G, Webster KA, Weiss D, Perez-Mancera PA, Krauthammer M, Halaban R, Provero P, Adams DJ, Tuveson DA, Pandolfi PP. In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell. 2011 Oct 14;147(2):382-95.

Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep. 2007 Aug;8(8):763-9.

Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JA, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Ørom UA, Lund AH, Perrakis A, Raz E, Agami R. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell. 2007 Dec 28;131(7):1273-86.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007 Oct;39(10):1278-84.

Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. 2001 Oct 15;15(20):2654-9.

Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009 Feb;10(2):126-39.

Kloosterman WP, Plasterk RH. The diverse functions of microRNAs in animal development and disease. Dev Cell. 2006a Oct;11(4):441-50.

Kloosterman WP, Steiner FA, Berezikov E, de Bruijn E, van de Belt J, Verheul M, Cuppen E, Plasterk RH. Cloning and expression of new microRNAs from zebrafish. Nucleic Acids Res. 2006b May 12;34(9):2558-69.

Kloosterman WP, Wienholds E, Ketting RF, Plasterk RH. Substrate requirements for let-7 function in the developing zebrafish embryo. Nucleic Acids Res. 2004 Dec 7;32(21):6284-91.

Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011 Jan;39(Database issue):D152-7.

Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. Nat Genet. 2005 May;37(5):495-500.

Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS. A microRNA array reveals extensive regulation of microRNAs during brain development. RNA. 2003 Oct;9(10):1274-81.

Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. Silencing of microRNAs in vivo with 'antAgomirs'. Nature. 2005 Dec 1;438(7068):685-9.

Kuchenbauer F, Morin RD, Argiropoulos B, Petriv OI, Griffith M, Heuser M, Yung E, Piper J, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Hansen CL, Marra MA, Humphries RK. In-depth characterization of the microRNA transcriptome in a leukemia progression model. Genome Res. 2008 Nov;18(11):1787-97.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science. 2001 Oct 26;294(5543):853-8.

Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. Curr Biol. 2002 Apr 30;12(9):735-9.

Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science. 2001 Oct 26;294(5543):858-62.

Le MT, Xie H, Zhou B, Chia PH, Rizk P, Um M, Udolph G, Yang H, Lim B, Lodish HF. MicroRNA-125b promotes neuronal differentiation in human cells by repressing multiple targets. Mol Cell Biol. 2009 Oct;29(19):5290-305.

Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, Ganat YM, Menon J, Shimizu F, Viale A, Tabar V, Sadelain M, Studer L. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. Nature. 2009 Sep 17;461(7262):402-6.

Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, Wang K. Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA. 2010 Nov;16(11):2170-80.

Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science. 2001 Oct 26;294(5543):862-4.

Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993 Dec 3;75(5):843-54.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003 Sep 25;425(6956):415-9.

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004 Oct 13;23(20):4051-60.

Lehrbach NJ, Miska EA. Regulation of pre-miRNA processing. Adv Exp Med Biol. 2010;700:67-75.

Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005 Jan 14;120(1):15-20.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell. 2003 Dec 26;115(7):787-98.

Li SC, Liao YL, Ho MR, Tsai KW, Lai CH, Lin WC. miRNA arm selection and isomiR distribution in gastric cancer. BMC Genomics. 2012;13 Suppl 1:S13.

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 MA, Pei D. MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. J Biol Chem. 2011 May 13;286(19):17359-64.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005 Feb 17;433(7027):769-73.

Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. The microRNAs of Caenorhabditis elegans. Genes Dev. 2003 Apr 15;17(8):991-1008.

Lin SL, Chang DC, Chang-Lin S, Lin CH, Wu DT, Chen DT, Ying SY. Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. RNA 2008;14(10):2115-24.

Lin SL, Chang DC, Lin CH, Ying SY, Leu D, Wu DT. Regulation of somatic cell reprogramming through inducible mir-302 expression. Nucleic Acids Res 2011;39(3):1054-65.

Lipchina I, Elkabetz Y, Hafner M, Sheridan R, Mihailovic A, Tuschl T, Sander C, Studer L, Betel D. Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. Genes Dev. 2011 Oct 15;25(20):2173-86.

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua Tor L, Hannon GJ. Argonaute2 is the catalytic engine of mammalian RNAi. Science 2004;305, 1437-1441.

Liu N, Abe M, Sabin LR, Hendriks GJ, Naqvi AS, Yu Z, Cherry S, Bonini NM. The exoribonuclease Nibbler controls 3' end processing of microRNAs in Drosophila. Curr Biol. 2011 Nov 22;21(22):1888-93.

Llorens F, Bañez-Coronel M, Pantano L, Del Río JA, Ferrer I, Estivill X, Martí E. A highly expressed miR-101 isomiR is a functional silencing small RNA. BMC Genomics. 2013 Feb 15;14(1):104.

Lukiw WJ. Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus. Neuroreport 2007; 18:297-300.

Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc Natl Acad Sci U S A. 2007 Jun 5;104(23):9667-72.

Ma L, Young J, Prabhala H, Pan E, Mestdagh P, Muth D, Teruya-Feldstein J, Reinhardt F, Onder TT, Valastyan S, Westermann F, Speleman F, Vandesompele J, Weinberg RA. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nat Cell Biol. 2010 Mar;12(3):247-56.

Majoros WH, Ohler U. Spatial preferences of microRNA targets in 3' untranslated regions. BMC Genomics. 2007 Jun 7;8:152.

Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Vergoulis T, Koziris N, Sellis T, Tsanakas P, Hatzigeorgiou AG. DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W273-6.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, Sharp PA, Bartel DP, Jaenisch R, Young RA.

Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008 Aug 8;134(3):521-33.

Martin EC, Elliott S, Rhodes LV, Antoon JW, Fewell C, Zhu Y, Driver JL, Jodari-Karimi M, Taylor CW, Flemington EK, Beckman BS, Collins-Burow BM, Burow ME. Preferential star strand biogenesis of pre-miR-24-2 targets PKC-alpha and suppresses cell survival in MCF-7 breast cancer cells. Mol Carcinog. 2012 Aug 21.doi: 10.1002/mc.21946. [Epub ahead of print] PubMed PMID: 22911661.

Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A. 1981 Dec;78(12):7634-8.

Martinez NJ, Gregory RI. MicroRNA gene regulatory pathways in the establishment and maintenance of ESC identity. Cell Stem Cell. 2010 Jul 2;7(1):31-5.

Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros VR, Walhout AJ. A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity. Genes Dev. 2008 Sep 15;22(18):2535-49.

Meister G, Landthaler M, Dorsett Y, Tuschl T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA 2004;10: 544–550.

Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature. 2010 Feb 4;463(7281):621-6.

Miska EA, Alvarez-Saavedra E, Abbott AL, Lau NC, Hellman AB, McGonagle SM, Bartel DP, Ambros VR, Horvitz HR. Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS Genet. 2007 Dec;3(12):e215.

Miyoshi N, Ishii H, Nagano H, et al. Reprogramming of mouse and human cells to pluripotency using mature microRNAs. Cell Stem Cell 2011;8(6):633-8.

Moreno-Manzano V, Rodríguez-Jiménez FJ, García-Roselló M, Laínez S, Erceg S, Calvo MT, Ronaghi M, Lloret M, Planells-Cases R, Sánchez-Puelles JM, Stojkovic M. Activated spinal cord ependymal stem cells rescue neurological function. Stem Cells. 2009 Mar;27(3):733-43.

Moretti A, Bellin M, Welling A, Jung CB, Lam JT, Bott-Flügel L, Dorn T, Goedel A, Höhnke C, Hofmann F, Seyfarth M, Sinnecker D, Schömig A, Laugwitz KL. Patient-specific induced pluripotent stem-cell models for long-QT syndrome. N Engl J Med. 2010 Oct 7;363(15):1397-409.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008 Apr;18(4):610-21.

Murchison EP, Hannon GJ. miRNAs on the move: miRNA biogenesis and the RNAi machinery. Curr Opin Cell Biol. 2004 Jun;16(3):223-9.

Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P. Molecular basis for interaction of let-7 microRNAs with Lin28. Cell. 2011 Nov 23;147(5):1080-91.

Neilsen CT, Goodall GJ, Bracken CP. IsomiRs--the overlooked repertoire in the dynamic microRNAome. Trends Genet. 2012 Nov;28(11):544-9.

Neumüller RA, Betschinger J, Fischer A, Bushati N, Poernbacher I, Mechtler K, Cohen SM, Knoblich JA. Mei-P26 regulates microRNAs and cell growth in the Drosophila ovarian stem cell lineage. Nature. 2008 Jul 10;454(7201):241-5.

Newman MA, Mani V, Hammond SM. Deep sequencing of microRNA precursors reveals extensive 3' end modification. RNA. 2011 Oct;17(10):1795-803.

Odorico JS, Kaufman DS, Thomson JA. Multilineage differentiation from human embryonic stem cell lines. Stem Cells. 2001;19(3):193-204.

Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. Nat Struct Mol Biol. 2008 Apr;15(4):354-63.

Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.

Otaegi G, Pollock A, Sun T. An Optimized Sponge for microRNA miR-9 Affects Spinal Motor Neuron Development in vivo. Front Neurosci. 2011;5:146.

Pantano L, Estivill X, Martí E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res. 2010 Mar;38(5):e34.

Paoloni-Giacobino A, Chen H, Antonarakis SE. Cloning of a novel human neural cell adhesion molecule gene (NCAM2) that maps to chromosome region 21q21 and is potentially involved in Down syndrome. Genomics. 1997 Jul 1;43(1):43-51.

Park CY, Choi YS, McManus MT. Analysis of microRNA knockouts in mice. Hum Mol Genet. 2010 Oct 15;19(R2):R169-75.

Parolini O, Alviano F, Bagnara GP, Bilic G, Bühring HJ, Evangelista M, Hennerbichler S, Liu B, Magatti M, Mao N, Miki T, Marongiu F, Nakajima H, Nikaido T, Portmann-Lanz CB, Sankar V, Soncini M, Stadler G, Surbek D, Takahashi TA, Redl H, Sakuragawa N, Wolbank S, Zeisberger S, Zisch A, Strom SC. Concise review: isolation and characterization of cells from human term placenta: outcome of the first international Workshop on Placenta Derived Stem Cells. Stem Cells. 2008 Feb;26(2):300-11.

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P, Spring J, Srinivasan A, Fishman M,

Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature. 2000 Nov 2;408(6808):86-9.

Pasquinelli AE. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet. 2012 Mar 13;13(4):271-82.

Pesce M, Marin Gomez M, Philipsen S, Schöler HR. Binding of Sp1 and Sp3 transcription factors to the Oct-4 gene promoter. Cell Mol Biol (Noisy-le-grand). 1999 Jul;45(5):709-16.

Pietrzykowski AZ, Friesen RM, Martin GE, Puig SI, Nowak CL, Wynne PM, Siegelmann HT, Treistman SN. Posttranscriptional regulation of BK channel splice variant stability by miR-9 underlies neuroadaptation to alcohol. Neuron. 2008 Jul 31;59(2):274-87.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010 Jun 24;465(7301):1033-8.

Qu B, Han X, Tang Y, Shen N. A novel vector-based method for exclusive overexpression of star-form microRNAs. PLoS One. 2012;7(7):e41504.

Rajewsky N, Socci ND. Computational identification of microRNA targets. Dev Biol. 2004 Mar 15;267(2):529-35.

Rajewsky N. microRNA target predictions in animals. Nat Genet. 2006 Jun;38 Suppl:S8-13.

Rao PK, Kumar RM, Farkhondeh M, Baskerville S, Lodish HF. Myogenic factors that regulate expression of muscle-specific microRNAs. Proc Natl Acad Sci U S A. 2006 Jun 6;103(23):8721-6.

Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA, 10:1507-1517, 2004.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000 Feb 24;403(6772):901-6.

Reynolds BA, Tetzlaff W, Weiss S. A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes. J Neurosci. 1992 Nov;12(11):4565-74.

Ritchie W, Flamant S, Rasko JE. Predicting microRNA targets and functions: traps for the unwary. Nat Methods. 2009 Jun;6(6):397-8.

Ro S, Park C, Young D, Sanders KM, Yan W. Tissue-dependent paired expression of miRNAs. Nucleic Acids Res. 2007;35(17):5944-53.

Robinton DA, Daley GQ. The promise of induced pluripotent stem cells in research and therapy. Nature. 2012 Jan 18;481(7381):295-305.

Ronaghi M, Erceg S, Moreno-Manzano V, Stojkovic M. Challenges of stem cell therapy for spinal cord injury: human embryonic stem cells, endogenous neural stem cells, or induced pluripotent stem cells? Stem Cells. 2010 Jan;28(1):93-9.

Rosa A, Brivanlou AH. A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation. EMBO J. 2011 Jan 19;30(2):237-48.

Rosa A, Spagnoli FM, Brivanlou AH. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. Dev Cell. 2009 Apr;16(4):517-27.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. 2007 Dec;17(12):1850-64.

Saetrom P, Heale BS, Snøve O Jr, Aagaard L, Alluin J, Rossi JJ. Distance constraints between microRNA target sites dictate efficacy and cooperativity. Nucleic Acids Res. 2007;35(7):2333-42.

Sandhu R, Rivenbark AG, Coleman WB. Loss of post-transcriptional regulation of DNMT3b by microRNAs: a possible molecular mechanism for the hypermethylation defect observed in a subset of breast cancer cell lines. Int J Oncol. 2012 Aug;41(2):721-32.

Schirle NT, MacRae IJ. The crystal structure of human Argonaute2. Science. 2012 May 25;336(6084):1037-40.

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell. 2003 Oct 17;115(2):199-208.

Shi XB, Tepper CG, deVere White RW. Cancerous miRNAs and their regulation. Cell Cycle. 2008 Jun 1;7(11):1529-38.

Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP. Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell. 2010 Jun 25;38(6):789-802.

Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, Hargus G, Blak A, Cooper O, Mitalipova M, Isacson O, Jaenisch R. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell. 2009 Mar 6;136(5):964-77.

Stark A, Brennecke J, Russell RB, Cohen SM. Identification of Drosophila MicroRNA targets. PLoS Biol. 2003 Dec;1(3):E60.

Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. Genome Res. 2007 Dec;17(12):1865-79.

Subramanyam D, Lamouille S, Judson RL, Liu JY, Bucay N, Derynck R, Blelloch R. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. Nat Biotechnol 2011;29(5):443-8.

Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, Kim VN, Kim KS. Human embryonic stem cells express a unique set of microRNAs. Dev Biol. 2004 Jun 15;270(2):488-98.

Sumazin P, Yang X, Chiu HS, Chung WJ, Iyer A, Llobet-Navas D, Rajbhandari P, Bansal M, Guarnieri P, Silva J, Califano A. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell. 2011 Oct 14;147(2):370-81.

Sun H, Lesche R, Li DM, Liliental J, Zhang H, Gao J, Gavrilova N, Mueller B, Liu X, Wu H. PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. Proc Natl Acad Sci U S A. 1999 May 25;96(11):6199-204.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126(4):663-76.

Takahashi S, Kato K, Nakamura K, Nakano R, Kubota K, Hamada H. Neural cell adhesion molecule 2 as a target molecule for prostate and breast cancer gene therapy. Cancer Sci. 2011 Apr;102(4):808-14.

Tan GC, Dibb NJ. MicroRNA-Induced Pluripotent Stem Cells. Malaysian J Pathol 2012;34(2):167-8.

Thadani R, Tammi MT. MicroTar: predicting microRNA targets from RNA duplexes. BMC Bioinformatics. 2006 Dec 18;7 Suppl 5:S20.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. Science. 1998 Nov 6;282(5391):1145-7.

Tong L, Lin L, Wu S, Guo Z, Wang T, Qin Y, Wang R, Zhong X, Wu X, Wang Y, Luan T, Wang Q, Li Y, Chen X, Zhang F, Zhao W, Zhong Z. MiR-10a* up-regulates coxsackievirus B3 biosynthesis by targeting the 3D-coding sequence. Nucleic Acids Res. 2013 Feb 6. [Epub ahead of print] PubMed PMID: 23389951.

Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA. Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev. 1999 Dec 15;13(24):3191-7.

Uccelli A, Moretta L, Pistoia V. Mesenchymal stem cells in health and disease. Nat Rev Immunol. 2008 Sep;8(9):726-36.

van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN. Control of stress dependent cardiac growth and gene expression by a microRNA. Science. 2007 Apr 27;316(5824):575-9.

Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. Genes Dev. 2004 Jan 15;18(2):132-7.

Vescovi AL, Reynolds BA, Fraser DD, Weiss S. bFGF regulates the proliferative fate of unipotent (neuronal) and bipotent (neuronal/astroglial) EGF-generated CNS progenitor cells. Neuron. 1993 Nov;11(5):951-66.

Voellenkle C, Rooij Jv, Guffanti A, Brini E, Fasanaro P, Isaia E, Croft L, David M, Capogrossi MC, Moles A, Felsani A, Martelli F. Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs. RNA. 2012 Mar;18(3):472-84.

Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nat Genet. 2008 Dec;40(12):1478-83.

Watahiki A, Wang Y, Morris J, Dennis K, O'Dwyer HM, Gleave M, Gout PW, Wang Y. MicroRNAs associated with metastatic prostate cancer. PLoS One. 2011;6(9):e24950.

Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell. 1993 Dec 3;75(5):855-62.

Winther M, Berezin V, Walmod PS. NCAM2/OCAM/RNCAM: cell adhesion molecule with a role in neuronal compartmentalization. Int J Biochem Cell Biol. 2012 Mar;44(3):441-6.

Wu DY, Yao Z. Functional analysis of two Sp1/Sp3 binding sites in murine Nanog gene promoter. Cell Res. 2006 Mar;16(3):319-22.

Wu H, Ye C, Ramirez D, Manjunath N. Alternative processing of primary microRNA transcripts by Drosha generates 5’ end variation of mature microRNA. Plos One 2009;4:e7566

Wu SM, Hochedlinger K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat Cell Biol. 2011 May;13(5):497-505.

Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, Krouse MA, Webster PJ, Tewari M. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Res. 2011 Sep;21(9):1450-61.

Yang JS, Phillips MD, Betel D, Mu P, Ventura A, Siepel AC, Chen KC, Lai EC. Widespread regulatory activity of vertebrate microRNA* species. RNA. 2011 Feb;17(2):312-26.

Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat Struct Mol Biol. 2006 Jan;13(1):13-21.

Yazawa M, Hsueh B, Jia X, Pasca AM, Bernstein JA, Hallmayer J, Dolmetsch RE. Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. Nature. 2011 Mar 10;471(7337):230-4.

Yeom KH, Lee Y, Han J, Suh MR, Kim VN. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res. 2006;34(16):4622-9.

Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003 Dec 15;17(24):3011-6.

Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, Lee-Messer C, Dolmetsch RE, Tsien RW, Crabtree GR. MicroRNA-mediated conversion of human fibroblasts to neurons. Nature. 2011 Jul 13;476(7359):228-31.

Yu J, Wang F, Yang GH, Wang FL, Ma YN, Du ZW, Zhang JW. Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. Biochem Biophys Res Commun. 2006 Oct 13;349(1):59-68.

Yuva-Aydemir Y, Simkin A, Gascon E, Gao FB. MicroRNA-9: functional evolution of a conserved small regulatory RNA. RNA Biol. 2011 Jul-Aug;8(4):557-64.

Appendix

	hESC		NSC		MSC	
hsa-miR-	Sequence	SN	Sequence	SN	Sequence	SN
7			TGGAAGACTAGTGATTTTGTTGTT	112		
			TGGAAGACTAGTGATTTTGTTGT	31		
9			TCTTTGGTTATCTAGCTGTATGA	587		
			CTTTGGTTATCTAGCTGTATGA	116		
			TCTTTGGTTATCTAGCTGTATG	49		
			TCTTTGGTTATCTAGCTGTATGAA	37		
			CTTTGGTTATCTAGCTGTATGAA	15		
			TCTTTGGTTATCTAGCTGTAT	13		
			TCTTTGGTTATCTAGC	13		
			TTTGGTTATCTAGCTGTATGA	10		
9*			ATAAAGCTAGATAACCGAAAGT	643		
			TAAAGCTAGATAACCGAAAGTA	216		
			ATAAAGCTAGATAACCGAAAGTA	145		
			TAAAGCTAGATAACCGAAAGT	71		
			TAAAGCTAGATAACCGAAAGTAT	66		
			TAAAGCTAGATAACCGAAAGTAA	53		
			ATAAAGCTAGATAACCGAAAGTAT	28		
			ATAAAGCTAGATAACCGAAAG	19		
			ATAAAGCTAGATAACCGAAAGTAA	14		
			ATAAAGCTAGATAACCGAAA	10		
10a			TACCCTGTAGATCCGAATTTGT	14	TACCCTGTAGATCCGAATTTGT	13
			TACCCTGTAGATCCGAATTTGTG		ACCCTGTAGATCCGAATTTGTG	8
					ACCCTGTAGATCCGAATTTGT	4
					TACCCTGTAGATCCGAAT	3
					TACCCTGTAGATCCGAATTTGTG	3
					ACCCTGTAGATCCGAATTTG	1
					ACCCTGTAGATCCGAATT	1
					TACCCTGTAGATCCGAATT	1
					CCCTGTAGATCCGAATTTGTG	1
10a*					CAAATTCGTATCTAGGGGAAT	2
					CAAATTCGTATCTAGGGGAATA	
15a			TAGCAGCACATAATGGTTTGTG	12	TAGCAGCACATAATGGTTTGTG	24
			TAGCAGCACATAATGGTTTGT	12	TAGCAGCACATAATGGTTTGT	4
15b					TAGCAGCACATCATGGTTTACA	101
					TAGCAGCACATCATGGTTTAC	31
					TAGCAGCACATCATGGTTTACAT	4
					TAGCAGCACATCATGGTT	2
					TAGCAGCACATCATGGTTTA	2
					TAGCAGCACATCATGGT	1
16-2*					ACCAATATTACTGTGCTGCTTTA	4
					CCAATATTACTGTGCTGCTTTA	2
					ACCAATATTACTGTGCTGCTTT	1
					CCAATATTACTGTGCTGCTTT	1
					ACCAATATTACTGTGCTG	1
16	TAGCAGCACGTAAATATTGGCG	19	TAGCAGCACGTAAATATTGGCG	107	TAGCAGCACGTAAATATTGGCG	47
	TAGCAGCACGTAAATATTGGCGT	11	TAGCAGCACGTAAATATTGGCGT	43	TAGCAGCACGTAAATATTGGCGT	32
					TAGCAGCACGTAAATATTGGC	25
					TAGCAGCACGT	3
					TAGCAGCACGTAAATATTG	1
					TAGCAGCACGTAAATATT	1
					TAGCAGCACGTAAA	1
					TAGCAGCACGTA	1
					AGCAGCACGTAAATATTGGCGT	1
17	CAAAGTGCTTACAGTGCAGGTAG	39	CAAAGTGCTTACAGTGCAGGTAG	41	CAAAGTGCTTACAGTGCAGGTAG	111
	CAAAGTGCTTACAGTGCAGGT	19			CAAAGTGCTTACAGTGCAGGTA	8
					CAAAGTGCTTACAGTGCAGGTAGT	3
					TCAAAGTGCTTACAGTGCAGGT	3
					CAAAGTGCTTACAGTGCAGGT	1
					TCAAAGTGCTTACAGTGCAGGTA	1
18a					TAAGGTGCATCTAGTGCAGATAG	11
					TAAGGTGCATCTAGTGCAGAT	4
19b	TGTGCAAATCCATGCAAAACTGA	10	TGTGCAAATCCATGCAAAACTGA	11	TGTGCAAATCCATGCAAAACTGA	14
					TGTGCAAATCCATGCAAAACTG	1
					GTGCAAATCCATGCAAAACTGA	1
20a			TAAAGTGCTTATAGTGCAGGTAG	10	TAAAGTGCTTATAGTGCAGGTAG	2
					TAAAGTGCTTATAGTGCAGG	1
20a*					ACTGCATTATGAGCACTTAAAGT	1
					ACTGCATTATGAGCACTTAAAG	
20b	CAAAGTGCTCATAGTGCAGGTAG	33				
21	TAGCTTATCAGACTGATGTTGAC	324	TAGCTTATCAGACTGATGTTGAC	2452	TAGCTTATCAGACTGATGTTGAC	496
	TAGCTTATCAGACTGATGTTGA	80	TAGCTTATCAGACTGATGTTGA	462	TAGCTTATCAGACTGATGTTGA	337
	TAGCTTATCAGACTGATGTTGACA	16	TAGCTTATCAGACTGATGTTGACA	112	TAGCTTATCAGACTGATGTTG	22
	TAGCTTATCAGACTGATGTTGAA	10	TAGCTTATCAGACTGATGTTGACT	49	TAGCTTATCAGACTGATGT	9
			AGCTTATCAGACTGATGTTGAC	21	TAGCTTATCAGACTGATGTTGACT	7
			TAGCTTATCAGACTGATGTTGAA	20	TAGCTTATCAGACTGATGTT	4
			TTATCAGACTGATGTTGACTAGC	11	TAGCTTATCAGAC	1
					TAGCTTATCAGA	1

						AGCTTATCAGACTGATGTTGA	1
						TAGCTTATCAGACTGATG	1
						TAGCTTATCAGACTGA	1
21*						CAACACCAGTCGATGGGCTGTC	3
						CAACACCAGTCGATGGGCTGT	
22						AAGCTGCCAGTTGAAGAACTGT	400
						AAGCTGCCAGTTGAAGAACTG	16
						AGCTGCCAGTTGAAGAACTGT	5
						AAGCTGCCAGTTGAAGAACT	3
						AAGCTGCCAGTTGAAGAAC	2
						AAGCTGCCAGTTGAAGAA	1
						AAAGCTGCCAGTTGAAGAACTGT	1
						AAGCTGCCAGTTGAAGAACTGTT	1
22*						AGTTCTTCAGTGGCAAGCTTTA	4
						AGTTCTTCAGTGGCAAGCTTT	2
23a			ATCACATTGCCAGGGATTTCCA	21		ATCACATTGCCAGGGATTTCCA	712
			ATCACATTGCCAGGGATTTCC	13		ATCACATTGCCAGGGATTTCC	601
			ATCACATTGCCAGGGATTTCCAA	12		ATCACATTGCCAGGGATTTCCAA	278
						TCACATTGCCAGGGATTTCCAAC	105
						ATCACATTGCCAGGGATTTC	88
						TCACATTGCCAGGGATTTCCA	64
						TCACATTGCCAGGGATTTCC	33
						ATCACATTGCCAGGGATTTCCAAC	31
						TCACATTGCCAGGGATTTCCAA	28
						TCACATTGCCAGGGATTTC	12
						ATCACATTGCCAGGGATTT	7
						AATCACATTGCCAGGGATTTCCA	6
						CAAATCACATTGCCAGGGATTTC	1
						CACATTGCCAGGGATTTCCAA	1
23b			ATCACATTGCCAGGGATTACCACT	13		ATCACATTGCCAGGGATTACCAC	50
			ATCACATTGCCAGGGATTACC			TCACATTGCCAGGGATTACCAC	10
						ATCACATTGCCAGGGATTACCA	6
						ATCACATTGCCAGGGATTACC	4
						ATCACATTGCCAGGGATTAC	3
						AATCACATTGCCAGGGATTACCAC	1
						TCACATTGCCAGGGATTAC	1
23b*						TGGGTTCCTGGCATGCTGATTT	1
24	TGGCTCAGTTCAGCAGGAACAG	11	TGGCTCAGTTCAGCAGGAACAGT	187		TGGCTCAGTTCAGCAGGAACAG	86
			TGGCTCAGTTCAGCAGGAACAG	66		TGGCTCAGTTCAGCAGGAAC	24
			TGGCTCAGTTCAGCAGGAAC	10		TGGCTCAGTTCAGCAGGAACA	5
						TGGCTCAGTTCAGCAGGA	2
						TGGCTCAGTTCAGCAGGAA	1
25	CATTGCACTTGTCTCGGTCTGA	98	CATTGCACTTGTCTCGGTCTGA	139		CATTGCACTTGTCTCGGTCTGA	2
						CATTGCACTTGTC	1
25*	AGGCGGAGACTTGGGCAATTGCT	23	AGGCGGAGACTTGGGCAATTGCT	14			
	AGGCGGAGACTTGGGCAATTGC	10	AGGCGGAGACTTGGGCAATTGC	10			
	AGGCGGAGACUUGGGCAAUUG		AGGCGGAGACUUGGGCAAUUG				
26a	TTCAAGTAATCCAGGATAGGCT	38	TTCAAGTAATCCAGGATAGGCT	289		TTCAAGTAATCCAGGATAGGCT	619
			TTCAAGTAATCCAGGATAGGCTAT	29		TTCAAGTAATCCAGGATAGGC	94
			TTCAAGTAATCCAGGATAGGCTA	26		TTCAAGTAATCCAGGATAGG	14
			TTCAAGTAATCCAGGATAGGCTT	23		TTCAAGTAATCCAGGATAG	5
			TTCAAGTAATCCAGGATAGGC	14		TTCAAGTAATCCAGGATA	3
						TCAAGTAATCCAGGATAGGCT	2
						ATTCAAGTAATCCAGGATAGGCT	1
						TTCAAGTAATCCAGGAT	1
						TTCAAGTAATCCAGGA	1
26b			TTCAAGTAATTCAGGATAGGTT	28		TTCAAGTAATTCAGGATAGGTT	3
			TTCAAGTAATTCAGGATAGGT			TTCAAGTAATTCAGGATAG	2
						TTCAAGTAATTCAGGATAGG	1
						TTCAAGTAATTCAGGATAGGT	
27a			TTCACAGTGGCTAAGTTCCG	25		TTCACAGTGGCTAAGTTCCGC	28
			TTCACAGTGGCTAAGTTCCGC	13		TTCACAGTGGCTAAGTTCCG	2
27b			TTCACAGTGGCTAAGTTCTGC	110		TTCACAGTGGCTAAGTTCTG	2
			TTCACAGTGGCTAAGTTCTG	20		TTCACAGTGGCTAAGTTCTGC	1
			TTCACAGTGGCTAAGTTCTGCA	10			
27b*						AGAGCTTAGCTGATTGGTGAACA	1
						AGAGCTTAGCTGATTGGTGAAC	
28-5p						AAGGAGCTCACAGTCTATTGAG	19
						AAGGAGCTCACAGTCTATTGA	18
						AAGGAGCTCACAGTCTATTG	2
						CAAGGAGCTCACAGTCTATTGA	1
						AGGAGCTCACAGTCTATTGAG	1
28-3p	CACTAGATTGTGAGCTCCTGGA	37	CACTAGATTGTGAGCTCCTGGA	39		CACTAGATTGTGAGCTCCTGGA	28
	CACTAGATTGTGAGCTCCTGGAA	10	CACTAGATTGTGAGCTCCTGGAA	13		CTAGATTGTGAGCTCCTGGAG	2
						CACTAGATTGTGAGCTCCTGG	2
28-5p	AAGGAGCTCACAGTCTATTGA	13					
	AAGGAGCTCACAGTCTATTGAG						
29a	TAGCACCATCTGAAATCGGTTA	30	TAGCACCATCTGAAATCGGTTA	511		TAGCACCATCTGAAATCGGTTA	116
			TAGCACCATCTGAAATCGGTT	24		TAGCACCATCTGAAATCGGTT	24
			TAGCACCATCTGAAATCGGTTAT	11		TAGCACCATCTGAAATCGGT	18
						TAGCACCATCTGAAATCGG	6
						TAGCACCATCTGAA	5
						TAGCACCATCTGAAATCGGTTAT	3
						CTAGCACCATCTGAAATCGGTTA	3

						CTAGCACCATCTGAAATCGGTT	2
						TAGCACCATCTGAAA	1
						TAGCACCATCTGAAATCG	1
29a*						ACTGATTTCTTTTGGTGTTCAG	3
						ACTGATTTCTTTTGGTGTTCAGA	1
29b						TAGCACCATTTGAAATCAGTGTT	18
						TAGCACCATTTGAAATCAGTGT	5
						TAGCACCATTTGAAATCAGT	4
						TAGCACCATTTGAAATCAGTGTTT	3
						TAGCACCATTTGAAATCAGTG	1
						TAGCACCATTTGAAATCAG	1
29c			TAGCACCATTTGAAATCGGTTA	15			
30a	TGTAAACATCCTCGACTGGAAGC	22	TGTAAACATCCTCGACTGGAAGCT	71		TGTAAACATCCTCGACTGGAAGC	22
	TGTAAACATCCTCGACTGGAAGCT	22	TGTAAACATCCTCGACTGGAAGC	39		TGTAAACATCCTCGACTGGAAGCT	4
	TGTAAACATCCTCGACTGGAAG		TGTAAACATCCTCGACTGGAAG			TGTAAACATCCTCGACTGGAA	1
						TGTAAACATCCTCGACTGGAAG	
30a*	CTTTCAGTCGGATGTTTGCAGC	50	CTTTCAGTCGGATGTTTGCAGC	83		CTTTCAGTCGGATGTTTGCAGC	16
	CTTTCAGTCGGATGTTTGCAGT	38	CTTTCAGTCGGATGTTTGCAGT	23		CTTTCAGTCGGATGTTTGCAG	3
			TTTCAGTCGGATGTTTGCAGC	13		CTTTCAGTCGGATGTTTGCAGCT	1
30b						TGTAAACATCCTACACTCAGC	1
						TGTAAACATCCTACACTCAGCT	
30b*						CTGGGAGGTGGATGTTTACTTC	1
						CTGGGAGGTGGATGTTTAC	1
30c-2*	CTGGGAGAAGGCTGTTTACTCT	16	CTGGGAGAAGGCTGTTTACTCT	14		TGTAAACATCCTACACTCTCAGC	7
						TGTAAACATCCTACACTCTCAGCT	2
30d	TGTAAACATCCCCGACTGGAAGCT	22	TGTAAACATCCCCGACTGGAAGCT	140		TGTAAACATCCCCGACTGGAAG	9
	UGUAAACAUCCCCCGACUGGAAG		TGTAAACATCCCCGACTGGAAGC	36		TGTAAACATCCCCGACTGGAAGCT	9
			TGTAAACATCCCCGACTGGAAGCTT	11		TGTAAACATCCCCGACTGGAAGC	6
			TGTAAACATCCCCGACTGGAAG	10			
30e						TGTAAACATCCTTGACTGGAAGCT	3
						TGTAAACATCCTTGACTGGAAGC	2
						TGTAAACATCCTTGACTGGAAG	
30e*	CTTTCAGTCGGATGTTTACAGT	82	CTTTCAGTCGGATGTTTACAGC	117		CTTTCAGTCGGATGTTTACAG	4
	CTTTCAGTCGGATGTTTACAGC	55	CTTTCAGTCGGATGTTTACAGT	84		CTTTCAGTCGGATGTTTACA	3
	CTTTCAGTCGGATGTTTACAG	14	CTTTCAGTCGGATGTTTACAG	14		CTTTCAGTCGGATGTTTACAGC	1
31	AGGCAAGATGCTGGCATAGCTG	31				AGGCAAGATGCTGGCATAGCTGT	30
	AGGCAAGATGCTGGCATAGCT	27				AGGCAAGATGCTGGCATAGCTG	29
	AGGCAAGATGCTGGCATAGCTGT	19				AGGCAAGATGCTGGCATAGCT	19
						AGGCAAGATGCTGGCATAGC	5
						GGCAAGATGCTGGCATAGCTG	3
						GGCAAGATGCTGGCATAGC	1
						GGCAAGATGCTGGCATAGCT	1
						GGCAAGATGCTGGC	1
92a	TATTGCACTTGTCCCGGCCTGT	18	TATTGCACTTGTCCCGGCCTGT	34		TATTGCACTTGTCCCGGCCTGT	5
						TATTGCACTTGTCCCGGCC	1
						ATTGCACTTGTCCCGGCCTGT	1
92b	TATTGCACTCGTCCCGGCCTCC	19	TATTGCACTCGTCCCGGCCTCC	530			
			TATTGCACTCGTCCCGGCCTC	96			
			TATTGCACTCGTCCCGGCCTCCT	35			
			TATTGCACTCGTCCCGGCCT	29			
			TATTGCACTCGTCCCGGCCTCA	28			
			TATTGCACTCGTCCCGGCCTCT	23			
			TATTGCACTCGTCCCGGCCTCCAT	22			
			TATTGCACTCGTCCCGGCCTCCTAT	22			
			TATTGCACTCGTCCCGGCCTCCA	17			
			TATTGCACTCGTCCCGGCCTCCATC	17			
			TATTGCACTCGTCCCGGCC	12			
92b*	AGGGACGGGACGCGGTGCAGTGTT	12	AGGGACGGGACGCGGTGCAGTGT	77			
	AGGGACGGGACGCGGTGCAGTGT	10	AGGGACGGGACGCGGTGCAGTGTT	60			
	AGGGACGGGACGCGGUGCAGUG		AGGGACGGGACGCGGTGCAGT	18			
			AGGGACGGGACGCGGTGCAGTGTTT	18			
			AGGGACGGGACGCGGUGCAGUG				
93	CAAAGTGCTGTTCGTGCAGGTAG	81	CAAAGTGCTGTTCGTGCAGGTAG	66		CAAAGTGCTGTTCGTGCAGGTAG	4
						CAAAGTGCTGTTCGTGCAGGT	1
						CAAAGTGCTGTTCGTGCAG	1
98			TGAGGTAGTAAGTTGTATTGTT	18		TGAGGTAGTAAGTTGTATTGT	11
						TGAGGTAGTAAGTTGTATTGTT	10
						TGAGGTAGTAAG	2
						TGAGGTAGTAA	2
						TGAGGTAGTAAGTTGTATTG	1
99a			AACCCGTAGATCCGATCTTGTG	30			
			AACCCGTAGATCCGATCTTGT	14			
99b	CACCCGTAGAACCGACCTTGC	25	CACCCGTAGAACCGACCTTGC	63			
	CACCCGTAGAACCGACCTTGCG	22	CACCCGTAGAACCGACCTTGCG	54			
100			AACCCGTAGATCCGAACTTGTG	91		AACCCGTAGATCCGAACTTGTG	36
			AACCCGTAGATCCGAACTTGT	45		AACCCGTAGATCCGAACTTGT	15
			AACCCGTAGATCCGAACTTGTGA	16		AACCCGTAGATCCGAACTT	3
			AACCCGTAGATCCGAAC	10		ACCCGTAGATCCGAACTTGTG	2
			AACCCGTAGATCCGAACTTGTGT	10		AACCCGTAGATCCGAACTTG	2
						AAACCCGTAGATCCGAACTTGTG	1
						AAACCCGTAGATCCGAACT	1

						AACCCGTAGATCCGAACT	1
101	TACAGTACTGTGATAACTGAAG	165	TACAGTACTGTGATAACTGAAG	334	TACAGTACTGTGATAACTGAA	2
	GTACAGTACTGTGATAACTGAA	80	TACAGTACTGTGATAACTGAAT	112	GTACAGTACTGTGATAACTGAA	1
	GTACAGTACTGTGATAACTGAAA	32	GTACAGTACTGTGATAACTGAA	102	TACAGTACTGTGATAACTGA	1
	TACAGTACTGTGATAACTGAA	30	TACAGTACTGTGATAACTGAA	63		
	TACAGTACTGTGATAACTGAAT	29	TACAGTACTGTGATAACTGAAA	53		
	GTACAGTACTGTGATAACTGA	23	GTACAGTACTGTGATAACTGAAA	25		
	TACAGTACTGTGATAACTGAAA	20	GTACAGTACTGTGATAACTGA	22		
			TACAGTACTGTGATAACTGAAGT	22		
			TACAGTACTGTGATAACTGAAGA	13		
103	AGCAGCATTGTACAGGGCTATGA	393	AGCAGCATTGTACAGGGCTATGA	196	AGCAGCATTGTACAGGGCTATGA	24
	AGCAGCATTGTACAGGGCTAT	104	AGCAGCATTGTACAGGGCTAT	178	AGCAGCATTGTACAGGGCTATG	3
	AGCAGCATTGTACAGGGCTATG	26	AGCAGCATTGTACAGGGCTATGAT	18	AGCAGCATTGTACAGGGCTATGAA	3
	AGCAGCATTGTACAGGGCTATGAT	11	AGCAGCATTGTACAGGGCTATG	17	GCAGCATTGTACAGGGCTATGA	1
106b	TAAAGTGCTGACAGTGCAGATA	38	TAAAGTGCTGACAGTGCAGATA	36	TAAAGTGCTGACAGTGCAGAT	8
	TAAAGTGCTGACAGTGCAGAT	33	TAAAGTGCTGACAGTGCAGAT	35	TAAAGTGCTGACAGTGCAGA	4
			TAAAGTGCTGACAGTGCAGA	11	TAAAGTGCTGACAGTGCAGATA	4
					TAAAGTGCTGACAGTGCAG	1
					TAAAGTGCTGACAGTGCAGATAGT	1
106b*	CCGCACTGTGGGTACTTGCTGC	10				
107	AGCAGCATTGTACAGGGCTATCA	10			AGCAGCATTGTACAGGGCT	3
					AGCAGCATTGTACAGGGC	2
					AGCAGCATTGTACAGGGCTAT	1
					AGCAGCATTGTACAGGGCTA	1
					AGCAGCATTGTACAGGGCTATCA	
124	TAAGGCACGCGGTGAATGCCAA	25	TAAGGCACGCGGTGAATGCCAA	23		
	TAAGGCACGCGGTGAATGCC		TAAGGCACGCGGTGAATGCC			
125a-5p					TCCCTGAGACCCTTTAACCTGTGA	3
125a-3p					ACAGGTGAGGTTCTTGGGAGC	2
					ACAGGTGAGGTTCTTGGGAGCC	1
125b-1*					ACGGGTTAGGCTCTTGGGAGCT	7
					ACGGGTTAGGCTCTTGGGAGC	3
					ACGGGTTAGGCTCTTGGGAG	2
125b			TCCCTGAGACCCTAACTTGTGA	246	TCCCTGAGACCCTAACTTGTGA	78
			TCCCTGAGACCCTAACTTGTG	17	TCCCTGAGACCCTAACTTGTG	7
			TCCCTGAGACCCTAAC	10	TCCCTGAGACCCTAACTTGT	6
					TCCCTGAGACCCTAACTTGTGAT	2
					TCCCTGAGACCCTAAC	2
					CCCTGAGACCCTAACTTGTGAT	1
					TCCCTGAGACCCTAACTTG	1
127-3p					TCGGATCCGTCTGAGCTTGGCT	6
					TCGGATCCGTCTGAGCTTGGC	3
					CGGATCCGTCTGAGCTTGGCT	1
128	TCACAGTGAACCGGTCTCTTT	12	TCACAGTGAACCGGTCTCTTT	40		
130a	CAGTGCAATGTTAAAAGGGCAT	888	CAGTGCAATGTTAAAAGGGCAT	509	CAGTGCAATGTTAAAAGGGCAT	116
	CAGTGCAATGTTAAAAGGGCATA	52	CAGTGCAATGTTAAAAGGGCAC	81	CAGTGCAATGTTAAAAGGGCA	13
	CAGTGCAATGTTAAAAGGGCATT	20	CAGTGCAATGTTAAAAGGGCATT	35	CAGTGCAATGTTAAAAGGGCATT	10
	CAGTGCAATGTTAAAAGGGC	14	CAGTGCAATGTTAAAAGGGCA	25	CAGTGCAATGTTAAAAGGGC	9
	CAGTGCAATGTTAAAAGGGCA	12	CAGTGCAATGTTAAAAGGGCATA	18	CAGTGCAATGTTAAAAG	1
			CAGTGCAATGTTAAAAGGGC	13	CAGTGCAATGTTAAAAGGG	1
					CAGTGCAATGTTAAAA	1
130b	CAGTGCAATGATGAAAGGGCAT	257	CAGTGCAATGATGAAAGGGCAT	208	CAGTGCAATGATGAAAGGGCAT	79
	CAGTGCAATGATGAAAGGGCATT	26	CAGTGCAATGATGAAAGGGCATT	20	CAGTGCAATGATGAAAGGGC	7
	CAGTGCAATGATGAAAGGGCATA	18	CAGTGCAATGATGAAAGGGCA	12	CAGTGCAATGATGAAAGGGCA	5
	CAGTGCAATGATGAAAGGGCA	15	CAGTGCAATGATGAAAGGGCATA	10	CAGTGCAATGATGAAAGGG	1
					AGTGCAATGATGAAAGGGCAT	1
132					TAACAGTCTACAGCCATGGTCG	14
					GTAACAGTCTACAGCCATGGTC	1
					TAACAGTCTACAGCCATGGT	1
134					TGTGACTGGTTGACCAGAGGGG	2
					TGTGACTGGTTGACCAGAGGG	1
135b			TATGGCTTTTCATTCCTATGTGA	20		
138			AGCTGGTGTTGTGAATCAGGCCG	11	AGCTGGTGTTGTGAATCAGGCCG	1
					AGCTGGTGTTGTGAATCAGGCCGTT	1
140	ACCACAGGGTAGAACCACGGAC	62	ACCACAGGGTAGAACCACGGAC	110	TACCACAGGGTAGAACCACGGA	3
	TACCACAGGGTAGAACCACGG		TACCACAGGGTAGAACCACGGAC	24	TACCACAGGGTAGAACCACGGAC	2
			CCACAGGGTAGAACCACGGAC	12	ACCACAGGGTAGAACCACGGAC	2
			ACCACAGGGTAGAACCACGGA	11	ACCACAGGGTAGAACCACGGA	1
			TACCACAGGGTAGAACCACGG		TACCACAGGGTAGAACCACGGACA	1
					TACCACAGGGTAGAACCACGG	
140-5p					CAGTGGTTTTACCCTATGGTAG	3
141	TAACACTGTCTGGTAAAGATGGC	25				
	UAACACUGUCUGGUAAAGAUGG					
143					TGAGATGAAGCACTGTAGCTC	68
					TGAGATGAAGCACTGTAGCT	14
					TGAGATGAAGCACTGTAGC	12
					TGAGATGAAGC	2

145						GTCCAGTTTTCCCAGGAATCCC	28
						GTCCAGTTTTCCCAGGAATCCCT	24
						GTCCAGTTTTCCCAGGAAT	20
						GTCCAGTTTTCCCAGGAA	13
						GTCCAGTTTTCCCAGGAATCCCTT	12
						GTCCAGTTTTCCCAGGAATCC	5
						GTCCAGTTTTCCCAGGAATC	4
						GTCCAGTTTTCCC	3
						GTCCAGTTTTCCCAGGA	2
						GTCCAGTTTTCCCAGG	1
						GTCCAGTTTTCCCAGGAATCCCTTA	1
146a						TGAGAACTGAATTCCATGGGTT	7
						TGAGAACTGAATTCCATGGGTTG	1
146b			TGAGAACTGAATTCCATAGGCTGT	57			
			TGAGAACTGAATTCCATAGGCTGG	31			
			UGAGAACUGAAUUCCAUAGGCU				
148a	TCAGTGCACTACAGAACTTTGT	110	TCAGTGCACTACAGAACTTTGT	79			
	TCAGTGCACTACAGAACTTTGTC	16					
148b			TCAGTGCATCACAGAACTTTGT	12			
151-3p	CTAGACTGAAGCTCCTTGAGG	290	CTAGACTGAAGCTCCTTGAGG	294	CTAGACTGAAGCTCCTTGAGGA	18	
	CTAGACTGAAGCTCCTTGAGGA	162	CTAGACTGAAGCTCCTTGAGGA	178	CTAGACTGAAGCTCCTTGAGG	6	
	CTAGACTGAAGCTCCTTGAGGAA	64	CTAGACTGAAGCTCCTTGAG	58	TACTAGACTGAAGCTCCTTGAG	1	
	CTAGACTGAAGCTCCTTGAG	49	CTAGACTGAAGCTCCTTGAGGT	54			
	CTAGACTGAAGCTCCTTGAGGT	25	CTAGACTGAAGCTCCTTGAGGAA	43			
	CTAGACTGAAGCTCCTTGAGGAAA	18	CTAGACTGAAGCTCCTTGAGT	25			
	CTAGACTGAAGCTCCTTGAGA	13	CTAGACTGAAGCTCCTTGAGGAT	14			
			CTAGACTGAAGCTCCTTGAGA	13			
			TACTAGACTGAAGCTCCTTGAGG	13			
			TACTAGACTGAAGCTCCTTGAG	12			
			CTAGACTGAAGCTCCTTGA	11			
			CTAGACTGAAGCTCCTTGAGGAAA	10			
151-5p	TCGAGGAGCTCACAGTCTAGT	12	TCGAGGAGCTCACAGTCTAGT	37	TCGAGGAGCTCACAGTCTAGT	1304	
			TCGAGGAGCTCACAGTCTAGTA	20	TCGAGGAGCTCACAGTCTAGTA	1135	
					TCGAGGAGCTCACAGTCTAG	91	
					TCGAGGAGCTCACAGTCTAGTAT	79	
					CTCGAGGAGCTCACAGTCTAGT	25	
					CTCGAGGAGCTCACAGTCTAGTA	19	
					TCGAGGAGCTCACAGTCTAGTATG	12	
					CTCGAGGAGCTCACAGTCTAGTAT	6	
					TCGAGGAGCTCACAGTCT	3	
					TCGAGGAGCTCACAGTCTA	3	
					TCGAGGAGCTCACAGT	2	
					TCGAGGAGCTCACAG	2	
					CTCGAGGAGCTCACAGTCTAG	1	
					CTCGAGGAGCTCACAGTCT	1	
					CGAGGAGCTCACAGTCTAGTATGT	1	
					CGAGGAGCTCACAGTCTAGTATG	1	
					CGAGGAGCTCACAGTCTAGTAT	1	
152					TCAGTGCATGACAGAACTTGG	21	
					TCAGTGCATGACAGAACTTGGG	8	
					TCAGTGCATGAC	1	
					CAGTGCATGACAGAACTTGGG	1	
154*					AATCATACACGGTTGACCTAT	1	
					AATCATACACGGTTGACCTATT		
154					TAGGTTATCCGTGTTGCCTTCG	1	
155					TTAATGCTAATCGTGATAGGGGTTT	1	
					TTAATGCTAATCGTGATAGGGGTT	1	
					TTAATGCTAATCGTGATAGGGGT	1	
181a			AACATTCAACGCTGTCGGTGAGTTT	104	AACATTCAACGCTGTCGGTGAGT	1	
			AACATTCAACGCTGTCGGTGAGTT	32	AACATTCAACGCTGTCGGTGAGTTT	1	
			AACATTCAACGCTGTCGGTGAGT	17	AACATTCAACGCTGTCGGTGAG	1	
181b			AACATTCATTGCTGTCGGTGGGTT	10	AACATTCATTGCTGTCGGTGGGTT	6	
			AACATTCATTGCTGTCGGTGGGT		AACATTCATTGCTGTCGGTGGGT	3	
					AACATTCATTGCTGTCGGTGGG	2	
					AACATTCATTGCTGTCGGTGGGTTT	1	
182	TTTGGCAATGGTAGAACTCACACT	432					
	TTTGGCAATGGTAGAACTCACACTGG	166					
	TTTGGCAATGGTAGAACTCACAC	148					
	TTTGGCAATGGTAGAACTCACACTG	95					
	TTTGGCAATGGTAGAACTCACACTGGT	20					
	TTTGGCAATGGTAGAACTCACA	13					
	TTTGGCAATGGTAGAACTCACACTGGA	12					
	TTTGGCAATGGTAGAACTCACACTGT	11					
183	ATGGCACTGGTAGAATTCACTG	133					
	ATGGCACTGGTAGAATTCACT	69					
	TATGGCACTGGTAGAATTCACT	66					
	TATGGCACTGGTAGAATTCACTG	45					
	ATGGCACTGGTAGAATTCACTGT	20					
	ATGGCACTGGTAGAATTCACTGA	11					
185	TGGAGAGAAAGGCAGTTCCTGA	29	TGGAGAGAAAGGCAGTTCCTGA	54			
186			CAAAGAATTCTCCTTTTGGGCTT	11	CAAAGAATTCTCCTTTTGGGC	3	
			CAAAGAATTCTCCTTTTGGGCT		CAAAGAATTCTCCTTTTGGGCTT	1	
					CAAAGAATTCTCCTTTTGGGCT		
186*					GCCCAAAGGTGAATTTTTTGGG	1	

191	CAACGGAATCCCAAAAGCAGCTG	76	CAACGGAATCCCAAAAGCAGCTG	49	CAACGGAATCCCAAAAGCAGCTG	275
	CAACGGAATCCCAAAAGCAGCTGT	25	CAACGGAATCCCAAAAGCAGCTGT	40	CAACGGAATCCCAAAAGCAGCT	117
	CAACGGAATCCCAAAAGCAGCT	23	CAACGGAATCCCAAAAGCAGCT	13	CAACGGAATCCCAAAAGCAGCTGT	115
	CAACGGAATCCCAAAAGCAGCTGA	11	CAACGGAATCCCAAAAGCAGCTGA	12	CAACGGAATCCCAAAAGCAGC	95
					CAACGGAATCCCAAAAGCAG	17
					AACGGAATCCCAAAAGCAGCTG	11
					CAACGGAATCCCAAAAGC	3
					CAACGGAATCCCAAAAGCA	1
					AACGGAATCCCAAAAGCAGC	1
					CAACGGAATCC	1
					AACGGAATCCCAAAAGCAGCTGT	1
193a-5p					TGGGTCTTTGCGGGCGAGATGA	2
193a	AACTGGCCTACAAAGTCCCAGT	15			AACTGGCCTACAAAGTCCCAGT	98
					AACTGGCCTACAAAGTCCCAG	4
					AACTGGCCTACAAAGTCCCAGTT	3
					CAACTGGCCTACAAAGTCCCAGT	1
					ACTGGCCTACAAAGTCCCAGT	1
194					TGTAACAGCAACTCCATGTGGAA	2
					TGTAACAGCAACTCCATGTGGA	1
195			TAGCAGCACAGAAATATTGGCA	17		
			UAGCAGCACAGAAAUAUUGGC			
199a-5p					CCCAGTGTTCAGACTACCTGTTC	10
					CCCAGTGTTCAGACTACCTGTT	2
					CCCAGTGTTCAGAC	1
199b-5p					CCCAGTGTTTAGACTATCTGTTC	3
199b-3p					ACAGTAGTCTGCACATTGGTTA	11
					ACAGTAGTCTGCACATTGGTT	3
					TACAGTAGTCTGCACATTGGTT	2
					TACAGTAGTCTGCACATTGGT	2
					CAGTAGTCTGCACATTGGTTA	1
					CAGTAGTCTGCACATTG	1
					CAGTAGTCTGCACATTGGT	1
210	CTGTGCGTGTGACAGCGGCTGA	10	CTGTGCGTGTGACAGCGGCTGA	199		
214					ACAGCAGGCACAGACAGGCAGT	13
					ACAGCAGGCACAGACAGGCAG	8
					TACAGCAGGCACAGACAGGCAG	3
					TACAGCAGGCACAGACAGGCAGT	3
					TACAGCAGGCACAGACAGGC	1
					ACAGCAGGCACAGACAGGCA	1
					CAGCAGGCACAGACAGGCAG	1
					TACAGCAGGCACAGACAGGCA	1
219-2-3p	AGAATTGTGGCTGGACATCTGT	20				
221	AGCTACATTGTCTGCTGGGTTTC	15			AGCTACATTGTCTGCTGGGTTTC	16
					AGCTACATTGTCTGCTGGGT	6
					AGCTACATTGTCTGCTGGGTTT	6
					AGCTACATTGTCTGCTGGGTTTCA	2
					GCTACATTGTCTGCTGGGTTT	1
					GCTACATTGTCTGCTGGGT	1
					AGCTACATTGTCTGCTGG	1
					GCTACATTGTCTGCTGGGTTTC	1
					AGCTACATTGTCTGCTGGGTT	1
					AGCTACATTGTCTGCTG	1
221*	ACCTGGCATACAATGTAGATTTCTGT	72	ACCTGGCATACAATGTAGATTTCTGT	93	ACCTGGCATACAATGTAGATTT	1
	ACCTGGCATACAATGTAGATTTCT	17	ACCTGGCATACAATGTAGATTTCT	22		
	ACCTGGCATACAATGTAGATTTC	10	ACCTGGCATACAATGTAGATTTCTG	14		
	ACCTGGCATACAATGTAGATTT		ACCTGGCATACAATGTAGATTT			
222	AGCTACATCTGGCTACTGGGTCT	21	AGCTACATCTGGCTACTGGGTCT	35	AGCTACATCTGGCTACTGGGTCTC	18
	AGCTACATCTGGCTACTGGGT		AGCTACATCTGGCTACTGGGT		AGCTACATCTGGCTACTGGGTCT	11
					AGCTACATCTGGCTACTGGGT	5
					AGCTACATCTGGCTACTGGGTC	5
					GCTACATCTGGCTACTGGGTCTC	3
					AGCTACATCTGGCTACTGGGTCTCT	3
					CTACATCTGGCTACTGGGTCT	1
					TACATCTGGCTACTGGGTCTC	1
					AGCTACATCTGGCTACTGGG	1
					AGCTACATCTGGCTACTGG	1
					GCTACATCTGGCTACTGGGTCTCT	1
222*					GGCTCAGTAGCCAGTGTAGATCC	1
					CTCAGTAGCCAGTGTAGATCCT	
299-3p					TATGTGGGATGGTAAACCGCT	2
					TATGTGGGATGGTAAACCGCTT	1
302a	TAAGTGCTTCCATGTTTTGGTGA	54				
	TAAGTGCTTCCATGTTTTGGTG	18				
	AAGTGCTTCCATGTTTTGGTGA	16				
302a*	TAAACGTGGATGTACTTGCTTT	1113				
	TAAACGTGGATGTACTTGCTT	491				
	CTTAAACGTGGATGTACTTGCTT	132				
	CTTAAACGTGGATGTACTTGCT	109				
	ACTTAAACGTGGATGTACTTGCT	96				
	TAAACGTGGATGTACTTGCTTTGA	86				
	ACTTAAACGTGGATGTACTTGC	80				
	TAAACGTGGATGTACTTGCT	67				
	TAAACGTGGATGTACTTGCTTTG	24				

	TAAACGTGGATGTACTTGCTTTGAAACT	23					
	TAAACGTGGATGTACTTGCTTTA	22					
	TAAACGTGGATGTACTTGCTTTGAAAC	20					
	TAAACGTGGATGGACTTGCTTT	16					
	TTAAACGTGGATGTACTTGCTT	15					
	TAAACGTGGATGTACTTGCTTTGAA	12					
302b	TAAGTGCTTCCATGTTTTAGTAG	60					
	TAAGTGCTTCCATGTTTTAGTA	26					
	TAAGTGCTTCCATGTTTTAGT	12					
	TAAGTGCTTCCATGTTTTAGTAGT	12					
302c	AAGTGCTTCCATGTTTCAGTGGT	51					
	AAGTGCTTCCATGTTTCAGTGG	11					
	TAAGTGCTTCCATGTTTCAGTGG						
302d	TAAGTGCTTCCATGTTTGAGTGT	63					
	AAGTGCTTCCATGTTTGAGTGT	12					
302d*	ACTTTAACATGGAGGCACTTGCT	18					
	ACTTTAACATGGAGGCACTTGC	15					
	ACTTTAACATGGAGGCACTTG	13					
320a	AAAAGCTGGGTTGAGAGGGCGA	221	AAAAGCTGGGTTGAGAGGGCGA	190	AAAAGCTGGGTTGAGAGGGCGA	3	
	AAAAGCTGGGTTGAGAGGGCGAA	62	AAAAGCTGGGTTGAGAGGGCGAT	89	AAGCTGGGTTGAGAGGGCGAAA	1	
	AAAAGCTGGGTTGAGAGGGCGAT	54	AAAAGCTGGGTTGAGAGGGCGAA	63	AAAAGCTGGGTTGAGAGGGCGAA	1	
	AAAAGCTGGGTTGAGAGGGCGT	24	AAAAGCTGGGTTGAGAGGGCGAAT	20	AAAGCTGGGTTGAGAGGGCGA	1	
	AAAAGCTGGGTTGAGAGGGCGAAT	10	AAAAGCTGGGTTGAGAGGGCGT	18	AAAAGCTGGGTTGAGAGGGCGAAA	1	
			AAAGCTGGGTTGAGAGGGCGA	13			
324-5p					CGCATCCCCTAGGGCATTGGTGT	2	
					CGCATCCCCTAGGGCATTGGTG	1	
329					AACACACCTGGTTAACCTCTT	6	
					ACACACCTGGTTAACCTCTT	2	
					AACACACCTGGTTAACCTCTTT		
335	TCAAGAGCAATAACGAAAAATG	65	TCAAGAGCAATAACGAAAAATG	51	TCAAGAGCAATAACGAAAAATGT	3	
	TCAAGAGCAATAACGAAAAAT	58	TCAAGAGCAATAACGAAAAAT	25	TCAAGAGCAATAACGAAA	2	
	TCAAGAGCAATAACGAAAAATGT	33	TCAAGAGCAATAACGAAAAATGT	25	TCAAGAGCAATAACGAAAAAT	2	
					TCAAGAGCAATAACGAAAAATG	1	
335*					TTTTTCATTATTGCTCCTGACC	8	
					TTTTTCATTATTGCTCCTGAC	2	
337-3p					TCCTATATGATGCCTTTCTTC	63	
					CTCCTATATGATGCCTTTCTTC	30	
					CTCCTATATGATGCCTTTCTT	7	
					TCCTATATGATGCCTTTCTT	6	
					TCCTATATGATGCCTTTCTTCA	1	
					CTCCTATATGATGCCTTTCT	1	
					TCCTATATGATGCCTTTCT	1	
339-5p					TCCCTGTCCTCCAGGAGCTCA	1	
					TCCCTGTCCTCCAGGAGCTCAC	1	
					TCCCTGTCCTCCAGGAGCTCACG	1	
340	TTATAAAGCAATGAGACTGATT	381	TTATAAAGCAATGAGACTGATT	487			
	TTATAAAGCAATGAGACTGAT	72	TTATAAAGCAATGAGACTGAT	149			
	TTATAAAGCAATGAGACTGA	15	TTATAAAGCAATGAGACTGA	32			
345					GCTGACTCCTAGTCCAGGGCTC	4	
					GCTGACTCCTAGTCCAGGGC	1	
					GCTGACTCCTAGTCCAGGGCT	1	
361-5p					TTATCAGAATCTCCAGGGGTAC	8	
					TTATCAGAATCTCCAGGGGTA	1	
					TTATCAGAATCTCCAGGGGTACT	1	
363	AATTGCACGGTATCCATCTGTA	12					
365					TAATGCCCCTAAAAATCC	1	
					TAATGCCCCTAAAAATCCT	1	
					TAATGCCCCTAAAAATCCTTA	1	
					TAATGCCCCTAAAAATCCTTAT		
367	AATTGCACTTTAGCAATGGTGA	34					
	AATTGCACTTTAGCAATGGT	11					
369-3p					AATAATACATGGTTGATCTTT	18	
					AATAATACATGGTTGATCTT	6	
					AATAATACATGGTTGATCTTTT	4	
					AATAATACATGGTTGATC	1	
					ATAATACATGGTTGATCTTT	1	
374a	TTATAATACAACCTGATAAGT	24	TTATAATACAACCTGATAAGT	35	TTATAATACAACCTGATAAGT	2	
	TTATAATACAACCTGATAAGTG		TTATAATACAACCTGATAAGTG		TTATAATACAACCTGATAAGTG	1	
374b	ATATAATACAACCTGCTAAGTG	24	ATATAATACAACCTGCTAAGTG	31			
376a					ATCATAGAGGAAAATCCACGT	207	
					AATCATAGAGGAAAATCCACGT	17	
					TCATAGAGGAAAATCCACGT	7	
					ATCATAGAGGAAAATCCAC	5	
					ATCATAGAGGAAAATCCACGTT	4	
					ATCATAGAGGAAAATCCACG	4	
376b					ATCATAGAGGAAAATCCATGT	116	
					ATCATAGAGGAAAATCCATGTT	27	
					ATCATAGAGGAAAATCCATGTTT	8	
					TCATAGAGGAAAATCCATGT	6	

						TCATAGAGGAAAATCCATGTTT	5
						ATCATAGAGGAAAATCCAT	2
						ATCATAGAGGAAAATCCATG	1
						ATCATAGAGGAAAATCCATGTTTT	1
						TCATAGAGGAAAATCCATGTT	1
376c						AACATAGAGGAAATTCCACGT	384
						ACATAGAGGAAATTCCACGT	25
						AACATAGAGGAAATTCCACGTT	19
						AACATAGAGGAAATTCCACG	5
						AACATAGAGGAAA	4
						AACATAGAGGAAATTCCAC	3
						AAACATAGAGGAAATTCCACGT	3
						AAACATAGAGGAAATTCCACG	2
						AACATAGAGGAAATTCCACGTTT	2
						AACATAGAGGA	2
						ACATAGAGGAAATTCCACGTT	1
						ACATAGAGGAAATTCCAC	1
						AACATAGAGGAAATTCC	1
						TAAACATAGAGGAAATTCCACG	1
						AACATAGAGGAAAT	1
377*						AGAGGTTGCCCTTGGTGAATTC	6
						AGAGGTTGCCCTTGGTGAAT	2
						AGAGGTTGCCCTTGGTGAATTCG	1
						AGAGGTTGCCCTTGGTGAATT	1
378	ACTGGACTTGGAGTCAGAAGGC	214	ACTGGACTTGGAGTCAGAAGGC	114	ACTGGACTTGGAGTCAGAAGG		14
	ACTGGACTTGGAGTCAGAAGGCA	86	ACTGGACTTGGAGTCAGAAGGCA	39	ACTGGACTTGGAGTCAGAAGGC		2
	ACTGGACTTGGAGTCAGAAGG	47	ACTGGACTTGGAGTCAGAAGG	37			
	CTGGACTTGGAGTCAGAAGGC	32	CTGGACTTGGAGTCAGAAGGC	27			
	ACTGGACTTGGAGTCAGAAGGCAT	21	ACTGGACTTGGAGTCAGAAGGCAT	21			
	ACTGGACTTGGAGTCAGAAG	19	CTGGACTTGGAGTCAGAAGGCT	13			
	ACTGGACTTGGAGTCAGAAGGCAA	12	ACTGGACTTGGAGTCAGAAG	11			
	CTGGACTTGGAGTCAGAAGGCA	11					
379			TGGTAGACTATGGAACGTAGG	21	TGGTAGACTATGGAACGTAGG		4
					ATGGTAGACTATGGAACGTAGG		1
					TGGTAGACTATGGAACGTAG		1
379*					TATGTAACATGGTCCACTAAC		12
					ATGTAACATGGTCCACTAACT		2
					TATGTAACATGGTCCACTAA		2
					TATGTAACATGGTCCACTAACT		
382					GAAGTTGTTCGTGGTGGATTCG		15
					GAAGTTGTTCGTGGTGGATTC		7
					GAAGTTGTTCGTGGTGGATT		5
					GAAGTTGTTCGTGGTGGAT		4
					AGTTGTTCGTGGTGGATTC		2
					AGTTGTTCGTGGTGGATTCGCT		1
					AAGTTGTTCGTGGTGGAT		1
					AAGTTGTTCGTGGTGGATTC		1
409-3p					CGAATGTTGCTCGGTGAACCCCT		2
					GAATGTTGCTCGGTGAACCCCT		2
					CGAATGTTGCTCGGTGAACCCC		1
					CGAATGTTGCTCGGTGAACCC		1
409-5p					GGTTACCCGAGCAAC		1
					AGGTTACCCGAGCAACTTTGCAT		
411					TAGTAGACCGTATAGCGTACG		1
411*					TATGTAACACGGTCCACTAAC		14
					TATGTAACACGGTCCACTAA		12
					TATGTAACACGGTCCACTAACC		
423	TGAGGGGCAGAGAGCGAGACTTT	529	TGAGGGGCAGAGAGCGAGACTTT	1512	AGCTCGGTCTGAGGCCCCTCAGT		5
	TGAGGGGCAGAGAGCGAGACTTTT	85	TGAGGGGCAGAGAGCGAGACTTTT	328			
	TGAGGGGCAGAGAGCGAGACTT	38	TGAGGGGCAGAGAGCGAGACTT	203			
	TGAGGGGCAGAGAGCGAGACT	31	TGAGGGGCAGAGAGCGAGACT	140			
	TGAGGGGCAGAGAGCGAGACTTTA	19	TGAGGGGCAGAGAGCGAGAC	42	TGAGGGGCAGAGAGCGAGAC		2
			TGAGGGGCAGAGAGCGAGACTTTA	38	TGAGGGGCAGAGAGCGAGACTTT		2
			TGAGGGGCAGAGAGCGAGA	31	TGAGGGGCAGAGAGCGAGACT		2
			TGAGGGGCAGAGAGCGAGACTTTTT	27	TGAGGGGCAGAGAGCGAGA		1
			TGAGGGGCAGAGAGCGAGACTTA	16	TGAGGGGCAGAGAGCGAGACTT		1
			TGAGGGGCAGAGAGCGAGACA	13	TGAGGGGCAGAGAGCGAGACTTTT		1
			TGAGGGGCAGAGAGCGAGAAA	11	TGAGGGGCAGAGA		1
424*					CAAAACGTGAGGCGCTGCTATA		2
					CAAAACGTGAGGCGCTGCTAT		1
424					CAGCAGCAATTCATGTTTTGAA		51
					CAGCAGCAATTCATGTTTTGA		20
					CAGCAGCAATTCATGTTTTG		1
425					AATGACACGATCACTCCCGTTGAGT		6
					ATGACACGATCACTCCCGTTGAGT		1
					AATGACACGATCACTCCCGTTGAG		1
					AATGACACGATCACTCCCGTTGA		
425*					TCGGGAATGTCGTGTCCGCCC		1
					TCGGGAATGTCGTGTCCGCC		1
					CATCGGGAATGTCGTGTCCGCC		1
					ATCGGGAATGTCGTGTCCGCCC		
432					TCTTGGAGTAGGTCATTGGGTGG		9
					TCTTGGAGTAGGTCATTGGGTG		2
					TTGGAGTAGGTCATTGGGTGG		1

433						ATCATGATGGGCTCCTCGGTGT	3
						ATCATGATGGGCTCCTCGGTG	1
						ATCATGATGGGCTCCTCGGT	1
450b			TTTTGCAATATGTTCCTGAAT	12			
			TTTTGCAATATGTTCCTGAATA				
542-3p			TGTGACAGATTGATAACTGAAA	16			
548f	AAAACTGTAATTACTTTTGGAC	11					
	AAAAACTGTAATTACTTTT						
455-3p						GCAGTCCATGGGCATATACAC	4
						ATGCAGTCCATGGGCATAT	3
						GCAGTCCATGGGCATATACA	2
						ATGCAGTCCATGGGCATATACAC	1
						TGCAGTCCATGGGCATATACAC	1
						GCAGTCCATGGGCATATACACT	1
						ATGCAGTCCATGGGCATATAC	1
483-5p						AGACGGGAGGAAAGAAGGGAGT	1
						AAGACGGGAGGAAAGAAGGGAG	1
484						TCAGGCTCAGTCCCCTCCCGAT	4
						TCAGGCTCAGTCCCCTCCCGA	2
485-5p						AGAGGCTGGCCGTGATGAATTCG	2
						AGAGGCTGGCCGTGATGAATT	2
						AGAGGCTGGCCGTGATGAATTC	
487b						AATCGTACAGGGTCATCCACTT	2
						ATCGTACAGGGTCATCCACTTT	1
490-5p						CCATGGATCTCCAGGTGGGT	3
						CCATGGATCTCCAGGTGGGTCA	1
						CCATGGATCTCCAGGTGGG	1
491-5p						AGTGGGGAACCCTTCCATGAGGA	26
						AGTGGGGAACCCTTCCATGAGG	2
						TGGGGAACCCTTCCATGAGGA	1
						TGGGGAACCCTTCCATGAGGAGT	1
						AGTGGGGAACCCTTCCATGA	1
						AGTGGGGAACCCTTCCATG	1
493*						TTGTACATGGTAGGCTTTCATT	91
						TTGTACATGGTAGGCTTTCAT	20
						TTGTACATGGTAGGCTTTCA	5
						TTGTACATGGTAGGCTTTC	3
494						TGAAACATACACGGGAAACCTCT	4
						TGAAACATACACGGGAAAC	2
						TGAAACATACACGGGAAACC	1
						TGAAACATACACGGGAAACCTC	
495						AAACAAACATGGTGCACTTCTT	90
						AAACAAACATGGTGCACTTCTTT	71
						AACAAACATGGTGCACTTCTT	21
						AAACAAACATGGTGCACTTCTTTT	16
						AACAAACATGGTGCACTTCTTT	13
						AAACAAACATGGTGCACTTCT	13
						AAACAAACATGGTGCACTT	4
						AACAAACATGGTGCACTTCTTTT	3
						AAACAAACATGGTGCACTTC	2
						ACAAACATGGTGCACTTCTTTTT	1
						AACAAACATGGTGCACTTCTTTTT	1
						ACAAACATGGTGCACTTCTTTT	1
						AACAAACATGGTGCACTTC	1
						AACAAACATGGTGCACTTCT	1
501-3p						ATGCACCCGGGCAAGGATTCT	1
						AATGCACCCGGGCAAGGATTCT	
501-5p						AATCCTTTGTCCCTGGGTGAGAGT	1
						AATCCTTTGTCCCTGGGTGAGA	
503	TAGCAGCGGGAACAGTTCTGAAA	13				TAGCAGCGGGAACAGTTCTGCA	5
	TAGCAGCGGGAACAGTTCTGCAG					TAGCAGCGGGAACAGTTCTGCAG	3
						TAGCAGCGGGAACAGTTCT	1
						TAGCAGCGGGAACAGTTCTGCAGT	1
532	CATGCCTTGAGTGTAGGACCGT	42	CATGCCTTGAGTGTAGGACCGT	68			
539						AGAAATTATCCTTGGTGTGTTC	1
						GGAGAAATTATCCTTGGTGTGT	1
						GGAGAAATTATCCTTGGTGTGTT	1
542-3p						TGTGACAGATTGATAACTGAAA	3
						TGTGACAGATTGATAACTGAAAG	1
652						AATGGCGCCACTAGGGTTGTGC	1
						ATGGCGCCACTAGGGTTGTG	1
						AATGGCGCCACTAGGGTTGTG	
654-5p						TGGTGGGCCGCAGAACATGTGC	3
						TGGTGGGCCGCAGAACATGTGCT	1
						TGGTGGGCCGCAGAACATGT	1
654-3p						TATGTCTGCTGACCATCACC	2
						CATATGTCTGCTGACCATCACC	1
						TATGTCTGCTGACCATCAC	1

					TATGTCTGCTGACCATCACCTT	
708	AAGGAGCTTACAATCTAGCTGGG	23			AAGGAGCTTACAATCTAGCTGGG	2
708*	AACTAGACTGTGAGCTTCTAGA CAACUAGACUGUGAGCUUCUAG	11				
744	TGCGGGGCTAGGGCTAACAGCA	21	TGCGGGGCTAGGGCTAACAGCA TGCGGGGCTAGGGCTAACAGC TGCGGGGCTAGGGCTAACAGCAA	74 40 12	TGCGGGGCTAGGGCTAACAGCA TGCGGGGCTAGGGCTAACAGC	2 1
769	TGAGACCTCTGGGTTCTGAGCT	11	TGAGACCTCTGGGTTCTGAGCT	29		
769-3p					TGGGATCTCCGGGGTCTTGGTT CTGGGATCTCCGGGGTCTTGGTT	1
1180					TTTCCGGCTCGCGTGGGTGTGT TTTCCGGCTCGCGTGGGTGTGTAG	11 1
1185					AGAGGATACCCTTTGTATGTTCA AGAGGATACCCTTTGTATGTTC AGAGGATACCCTTTGTATGTT	2 7
1255b					CGGATGAGCAAAGAAAGTGGTT	1
1260	ATCCCACCGCTGCCACCA	10	ATCCCACCGCTGCCACCA	18		
1270	CTGGAGATATGGAAGAGCTGTGT	27				
1275	GTGGGGGAGAGGCTGT GTGGGGGAGAGGCTGTC	13	GTGGGGGAGAGGCTGT GTGGGGGAGAGGCTGTC	12		
1286					TCTGGGCAACAAAGTGAGA TCTGGGCAACAAAGTGAGACCT	1
1296					TTAGGGCCCTGGCTCCATC TTAGGGCCCTGGCTCCATCT TTAGGGCCCTGGCTCCATCTCC	1 1
1301					TGCAGCTGCCTGGGAGTGAC TTGCAGCTGCCTGGGAGTGACTTC	1
1305					TTTCAACTCTAATGGGAGAGAC TTTTCAACTCTAATGGGAGAGA	1
1323	TCAAAACTGAGGGGCATTTTCT	53				
2110					TTGGGGAAACGGCCGCTGAGTGAG TTGGGGAAACGGCCGCTGAGTG	1
let-7a	TGAGGTAGTAGGTTGTATAGTT	20	TGAGGTAGTAGGTTGTATAGTT TGAGGTAGTAGGTTGTATAGT TGAGGTAGTAGGTTGTATAGTTA TGAGGTAGTAGGTTGTATAGTTT TGAGGTAGTAGGTTGTATAG TGAGGTAGTAGGTTGTATAGTTAA TGAGGTAGTAGGTTGTAT TGAGGTAGTAGGGTGTATAGTT	2798 525 146 100 29 12 11 10	TGAGGTAGTAGGTTGTATAGTT TGAGGTAGTAGGTTGTATAGT TGAGGTAGTAGGTTGTATAG TGAGGTAGTAGGTTGTATAGTTT TGAGGTAGTAGGTTGTATA GTGAGGTAGTAGGTTGTATAGT TGAGGTAGTAGGTTGTATAGTTTT ATGAGGTAGTAGGTTGTATAGTT TTGAGGTAGTAGGTTGTATAGT TTGAGGTAGTAGGTTGTATAGTT	1222 712 55 37 15 4 3 2 1 1
let-7b			TGAGGTAGTAGGTTGTGTGGTT TGAGGTAGTAGGTTGTGTGGT TGAGGTAGTAGGTTGTGTGGTTT TGAGGTAGTAGGTTGTGTGGTTA	253 50 38 24	TGAGGTAGTAGGTTGTGTGGTT TGAGGTAGTAGGTTGTGTGGT TGAGGTAGTAGGTTGTGTGGTTT TGAGGTAGTAGGTTGTGTGG TGAGGTAGTAGGTTGTGTG TGAGGTAGTAGGTTGTGT GAGGTAGTAGGTTGTGTGGTT	90 58 41 15 3 3 1
let-7c			TGAGGTAGTAGGTTGTATGGTT TGAGGTAGTAGGTTGTATGGTTA TGAGGTAGTAGGTTGTATGGT TGAGGTAGTAGGTTGTATGGTTT	399 35 23 18	TGAGGTAGTAGGTTGTATGGTT TGAGGTAGTAGGTTGTATGGT	12 1
let-7d			AGAGGTAGTAGGTTGCATAGTT AGAGGTAGTAGGTTGCATAGTTT AGAGGTAGTAGGTTGCATAGT	156 15 12	AGAGGTAGTAGGTTGCATAGTT AGAGGTAGTAGGTTGCATAGT AGAGGTAGTAGGTTGCATAGTTT GAGGTAGTAGGTTGCATAGTT AGAGGTAGTAGGTTGCATAG GAGGTAGTAGGTTGCATAGT AAGAGGTAGTAGGTTGCATAGTT AGAGGTAGTAGG AGAGGTAGTAGGTTGCATAGTTTT AAGAGGTAGTAGGTTGCATAG AGAGGTAGTAGGTTGCATA GAGGTAGTAGGTTGCATAGTTT	1072 99 72 34 30 5 3 3 2 1 1 1
let-7d*					TATACGACCTGCTGCCTTTCT CTATACGACCTGCTGCCTTTCT	1
let-7e			TGAGGTAGGAGGTTGTATAGTT TGAGGTAGGAGGTTGTATAGT TGAGGTAGGAGGTTGTATAGTTT TGAGGTAGGAGGTTGTATAGTTA	198 44 38 31		
let-7f			TGAGGTAGTAGATTGTATAGTT TGAGGTAGTAGATTGTATAGT TGAGGTAGTAGATTGTATAGTTA TGAGGTAGTAGATTGTATAGTTT TGAGGTAGTAGATTGTATAG	3135 649 261 150 37	TGAGGTAGTAGATTGTATAGTT TGAGGTAGTAGATTGTATAGT TGAGGTAGTAGA TGAGGTAGTAGATTGTATAG TGAGGTAGTAGATTGTATA	131 101 19 14 10

	TGAGGTAGTAGATTGTAT	19	ATGAGGTAGTAGATTGTATAGT	4
	TGAGGTAGTAGATTGTATAGTTATC	17	TGAGGTAGTAGATTGTATAGTTT	2
	TGAGGTAGTAGATTGTATA	17	ATGAGGTAGTAGATTGTATAGTT	2
	TGAGGTAGTAGATTGTATAGTTC	12	TGAGGTAGTAGATTGTAT	2
	TGAGGTAGTAGATTGTATAGTTG	11	TGAGGTAGTAGAT	1
			TGAGGTAGTAGATT	1
let-7g	TGAGGTAGTAGTTTGTACAGTT	146	TGAGGTAGTAGTTTGTACAGTT	91
	TGAGGTAGTAGTTTGTACAGT	40	TGAGGTAGTAGTTTGTACAGT	65
			TGAGGTAGTAGTTTGTACAG	8
			TGAGGTAGTAGTTTGTAC	4
			TGAGGTAGTAGTTTGTACAGTTT	3
let-7i	TGAGGTAGTAGTTTGTGCTGTT	95	TGAGGTAGTAGTTTGTGCTGTT	26
	TGAGGTAGTAGTTTGTGCTGT	16	TGAGGTAGTAGTTTGTGCTGT	7
	TGAGGTAGTAGTTTGTGCTGTTT	16	TGAGGTAGTAGTTTGTGC	5
	TGAGGTAGTAGTTTGTGCTGTTA	10	TGAGGTAGTAGTTTGTGCT	2
let-7i*			CTGCGCAAGCTACTGCCTTGCT	1

Table S3.1
Sequencing number of canonical microRNAs and their isomiRs in human embryonic stem cells (hESC) and neuronal stem cells (NSC) and mesenchymal stem cells (MSC). Canonical/ annotated (highlighted in yellow) and isomiRs were detected by Solexa (hESC and NSC) and 454 (MSC) sequencing. Sequencing frequency of microRNA was listed. SN - sequencing number. Cloning and sequencing was performed by Elcie Chan.

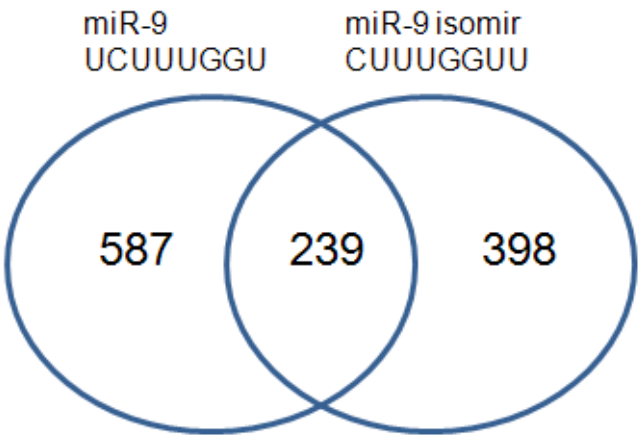
ES	5' different	5' end of miRNA	Canonical seed	IsomiR seed
101	1 addition	GUACAGUACU	ACAGUAC	UACAGUA
183	1 deletion	UAUGGCACUG	AUGGCAC	UGGCACU
302a	1 deletion	UAAGUGCUUC	AAGUGCU	AGUGCUU
302a*	1-3 deletion	ACUUAAACGUG	CUUAAAC	AAACGUG
302d	1 deletion	UAAGUGCUUC	AAGUGCU	AGUGCUU
378	1 deletion	ACUGGACUUG	CUGGACU	UGGACUU

NS	5' different	Seed sequence	Canonical seed	IsomiR seed
9	1-2 deletion	UCUUUGGUUA	CUUUGGU	UUUGGUU
9*	1 deletion	AUAAAGCUAG	UAAAGCU	AAAGCUA
21	1,4 deletion	UAGCUUAUCA	AGCUUAU	GCUUAUC
30a*	1 deletion	CUUUCAGUCG	UUUCAGU	UUCAGUC
101	1 addition	GUACAGUACU	ACAGUAC	UACAGUA
140	1-2 deletion	UACCACAGGG	ACCACAG	CCACAGG
151-3p	2 addition	UACUAGACUGA	UAGACUG	ACUAGAC
320a	1 deletion	AAAAGCUGGGU	AAAGCUG	AAGCUGG
378	1 deletion	ACUGGACUUG	CUGGACU	UGGACUU

MS	5' different	Seed sequence	Canonical seed	IsomiR seed
10a	1 deletion	UACCCUGUAG	ACCCUGU	CCCUGUA
16*	1 addition	ACCAAUAUUA	CAAUAUU	CCAAUAU
17	1 addition	UCAAAGUGCU	AAAGUGC	CAAAGUG
23a	1 deletion	AUCACAUUGC	UCACAUU	CACAUUG

23b	1 deletion	AUCACAUUGC	UCACAUU	CACAUUG
26a	1 deletion	UUCAAGUAAU	UCAAGUA	CAAGUAA
28-3p	2 deletion	CACUAGAUUGU	ACUAGAU	UAGAUUG
29a	1 addition	CUAGCACCAU	AGCACCA	UAGCACC
31	1 deletion	AGGCAAGAUG	GGCAAGA	GCAAGAU
100	1 deletion	AACCCGUAGA	ACCCGUA	CCCGUAG
140-3p	1 deletion	UACCACAGGG	ACCACAG	CCACAGG
151-5p	1 addition	CUCGAGGAGC	CGAGGAG	UCGAGGA
191	1 deletion	CAACGGAAUC	AACGGAA	ACGGAAU
199b-3p	1 addition	UACAGUAGUC	CAGUAGU	ACAGUAG
214	1 addition	UACAGCAGGC	CAGCAGG	ACAGCAG
222	1 deletion	AGCUACAUCU	GCUACAU	CUACAUC
337-3p	1 deletion	CUCCUAUAUG	UCCUAUA	CCUAUAU
376a	1-2 deletion	AAUCAUAGAGG	AUCAUAG	CAUAGAG
376b	1 deletion	AUCAUAGAGGA	UCAUAGA	CAUAGAG
376c	1 deletion	AACAUAGAGG	ACAUAGA	CAUAGAG
379*	1 deletion	UAUGUAACAU	AUGUAAC	UGUAACA
382	2 deletion	GAAGUUGUUCG	AAGUUGU	GUUGUUC
409-3p	1 deletion	CGAAUGUUGC	GAAUGUU	AAUGUUG
455-3p	2 addition	AUGCAGUCCAU	CAGUCCA	UGCAGUC
495	1 deletion	AAACAAACAU	AACAAAC	ACAAACA
let7a	1 addition	AUGAGGUAGUA	GAGGUAG	UGAGGUA
let7d	1 deletion	AGAGGUAGUAG	GAGGUAG	AGGUAGU
let7f	1 addition	AUGAGGUAGUA	GAGGUAG	UGAGGUA

Table S3.2
The commonest isomiRs that show 5' differences. Analysis was based on the deep sequencing results, generated from hESC, NSC and MSC (see Table S3.1). The listed canonical and isomiR seed regions were used to generate target predictions (See Figure 3.6)



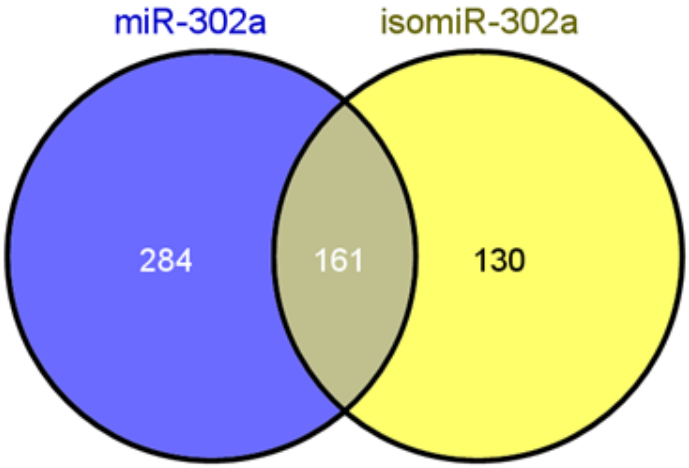
miR-9 CUUUGGU			Common targets			IsomiR-9 UUUGGUU	
ABCA1	FNDC3B	PTMA	ACVR1B	SLC8A1		A2BP1	LOC51136
ABCD1	FOXF2	PXDN	ADAM11	SLC9A1		AAK1	LPGAT1
ACCN2	FOXG1B	PYGO2	ADAMTS6	SMARCD2		ABHD3	LRIG3
ACOT7	FOXN2	QKI	ALCAM	SMURF2		ACSL1	LRP8
ACTR1A	FOXO1A	RAB34	AP1S2	SNIP1		ACSL4	LRRC41
ACVR2B	FOXO3A	RAB38	AP3B1	SNRK		ADCY1	LRRN3
ADAM10	FRMD4A	RAB40B	ARHGAP24	SOCS5		ADRA2C	LRRTM3
ADAMTS3	FRY	RAB5B	ARID1A	SORT1		ADSS	LYSMD2
ADAMTS5	FURIN	RAG1AP1	ARID1B	STARD13		AKAP2	MAFB
ADCY9	FXR1	RAI14	ARL4C	STK3		ALKBH5	MAFG
ADPGK	FYCO1	RALB	ARNT	STS-1		ANKRD40	MAGI2
AEBP2	FYTTD1	RALGDS	ATBF1	SYT9		ANKRD50	MAML1
AFAP1	FZD8	RAP1B	ATP1B1	TESK2		AP3S1	MAP2K4
AFF1	GAB2	RAP2A	ATP7A	TGFBI		ARHGAP5	MAP3K2
AK3L1	GABBR2	RASD2	BACE1	TLK1		ARHGEF12	MAP3K5
AKAP7	GAD1	RBM24	BAHD1	TMCC1		ARHGEF9	MAP4K4
ALS2CR13	GALNT1	RBM5	BCL6	TMEM16A		ARID4B	MAP9
AMMECR1L	GALNT17	RBM9	BCLAF1	TNKS		ARPC5	MAPK9
AMOTL1	GALNT3	RBMS3	C14orf129	TNRC6A		ARPP-21	MAX
AMOTL2	GCH1	REEP4	C18orf25	TNRC6B		ATF7	MEF2D
ANK2	GIT2	RERE	C3orf58	TRAM1		ATG5	METAP2
ANP32B	GOLPH3	RFX1	C9orf150	TRIM2		ATP10A	MGC4655
ANXA2	GOPC	RFXDC2	CACNB2	TRPM7		ATP2B1	MITF
AP2M1	GOSR1	RHOJ	CAMTA1	TSC22D2		ATP2B2	MLL3
AP4E1	GOT1	RIMS2	CAPZA1	UBE2H		ATRX	MLLT3
APPBP2	GPAM	RIMS3	CARM1	UBE2R2		AUTS2	MLSTD2
ARCN1	GPATCH8	RIMS4	CBFA2T2	UBE3C		AVPR1A	MTF1
ARFGEF1	GPBP1L1	RIPK5	CCNG1	UTRN		B3GNT2	MTMR10
ARFIP2	GPC6	RNF11	CELSR2	VANGL1		BAAT	MTMR4
ARHGDIA	GPR124	RNF128	ChGn	VCL		BACE2	MTUS1
ARHGEF17	GPR158	RNF144	CLOCK	XYLT1		BAZ2A	MTX2
ARHGEF2	GPR85	RNF150	CNNM1	YPEL2		BAZ2B	MXD1
ARID3B	GPRASP2	RNF19	CNNM2	ZBTB39		BCL9	MXI1
ARL1	GRSF1	RNF44	CNOT6L	ZBTB41		BDNF	N4BP3
ARMCX2	GZF1	RPS6KA2	CNTN3	ZFHX4		BICD2	NAGS
ARPC1A	HCN2	RPS6KA4	COL15A1	ZKSCAN1		BRD1	NARG1
ASXL1	HDAC4	RTN4RL1	CPEB1	ZNF236		C10orf12	NCAM2
ATF1	HECW2	RUNX1	CPEB2	ZNF319		C17orf63	NDST3
ATOH8	HIAT1	RUTBC2	CPEB4	ZNF354A		C19orf22	NEDD9
ATP11A	HISPPD1	RXRA	CREB3L2	ZNF618		C20orf23	NEGR1
ATP11C	HISPPD2A	S100PBP	CREB5			C20orf77	NEUROG2
ATP8B2	HMGA2	SCN2B	CREBZF			C5orf5	NFATC1
ATXN1	HN1L	SCUBE3	CRIM1			C9orf25	NIPA1
ATXN3	HNRPA3	SCYL3	CSDA			C9orf58	NLK
ATXN7	HOXC13	SDC2	CSNK1A1			CACNB1	NR2F6
AUH	HRB	SEC31A	CXCR4			CACNG3	NR3C2
AXL	HRBL	SEMA6D	DCUN1D4			CALD1	NRBF2
B4GALNT1	ICMT	SENP1	DCX			CALM1	NRK
BAG4	IGF2BP1	SERTAD2	DDX17			CAP1	NUP153
BAIAP2	IKZF2	SEZ6	DDX3X			CAPN5	NUP160
BAZ2B	IKZF5	SFRS10	DEDD			CAST	OTUD5
BCAT2	IPO11	SFRS6	DHX40			CBLN2	OTX2
BCL11A	IPO13	SFXN2	DIAPH2			CCNC	PALLD
BCL2L11	IPO4	SGCD	DIO2			CCND1	PALM2-AKAP2
BICC1	ISL1	SGMS1	DLX3			CCNJ	PCDH9
BIN1	ITGA6	SH2B3	DOCK9			CDC14A	PCGF2
BMPER	ITPKC	SH3BP4	DTNA			CDC2L6	PDZRN3
BRPF3	JAKMIP2	SH3GLB1	DYNC1LI2			CDH2	PER2

BRUNOL4	JMJD2C	SHB	DYRK1A	CDYL	PER3
BRUNOL6	JUB	SHC1	EDEM3	CGI-09	PFN2
BSN	JUP	SHC2	EIF4E	CHD4	PHF16
BTBD10	KCNK4	SHROOM3	ERBB2IP	CHD6	PHF17
BTBD14B	KCNMB2	SIN3A	ERC2	CHIC1	PHF6
BTBD7	KIAA0174	SIRT1	ESR1	CITED4	PHYHIPL
BTG2	KIAA0284	SIX5	EVI5L	CKAP4	PLCB1
C11orf58	KIAA0323	SLC10A3	FAM107B	CLCN5	PLEC1
C11orf9	KIAA0423	SLC16A2	FAM19A4	CLINT1	POU2F3
C14orf101	KIAA0562	SLC19A2	FAM46C	CLIP3	POU3F1
C14orf147	KIAA0738	SLC1A1	FAM55C	CLP1	POU4F1
C16orf70	KIAA0831	SLC20A2	FBN1	CNOT6	PPARBP
C17orf85	KIAA1217	SLC22A17	FBXW2	COL1A1	PPAT
C1QL1	KIAA1219	SLC26A7	FGF12	COPS2	PPP2CA
C1QL3	KIAA1468	SLC27A4	FLJ11171	CPLX2	PPP2R2D
C20orf39	KIAA1539	SLC30A3	FLRT3	CPNE8	PPP2R5A
C21orf25	KIAA1553	SLC30A8	FNIP1	CPT1A	PPP2R5E
C21orf91	KIAA1913	SLC33A1	FOXP1	CSMD2	PRKACB
C22orf9	KIF13B	SLC39A13	FOXP4	CSN3	PRKG1
C3orf57	KIF21A	SLC39A14	FREM2	CSNK1A1L	PSMF1
C5orf30	KITLG	SLC5A3	FRMD6	CSNK1G1	PTPN9
C6orf106	KLF12	SLC6A6	FRYL	CSNK2A2	PURB
C7orf42	KLF13	SLC7A8	FSTL1	CTDSPL	PWWP2
C8ORFK32	KLHDC5	SLITRK2	GABRB2	CTGF	RAB11FIP2
CA7	KLHL1	SMARCE1	GALNAC4S-6ST	CTTNBP2NL	RAB30
CAMK2N1	KLHL24	SMC1A	GLCCI1	CUL4B	RAB31
CAMKK2	KLHL3	SMEK1	GMEB2	CXXC5	RAB8B
CBL	KSR1	SMPD3	GNPNAT1	CYP2U1	RAD23B
CBLN4	LAMP1	SNAP23	GRHL1	DCUN1D1	RALBP1
CBX1	LDLRAD3	SNF1LK	HDAC5	DDHD1	RANBP9
CBX6	LEPRE1	SNX16	HIC2	DDX3Y	RAPH1
CC2D1A	LHFP	SNX25	HIPK1	DDX50	RARG
CC2D1B	LIFR	SNX7	HIST1H4A	DICER1	RASAL2
CCDC126	LIN7C	SOAT1	HLCS	DLG2	RBPMS2
CCDC43	LMBRD2	SON	HMG20A	DLG3	RFFL
CCDC6	LMX1A	SORBS3	HOXA11	DMXL1	RFT1
CCNE1	LOC153222	SP4	HUNK	DMXL2	RHOA
CCNE2	LOC162073	SP7	ID4	DNAJA5	RICS
CD47	LOC201725	SPECC1L	IGF2BP2	DNAJC6	RICTOR
CDC73	LOC401720	SPG20	IGF2BP3	DNAL4	RNF2
CDH1	LPHN1	SPIN1	IKZF4	DNMT3B	RNF6
CENTG2	LPP	SPTLC2	INSIG1	DSCAML1	ROCK1
CEP350	LRCH4	SRC	ITM2B	DST	RP11-493K23.2
CHMP2B	LSM14A	SRCAP	ITM2C	DUSP10	RP11-493K23.3
CHSY1	LUC7L2	SRF	KALRN	DUSP5	RP2
CLCN5	LZTS2	SRGAP3	KCNA1	DYNLL2	RPIA
CLCN6	MAEA	SRPK1	KCNJ2	EDG3	RRBP1
CLDN2	MAP1B	SSX2IP	KCTD12	EEF1A1	RUTBC3
CMTM6	MAP2K3	ST8SIA4	KIAA0240	EFCBP1	SATB2
CNNM4	MAP2K7	STAM	KIAA1128	EHF	SCN1A
CNOT1	MAP3K3	STC1	KIAA1598	EIF3S1	SCN2A
CNOT7	MAPKAPK2	STCH	KIAA1967	EIF4G3	SCN8A
CNTFR	MAR06/	STK38L	KIF1C	EIF5A2	SCOC
CNTN4	MDGA1	STMN1	KLF5	EML5	SEC61B
COL12A1	MESDC1	STX1A	KLHL18	ENOX2	SEP15/
COL18A1	MGA	SYAP1	KPNB1	ENTPD7	SETBP1
COL4A2	MGC13057	SYNJ1	LARP1	EPHB1	SETD7
COL5A1	MKNK2	SYNJ2BP	LDLRAP1	ERF	SETD8
COL9A1	MLXIP	SYNPR	LIN28	EYA1	SF3A3
COLEC12	MME	SYT10	LIN28B	FAM13A1	SLC12A2
CPEB3	MMP15	SYT4	LMNA	FAM63B	SLC17A6

CSNK1G2	MMP16	T,brachuury	LOC196463	FAM80A	SLC25A26
CTDP1	MNT	TARDBP	LRRC1	FBXL3	SLC30A7
CTDSP2	MRFAP1	TBC1D22A	LRRTM4	FBXO10	SLC38A2
CTHRC1	MTHFD1L	TBC1D4	M6PR	FBXO11	SLC39A10
CTLA4	MTMR2	TBL1XR1	MAF	FGL2	SLC6A8
CUL4A	MTPN	TBPL1	MAMDC1	FJX1	SLITRK5
D4ST1	MYH1	tcag7.1017	MAN1A2	FLJ20366	SLMAP
DACT3	MYH9	TCF2	MAP1A	FLJ39378	SMAD2
DBNL	MYO1C	TCF7	MAP3K7IP3	FLJ45557	SMARCC1
DBT	MYO1D	TES	MAP7	FMNL2	SMPD4
DCBLD2	MYOHD1	TFRC	MBNL1	FRMPD4	SNIP
DCC	NAGPA	TGFBR2	MIER3	G3BP2	SNX7
DENND1A	NCOA3	TGOLN2	MTHFD2	GABARAPL1	SNX9
DGKB	NCOA7	THBS2	MUM1L1	GABPA	SOBP
DIP	NDRG1	TLN1	MYPN	GADD45A	SOCS2
DIXDC1	NEFH	TM9SF3	NAP1L1	GALNACT-2	SOCS7
DKFZp667G2110	NEK1	TMEM109	NCOA1	GALNT7	SORL1
DLGAP2	NEK9	TMEM28	NCOR2	GATA3	SP1
DNAJA4	NID2	TNFRSF21	NEDD4	GFRA2	SP3
DNAJB1	NMT1	TNS1	NFASC	GJA3	SPG21
DNAJC14	NMT2	TOX2	NFIB	GLCE	SPRED2
DNAJC8	NOX4	TRIM55	NFKB1	GLS	SPTY2D1
DOLPP1	NPY2R	TSNAX	NHLH2	GORASP2	SSU72
DPP4	NRF1	TSPAN9	NOTCH2	GOT2	ST6GALNAC3
DR1	NRP1	TTYH2	NPTX1	GPR177	STAG2
DRD2	NTNG1	TULP4	NR5A2	GRAMD3	STAT5B
DSCR1L1	NTRK3	TXNDC5	ONECUT1	GRIA2	STIM1
DSE	NUTF2	UBE2Q1	ONECUT2	GRIA4	STX16
DTD1	ODZ1	UBE2Z	OSBPL3	GRIN2A	SUMO1
DUSP6	OPCML	UBE4B	OTUD7B	GRM7	SV2A
DYNC1LI2	OSBPL11	UBFD1	OXSR1	GTF2H1	SYNE2
DYRK1B	OTUD4	UBN1	P4HA2	H2AFZ	TBC1D2B
EDD1	PACSIN1	UBR1	PAK2	HBP1	TBL1X
EFNA1	PAK6	UBXD8	PAK4	HIVEP2	TCF7L2
EGLN3	PALMD	UHRF1	PANK3	HLF	TEAD1
EGR3	PB1	ULK2	PARG	HNRPH3	TFAP2B
EHD4	PCBP4	UNC13A	PCGF5	HP1BP3	TFG
EIF5	PCDH10	USP31	PCSK2	HRNBP3	TGFBR1
ELAVL1	PCGF6	VAMP3	PGAP1	HS2ST1	THAP4
ELAVL2	PCNP	VAT1	PGRMC2	HYOU1	TMED5
ELF1	PCNX	VAV3	PHF20L1	IDH1	TMED7
ELL	PCSK6	VDAC3	PHF21A	IFRD2	TMEM1
ELOVL4	PDE11A	VGLL4	PHLDB2	IGF2R	TMEM150
ELOVL7	PDE7B	WAPAL	PI4KII	INOC1	TMEM32
EMB	PDGFC	WDR20	PIK3R3	INPP5A	TMTC2
EN1	PDGFRB	WDTC1	PITPNM2	IQCH	TOB2
EN2	PDK4	WIPF1	PMP22	IQGAP2	TP53INP1
ENTPD5	PERQ1	WNT4	POU2F2	IRS2	TPPP
EPAS1	PHF13	WSB1	POU3F2	IRX3	TRAF3
EPHA7	PHF15	XPO4	PRDM1	ITGA10	TRIM36
EPHB2	PHF8	XRN1	PRUNE	ITPKC	TRIM39
ERG	PHLDB1	YAP1	PURA	JAK1	TTC7B
ETS1	PHOSPHO1	YIPF4	RAB8A	JAKMIP2	TTYH3
EXTL3	PHTF2	ZBTB38	RANBP17	JAZF1	UBE2J2
FAM117A	PIAS3	ZBTB4	RANBP2	JMJD2A	UBL3
FAM123B	PIGZ	ZC3H10	RAP2C	JOSD1	UBQLN2
FAM13C1	PIM3	ZC3H12B	RCOR1	JPH1	UBQLN3
FAM19A5	PIP5K2B	ZDHHC5	RHOBTB1	KCNK2	USP37
FAM43B	PJA2	ZDHHC7	RHOQ	KCNS2	USP46
FAM46A	PLD3	ZIC5	RNF111	KCTD6	USP9X
FAM58A	PLEKHA1	ZNF131	RNF24	KIAA0152	VCP

FAM62A	PLEKHA6	ZNF248	RNF38	KIAA0355	VDAC1
FAM62B	POLR3G	ZNF263	ROD1	KIAA0427	VIP
FAM73B	PPAPDC2	ZNF324	RP11-130N24.1	KIAA0494	WDFY3
FAM91A1	PPARA	ZNF365	RP11-217H1.1	KIAA0523	WHSC2
FBN2	PPARD	ZNF395	RUNX2	KIAA1409	XKR6
FBS1	PPP1R13B	ZNF407	RYBP	KIAA1522	XLKD1
FBXL11	PPP2R5D	ZNF629	SACS	KLF9	XPO1
FBXL16	PRDM10	ZNRF2	SAPS3	L3MBTL3	YPEL1
FBXL2	PRKCA		SCRIB	LASS2	YTHDF3
FBXL3	ProSAPiP1		SDC1	LATS1	ZBTB7A
FBXO33	PRRT2		SFRS1	LCOR	ZCCHC8
FBXW11	PRRT3		SH3BGRL2	LENG8	ZDHHC17
FCMD	PRRX1		SHANK2	LEPR	ZDHHC21
FGF5	PSD3		SHROOM4	LEPROTL1	ZIC1
FLI1	PSEN1		SIX4	LETMD1	ZMIZ1
FLJ20294	PTBP1		SLC12A5	LOC133308	ZMYM2
FLJ25476	PTBP2		SLC25A24	LOC152485	ZNF148
FMR1	PTCH1		SLC31A2	LOC339745	ZNF364
FNBP1	PTER		SLC35B3	LOC374395	ZNF664
	PTGFRN				

Table S3.3A
Predicted targets of miR-9 and isomiR-9. Tested targets were highlighted in yellow. Target prediction was performed using TargetScan Human and TargetScan Custom.



miR-302a		Common targets		isomiR-302a
ABHD3	MTERFD2	ABCA1	TMUB2	A2BP1
ACAD9	MTMR4	AEBP2	TNFAIP1	ABR
ACTL6A	MUTED	ANKRD13B	TOX	ADCYAP1
ACVR1	MYBL1	APP	TSHZ3	AP3M1
ALKBH5	MYCN	ARHGEF17	UBE2B	ARF1
ALX4	MYLK	ARID4A	UBE2Q2	ARFIP2
ANK2	MYST3	ARID4B	UBFD1	ARHGEF3
ANKRD13C	MYT1L	ASF1A	UNK	ARHGEF9
ANKRD54	NCOA7	ASF1B	USP42	ARMC8
AOF1	NEO1	ATAD2	USP46	ASH1L
ARHGAP24	NEUROD1	ATP2B2	VLDLR	BACH2

ARHGAP29	NEUROD6	BNC2	VSX1	BAGE2
ARHGEF18	NFIA	C10orf12	WDR37	BAGE3
ASXL2	NFYA	C11orf9	YOD1	BAGE4
ATP2C1	NPAS3	C16orf28	YTHDF3	BAGE5
ATXN1	NR2F2	C6orf107	ZBTB11	BAZ1A
BAHD1	NR4A2	C6orf85	ZBTB41	BTBD3
BAMBI	NUFIP2	C7orf43	ZBTB44	C12orf34
BCL11A	ODZ2	CAMK2N1	ZBTB47	C13orf27
BCL11B	ORMDL3	CC2D1A	ZDHHC17	C19orf48
BCL2L11	OSBPL5	CCND1	ZFP91	C1orf21
BCL6	OTUD4	CCND2	ZFPM2	C5orf25
BLCAP	OXR1	CDC2L6	ZFYVE26	C6orf35
BRMS1L	PARP8	CDCA7	ZKSCAN1	CACNG3
BRP44L	PCAF	CFL2	ZNF148	CCDC6
BTG1	PCDHA1	CNOT6	ZNF385	CDC25B
C10orf104	PCDHA10	CREB5	ZNF436	CDH2
C11orf30	PCDHA12	CREBL1	ZNF650	CDH4
C14orf101	PCDHA13	CRTC2	ZNF800	CDKN1B
C16orf72	PCDHA2	CUGBP2	ZNFX1	CEP350
C19orf43	PCDHA3	CXXC6	ZRANB1	CLDN2
C1orf9	PCDHA4	CYP26B1		CNOT2
C21orf25	PCDHA5	DDHD1		CNR1
CADM1	PCDHA6	DERL2		CPD
CADM2	PCDHA7	DNAJA2		CXXC5
CAMTA1	PCDHA8	EDNRB		DDX3X
CCNJ	PCDHAC1	EPAS1		DLL4
CDK2	PCDHAC2	ERBB4		DNAJC13
CLCN4	PERQ1	ESR1		DOCK3
CLIP4	PFN2	FAM13C1		DPP4
COL23A1	PHF2	FAM46C		ELF4
CORO2B	PHF6	FAM57A		EMX2
CPEB1	PHLPPL	FGD5		FBXW11
CPEB2	PIGS	FLT1		FIGN
CRIM1	PLAGL2	FNDC3A		FNBP1L
CRK	PLEKHA3	GATAD2B		GAB2
CUTL1	PLEKHM1	GRHL2		GAS2
CXCR4	POLQ	HN1		GBAS
DAZAP2	PPP1R10	HRBL		H2AFY
DCUN1D1	PPP1R9A	IGF2BP1		HCN1
DHX40	PPP3R1	IKZF2		HCN4
DKK1	PPP6C	INTS6		HIVEP3
DMTF1	PRR16	IQSEC2		ITFG1
DPP8	PRRG1	IQWD1		ITGA9
DPYSL5	PRRX1	IRF2		ITPR3
DRD1	PTCHD1	JAZF1		KPNB1
DUSP2	PURA	KBTBD8		LASS3

DYRK2	R3HDM1	KIAA1212	LCOR
E2F7	RAB11FIP1	KIAA1522	LDOC1
ECT2	RAB5C	KIF26B	LETMD1
EIF2C1	RAB6A	KIF3B	LMBR1L
ELAVL2	RAB6C	KLHL18	LOC51035
EPHA2	RAB7A	KPNA1	MAT2A
ETV1	RAD23B	LEFTY1	MICALL1
FAM117A	RBBP6	LEFTY2	MLL5
FAM73B	RBBP7	LHX6	MTCH2
FAM78A	RBJ	LMO3	MUM1L1
FBXL11	RBM33	LOC153222	NEK1
FBXO10	RDBP	LYPD6	NKRF
FBXO11	RECK	MAP3K14	NOVA1
FBXO41	REEP3	MBNL2	NT5C3
FEM1C	RHOC	MECP2	OCRL
FGF9	RORB	MED12L	PCDH10
FILIP1L	RPS6KA1	MEF2C	PCGF5
FLJ25476	RPS6KA3	MINK1	PHF15
FLJ45187	RRAGD	MKRN1	PICALM
FNDC3B	RSBN1	MLL	POU4F1
FOXF2	RSBN1L	MMP23B	PPP2R3A
FOXJ2	RSRC2	MMP24	PPP3CA
FOXJ3	RTN1	MNT	PRC1
FOXL2	RUNX1	MTF1	PRKACB
FRMD4A	SCRT2	MTMR3	PTBP2
GABPB2	SDC1	MTUS1	PTPLB
GLIS3	SETD5	NAPE-PLD	QSER1
GNB5	SIPA1L3	NEK9	RAP1A
HBP1	SLAIN1	NFIB	RARB
HDAC4	SLC2A4	NR2C2	RASAL2
HIC2	SMAD2	NR4A3	RBM9
HIF1AN	SNIP	NTN4	RC3H1
HIPK3	SNRK	OLFM3	RDX
HIVEP2	SNX5	OPCML	RET
HLF	SOBP	OSBPL8	RFX4
HLX1	SP3	PAK7	RFXDC2
HNRPH3	SSX2IP	PAN3	RICTOR
HNRPUL1	ST8SIA2	PAPOLA	RKHD2
HNRPUL2	ST8SIA3	PBX3	ROCK1
IGF1R	SUV39H1	PGBD5	SEP02/
IHPK1	SYDE1	PKN2	SLC24A3
INHBB	TAL1	PLAG1	SLC38A2
INOC1	TAPT1	PLXNA1	SMAD9
ITGB4	TBC1D8B	PRDM4	SOCS3
ITGB8	TESK2	PRDM8	SOX11
JAKMIP1	TIAM1	PTPRD	SP1

JOSD1	TLE4	PURB	SPRED1
KBTBD2	TLOC1	RAB11A	SPTLC1
KCNA1	TMEM28	RAB22A	TBL1XR1
KCND2	TNKS2	RABGAP1	THRAP1
KIAA0157	TOB2	RALGDS	TMCO2
KIAA1267	TOX3	RAPGEFL1	TMEM87A
KLF12	TP53INP1	RASSF2	TMTC1
KLF13	TP53INP2	RBL1	TNRC6A
KLHL28	TP73L	RGL1	TNRC6B
KLHL3	TRIM36	RGMA	TOM1L2
KPNA2	TRIP11	RNF38	TUBG1
KREMEN1	TRPS1	RNF6	UBAP1
LAMP2	TRPV6	RPS6KA5	UBL3
LARP4	TUSC2	SAR1B	UHMK1
LASS6	TWF1	SENP1	VAMP3
LATS2	TXNIP	SLC6A9	VANGL2
LEF1	UBE2J1	SLITRK3	VASH2
LHX8	UBE2R2	SNF1LK	WDR26
LMX1A	UBE2W	SNX21	WNT5A
LOC130074	ULK1	SPOP	XYLT1
LOC150786	UNC5A	SS18L1	YTHDC1
LOC162427	VEGFA	SUV420H1	ZBTB26
LRP2	WDR20	SYNC1	ZFR
LUC7L2	WDR45	TARDBP	ZMYM2
LYCAT	WEE1	TFAP4	ZNF180
MAML1	YPEL2	TIPARP	ZNF462
MAP1B	YWHAZ	TMEM16F	ZNF654
MAP3K11	ZBTB4		
MAPK9	ZBTB6		
MAR08/	ZBTB7A		
MBNL1	ZBTB9		
MCCD1	ZDHHC9		
MCL1	ZFHX4		
MFAP3L	ZFP36L2		
MIER3	ZMYND11		
MLL3	ZNF2		
MLLT6	ZNF238		
MOBKL1A	ZNF512B		
MRPS25	ZNF697		

Table S3.3B
Predicted targets of miR-302a and isomiR-302a. Tested targets were highlighted in yellow. Target prediction was performed using TargetScan Human and TargetScan Custom.

hESCs	Predicted target to isomiR only		Predicted targets common to both canonical miRNA and isomiR	
101	158/681	23.2%	414/681	60.1%
183	366/636	57.5%	71/636	11.2%
302a	130/445	29.2%	161/445	36.2%
302a*	281/329	85.4%	13/329	4%
378	83/203	40.9%	21/203	10.3%
	Average	46.4%		24.4%
NSCs				
9	398/1224	32.5%	239/1224	19.5%
9*	233/849	27.4%	338/849	39.8%
21	119/329	36.1%	19/329	6%
30a*	281/329	85.4%	13/329	4%
101	158/681	23.2%	414/681	60%
140	344/689	50.0%	23/689	33.4%
151-3p	65/180	36.1%	4/180	2.2%
320a	130/916	14.2%	68/916	7.4%
378	83/203	40.9%	21/203	10.3%
	Average	38.4%		20.3%
MSCs				
10a	100/286	35.0%	43/286	15%
16*	121/520	23.3%	42/520	8.1%
17	178/1168	15.2%	224/1168	19.2%
23a	106/944	11.2%	518/944	54.9%
23b	106/944	11.2%	518/944	54.9%
26a	137/720	19.0%	365/720	50.7%
28-3p	89/166	53.6%	3/166	1.8%
29a	93/944	9.9%	357/944	37.8%
31	216/449	48.1%	31/449	6.9%
100	2/42	4.8%	0/42	0%
140-3p	317/557	56.9%	50/557	9.0%
151-5p	5/11	45.0%	3/11	27.3%
191	13/49	26.5%	9/49	18.4%
199b-3p	464/758	61.2%	110/758	14.5%
214	474/881	53.8%	104/881	11.8%
222	172/479	35.9%	36/479	7.5%
337-3p	44/190	23.2%	59/190	31.1%
376a	81/180	45.0%	8/180	4.4%
376b	26/168	15.5%	63/168	37.5%
376c	76/232	32.8%	13/232	5.6%
379*	203/443	45.8%	76/443	17.2%
382	58/184	31.5%	6/184	3.3%

409-3p	132/392	33.7%	35/392	8.9%
455-3p	176/365	48.2%	14/365	3.8%
495	555/1125	49.3%	192/1125	17.1%
let-7a	130/949	13.7%	516/949	54.4%
let-7d	87/906	9.6%	124/906	13.7%
let-7f	130/949	13.7%	516/949	54.4%
	Average	31.2%		21%
		38.7%		21.9%

Table S3.4
Table shows the percentage of predicted targets that is common to both canonical miRNA and isomiRs as well as the percentage of the predicted target that is confined to only the isomiRs. Analysis was performed on the miRNAs and their isomiRs collected from the deep sequencing results of 3 different stem cell lines i.e., human embryonic stem cells (hESCs), neural stem cells (NSCs) and mesenchymal stem cells (MSCs).

No	Gene	mRNA seed target site	microRNA	rhesus	mouse	rat	rabbit	dog	armadillo	elephant	opossum	zebrafish
1	CDH1	ccaaaga (A1+6)	miR-9	√	√	√	-	√	x	x	x	-
2	DNMT3B	accaaaa (A1+6)	isomiR-9	√	√	√	√	√	-	√	x	-
3	HMGA2	accaaaga (A1+7)	miR-9	√	√	√	√	√	√	√	√	√
4	NCAM2	accaaaa (A1+6)	isomiR-9	√	√	√	√	√	x	√	√	-
5	Lefty1	gcactta (A1+6)	miR-302a/ isomiR-302a	√	√	√	x	√	x	√	x	√
6	PTEN	gtgcaata (A1+7)	miR-367/ isomiR-367	√	√	√	√	√	√	√	√	√
7	BTG1	agcactta (A1+7)	miR-302a	√	√	√	√	√	√	√	√	-

Table S3.5
The predicted target sites of miRNA in the 3' UTRs of the listed mRNAs are reasonably conserved. The seed target site sequences in the second column are A1 plus 2 to 7 or 2 to 8 nucleotides. IsomiR-9 and isomiR-302a are 5' isomiRs and isomiR-367 is a 3' isomiR. CDH1 and HMGA2 are predicted targets of miR-9 but not isomiR-9. DNMT3B and NCAM2 are predicted targets of isomiR-9 but not miR-9. Lefty1 is a predicted target of both miR-302a and isomiR-302a. PTEN is a predicted target of both miR-367 and isomiR-367. BTG1 is a predicted target of miR-302a but not isomiR-302a. √ represents conserved, x represents not conserved and - is not available. UCSC Genome Browser (http://genome.ucsc.edu/) was used to analyse the seed target site conservation and generate the gene map.

Neural Progenitor Stem Cells

hsa-miR-	Sequence	Length	Seq.no.
let-7f	TGAGGTAGTAGATTGTATAGTT	22	3135
let-7a	TGAGGTAGTAGGTTGTATAGTT	22	2798
21i1	TAGCTTATCAGACTGATGTTGAC	23	2452
423	TGAGGGGCAGAGAGCGAGACTTT	23	1512
let-7fi1	TGAGGTAGTAGATTGTATAGT	21	649
9*	ATAAAGCTAGATAACCGAAAGT	22	643
9	TCTTTGGTTATCTAGCTGTATGA	23	587
92b	TATTGCACTCGTCCCGGCCTCC	22	530
let-7ai1	TGAGGTAGTAGGTTGTATAGT	21	525
29a	TAGCACCATCTGAAATCGGTTA	22	511
130a	CAGTGCAATGTTAAAAGGGCAT	22	509
340	TTATAAAGCAATGAGACTGATT	22	487
21	TAGCTTATCAGACTGATGTTGA	22	462
let-7c	TGAGGTAGTAGGTTGTATGGTT	22	399
101i1	TACAGTACTGTGATAACTGAAG	22	334
423i1	TGAGGGGCAGAGAGCGAGACTTTT	24	328
151-3p	CTAGACTGAAGCTCCTTGAGG	21	294
26a	TTCAAGTAATCCAGGATAGGCT	22	289
let-7fi2	TGAGGTAGTAGATTGTATAGTTA	23	261
let-7b	TGAGGTAGTAGGTTGTGTGGTT	22	253
125b	TCCCTGAGACCCTAACTTGTGA	22	246
9*i1	TAAAGCTAGATAACCGAAAGTA	22	216
130b	CAGTGCAATGATGAAAGGGCAT	22	208
423i2	TGAGGGGCAGAGAGCGAGACTT	22	203
210	CTGTGCGTGTGACAGCGGCTGA	22	199
let-7e	TGAGGTAGGAGGTTGTATAGTT	22	198
103	AGCAGCATTGTACAGGGCTATGA	23	196
320a	AAAAGCTGGGTTGAGAGGGCGA	22	190

24i1	TGGCTCAGTTCAGCAGGAACAGT	23	187
103i1	AGCAGCATTGTACAGGGCTAT	21	178
151-3pi1	CTAGACTGAAGCTCCTTGAGGA	22	178
let-7d	AGAGGTAGTAGGTTGCATAGTT	22	156
let-7fi3	TGAGGTAGTAGATTGTATAGTTT	23	150
340i1	TTATAAAGCAATGAGACTGAT	21	149
let-7ai2	TGAGGTAGTAGGTTGTATAGTTA	23	146
let-7g	TGAGGTAGTAGTTTGTACAGTT	22	146
9*i2	ATAAAGCTAGATAACCGAAAGTA	23	145
30di1	TGTAAACATCCCCGACTGGAAGCT	24	140
423i3	TGAGGGGCAGAGAGCGAGACT	21	140
25	CATTGCACTTGTCTCGGTCTGA	22	139
30e*	CTTTCAGTCGGATGTTTACAGC	22	117
9i1	CTTTGGTTATCTAGCTGTATGA	22	116
378i1	ACTGGACTTGGAGTCAGAAGGC	22	114
7i1	TGGAAGACTAGTGATTTTGTTGTT	24	112
21i2	TAGCTTATCAGACTGATGTTGACA	24	112
101i2	TACAGTACTGTGATAACTGAAT	22	112
27b	TTCACAGTGGCTAAGTTCTGC	21	110
140i1	ACCACAGGGTAGAACCACGGAC	22	110
16	TAGCAGCACGTAAATATTGGCG	22	107
181ai1	AACATTCAACGCTGTCGGTGAGTTT	25	104
101i3	GTACAGTACTGTGATAACTGAA	22	102
let-7ai3	TGAGGTAGTAGGTTGTATAGTTT	23	100
92bi1	TATTGCACTCGTCCCGGCCTC	21	96
let-7i	TGAGGTAGTAGTTTGTGCTGTT	22	95
221*i1	ACCTGGCATACAATGTAGATTTCTGT	26	93
100	AACCCGTAGATCCGAACTTGTG	22	91
320ai1	AAAAGCTGGGTTGAGAGGGCGAT	23	89
30e*i1	CTTTCAGTCGGATGTTTACAGT	22	84
30a*	CTTTCAGTCGGATGTTTGCAGC	22	83
130ai1	CAGTGCAATGTTAAAAGGGCAC	22	81
148a	TCAGTGCACTACAGAACTTTGT	22	79
92b*i1	AGGGACGGGACGCGGTGCAGTGT	23	77
744	TGCGGGGCTAGGGCTAACAGCA	22	74
9*i3	TAAAGCTAGATAACCGAAAGT	21	71
30ai1	TGTAAACATCCTCGACTGGAAGCT	24	71
532	CATGCCTTGAGTGTAGGACCGT	22	68
9*i4	TAAAGCTAGATAACCGAAAGTAT	23	66
24	TGGCTCAGTTCAGCAGGAACAG	22	66
93	CAAAGTGCTGTTCGTGCAGGTAG	23	66

99b	CACCCGTAGAACCGACCTTGC	21	63
101	TACAGTACTGTGATAACTGAA	21	63
320ai2	AAAAGCTGGGTTGAGAGGGCGAA	23	63
92b*i2	AGGGACGGGACGCGGTGCAGTGTT	24	60
151-3pi2	CTAGACTGAAGCTCCTTGAG	20	58
146bi1	TGAGAACTGAATTCCATAGGCTGT	24	57
99bi1	CACCCGTAGAACCGACCTTGCG	22	54
151-3pi3	CTAGACTGAAGCTCCTTGAGGT	22	54
185	TGGAGAGAAAGGCAGTTCCTGA	22	54
9*i5	TAAAGCTAGATAACCGAAAGTAA	23	53
101i4	TACAGTACTGTGATAACTGAAA	22	53
335i1	TCAAGAGCAATAACGAAAAATG	22	51
let-7bi1	TGAGGTAGTAGGTTGTGTGGT	21	50
9i2	TCTTTGGTTATCTAGCTGTATG	22	49
21i3	TAGCTTATCAGACTGATGTTGACT	24	49
191	CAACGGAATCCCAAAAGCAGCTG	23	49
100i1	AACCCGTAGATCCGAACTTGT	21	45
let-7ei1	TGAGGTAGGAGGTTGTATAGT	21	44
16i1	TAGCAGCACGTAAATATTGGCGT	23	43
151-3pi4	CTAGACTGAAGCTCCTTGAGGAA	23	43
423i4	TGAGGGGCAGAGAGCGAGAC	20	42
17	CAAAGTGCTTACAGTGCAGGTAG	23	41
128	TCACAGTGAACCGGTCTCTTT	21	40
191i1	CAACGGAATCCCAAAAGCAGCTGT	24	40
744i1	TGCGGGGCTAGGGCTAACAGC	21	40
let-7gi1	TGAGGTAGTAGTTTGTACAGT	21	40
28-3p	CACTAGATTGTGAGCTCCTGGA	22	39
30ai2	TGTAAACATCCTCGACTGGAAGC	23	39
378i2	ACTGGACTTGGAGTCAGAAGGCA	23	39
423i5	TGAGGGGCAGAGAGCGAGACTTTA	24	38
let-7bi2	TGAGGTAGTAGGTTGTGTGGTTT	23	38
let-7ei2	TGAGGTAGGAGGTTGTATAGTTT	23	38
9i3	TCTTTGGTTATCTAGCTGTATGAA	24	37
151-5p	TCGAGGAGCTCACAGTCTAGT	21	37
378	ACTGGACTTGGAGTCAGAAGG	21	37
let-7fi4	TGAGGTAGTAGATTGTATAG	20	37
30di2	TGTAAACATCCCCGACTGGAAGC	23	36

106bi1	TAAAGTGCTGACAGTGCAGATA	22	36
92bi2	TATTGCACTCGTCCCGGCCTCCT	23	35
106b	TAAAGTGCTGACAGTGCAGAT	21	35
130ai2	CAGTGCAATGTTAAAAGGGCATT	23	35
222i1	AGCTACATCTGGCTACTGGGTCT	23	35
374ai1	TTATAATACAACCTGATAAGT	21	35
let-7ci1	TGAGGTAGTAGGTTGTATGGTTA	23	35
92a	TATTGCACTTGTCCCGGCCTGT	22	34
181ai2	AACATTCAACGCTGTCGGTGAGTT	24	32
340i2	TTATAAAGCAATGAGACTGA	20	32
7	TGGAAGACTAGTGATTTTGTTGT	23	31
146bi2	TGAGAACTGAATTCCATAGGCTGG	24	31
374b	ATATAATACAACCTGCTAAGTG	22	31
423i6	TGAGGGGCAGAGAGCGAGA	19	31
let-7ei3	TGAGGTAGGAGGTTGTATAGTTA	23	31
99a	AACCCGTAGATCCGATCTTGTG	22	30
26ai1	TTCAAGTAATCCAGGATAGGCTAT	24	29
92bi3	TATTGCACTCGTCCCGGCCT	20	29
769	TGAGACCTCTGGGTTCTGAGCT	22	29
let-7ai4	TGAGGTAGTAGGTTGTATAG	20	29
9*i6	ATAAAGCTAGATAACCGAAAGTAT	24	28
26bi1	TTCAAGTAATTCAGGATAGGTT	22	28
92bi4	TATTGCACTCGTCCCGGCCTCA	22	28
378i3	CTGGACTTGGAGTCAGAAGGC	21	27
423i7	TGAGGGGCAGAGAGCGAGACTTTTT	25	27
26ai2	TTCAAGTAATCCAGGATAGGCTA	23	26
27ai1	TTCACAGTGGCTAAGTTCCG	20	25
101i5	GTACAGTACTGTGATAACTGAAA	23	25
130ai3	CAGTGCAATGTTAAAAGGGCA	21	25
151-3pi5	CTAGACTGAAGCTCCTTGAGT	21	25
335i2	TCAAGAGCAATAACGAAAAAT	21	25
335	TCAAGAGCAATAACGAAAAATGT	23	25
29ai1	TAGCACCATCTGAAATCGGTT	21	24
140i2	TACCACAGGGTAGAACCACGGAC	23	24
let-7bi3	TGAGGTAGTAGGTTGTGTGGTTA	23	24
26ai3	TTCAAGTAATCCAGGATAGGCTT	23	23
30a*i1	CTTTCAGTCGGATGTTTGCAGT	22	23
92bi5	TATTGCACTCGTCCCGGCCTCT	22	23
124i1	TAAGGCACGCGGTGAATGCCAA	22	23
let-	TGAGGTAGTAGGTTGTATGGT	21	23

7ci2			
92bi6	TATTGCACTCGTCCCGGCCTCCAT	24	22
92bi7	TATTGCACTCGTCCCGGCCTCCTAT	25	22
101i6	GTACAGTACTGTGATAACTGA	21	22
101i7	TACAGTACTGTGATAACTGAAGT	23	22
221*i2	ACCTGGCATACAATGTAGATTTCT	24	22
21i4	AGCTTATCAGACTGATGTTGAC	22	21
23ai1	ATCACATTGCCAGGGATTTCCA	22	21
378i4	ACTGGACTTGGAGTCAGAAGGCAT	24	21
379	TGGTAGACTATGGAACGTAGG	21	21
21i5	TAGCTTATCAGACTGATGTTGAA	23	20
27bi1	TTCACAGTGGCTAAGTTCTG	20	20
130bi1	CAGTGCAATGATGAAAGGGCATT	23	20
135b	TATGGCTTTTCATTCCTATGTGA	23	20
151-5pi1	TCGAGGAGCTCACAGTCTAGTA	22	20
320ai3	AAAAGCTGGGTTGAGAGGGCGAAT	24	20
9*i7	ATAAAGCTAGATAACCGAAAG	21	19
let-7fi5	TGAGGTAGTAGATTGTAT	18	19
92b*i3	AGGGACGGGACGCGGTGCAGT	21	18
92b*i4	AGGGACGGGACGCGGTGCAGTGTTT	25	18
98	TGAGGTAGTAAGTTGTATTGTT	22	18
103i2	AGCAGCATTGTACAGGGCTATGAT	24	18
130ai4	CAGTGCAATGTTAAAAGGGCATA	23	18
320ai4	AAAAGCTGGGTTGAGAGGGCGT	22	18
1260	ATCCCACCGCTGCCACCA	18	18
let-7ci3	TGAGGTAGTAGGTTGTATGGTTT	23	18
92bi8	TATTGCACTCGTCCCGGCCTCCA	23	17
92bi9	TATTGCACTCGTCCCGGCCTCCATC	25	17
103i3	AGCAGCATTGTACAGGGCTATG	22	17
125bi1	TCCCTGAGACCCTAACTTGTG	21	17
181a	AACATTCAACGCTGTCGGTGAGT	23	17
195i1	TAGCAGCACAGAAATATTGGCA	22	17
let-7fi6	TGAGGTAGTAGATTGTATAGTTATC	25	17
let-7fi7	TGAGGTAGTAGATTGTATA	19	17
100i2	AACCCGTAGATCCGAACTTGTGA	23	16
423i8	TGAGGGGCAGAGAGCGAGACTTA	23	16
542-3p	TGTGACAGATTGATAACTGAAA	22	16
let-7ii1	TGAGGTAGTAGTTTGTGCTGT	21	16
let-7ii2	TGAGGTAGTAGTTTGTGCTGTTT	23	16

9i3	CTTTGGTTATCTAGCTGTATGAA	23	15
29c	TAGCACCATTTGAAATCGGTTA	22	15
let-7di1	AGAGGTAGTAGGTTGCATAGTTT	23	15
9*i8	ATAAAGCTAGATAACCGAAAGTAA	24	14
10ai1	TACCCTGTAGATCCGAATTTGT	22	14
25*i1	AGGCGGAGACTTGGGCAATTGCT	23	14
26ai4	TTCAAGTAATCCAGGATAGGC	21	14
30c-2*	CTGGGAGAAGGCTGTTTACTCT	22	14
30e*i2	CTTTCAGTCGGATGTTTACAG	21	14
99ai1	AACCCGTAGATCCGATCTTGT	21	14
151-3pi6	CTAGACTGAAGCTCCTTGAGGAT	23	14
221*i3	ACCTGGCATACAATGTAGATTTCTG	25	14
9i4	TCTTTGGTTATCTAGCTGTAT	21	13
9i5	TCTTTGGTTATCTAGC	16	13
23a	ATCACATTGCCAGGGATTTCC	21	13
23b	ATCACATTGCCAGGGATTACCACT	24	13
27a	TTCACAGTGGCTAAGTTCCGC	21	13
28-3pi1	CACTAGATTGTGAGCTCCTGGAA	23	13
30a*i2	TTTCAGTCGGATGTTTGCAGC	21	13
101i8	TACAGTACTGTGATAACTGAAGA	23	13
130ai5	CAGTGCAATGTTAAAAGGGC	20	13
151-3pi7	CTAGACTGAAGCTCCTTGAGA	21	13
151-3pi8	TACTAGACTGAAGCTCCTTGAGG	23	13
191i2	CAACGGAATCCCAAAAGCAGCT	22	13
320ai5	AAAGCTGGGTTGAGAGGGCGA	21	13
378i5	CTGGACTTGGAGTCAGAAGGCT	22	13
423i9	TGAGGGGCAGAGAGCGAGACA	21	13
15a	TAGCAGCACATAATGGTTTGTG	22	12
15ai1	TAGCAGCACATAATGGTTTGT	21	12
23ai2	ATCACATTGCCAGGGATTTCCAA	23	12
92bi10	TATTGCACTCGTCCCGGCC	19	12
130bi2	CAGTGCAATGATGAAAGGGCA	21	12
140i3	CCACAGGGTAGAACCACGGAC	21	12
148b	TCAGTGCATCACAGAACTTTGT	22	12
151-3pi9	TACTAGACTGAAGCTCCTTGAG	22	12
191i3	CAACGGAATCCCAAAAGCAGCTGA	24	12
450bi1	TTTTGCAATATGTTCCTGAAT	21	12
744i2	TGCGGGGCTAGGGCTAACAGCAA	23	12
1275i1	GTGGGGGAGAGGCTGT	16	12
let-	TGAGGTAGTAGGTTGTATAGTTAA	24	12

7ai5			
let-7di2	AGAGGTAGTAGGTTGCATAGT	21	12
let-7fi8	TGAGGTAGTAGATTGTATAGTTC	23	12
19b	TGTGCAAATCCATGCAAAACTGA	23	11
21i6	TTATCAGACTGATGTTGACTAGC	23	11
29ai2	TAGCACCATCTGAAATCGGTTAT	23	11
30di3	TGTAAACATCCCCGACTGGAAGCTT	25	11
106bi2	TAAAGTGCTGACAGTGCAGA	20	11
138	AGCTGGTGTTGTGAATCAGGCCG	23	11
140i4	ACCACAGGGTAGAACCACGGA	21	11
151-3pi10	CTAGACTGAAGCTCCTTGA	19	11
186i1	CAAAGAATTCTCCTTTTGGGCTT	23	11
378i6	ACTGGACTTGGAGTCAGAAG	20	11
423i10	TGAGGGGCAGAGAGCGAGAAA	21	11
let-7ai6	TGAGGTAGTAGGTTGTAT	18	11
let-7fi9	TGAGGTAGTAGATTGTATAGTTG	23	11
9i6	TTTGGTTATCTAGCTGTATGA	21	10
9*i9	ATAAAGCTAGATAACCGAAA	20	10
20a	TAAAGTGCTTATAGTGCAGGTAG	23	10
24i2	TGGCTCAGTTCAGCAGGAAC	20	10
25*i2	AGGCGGAGACTTGGGCAATTGC	22	10
27bi2	TTCACAGTGGCTAAGTTCTGCA	22	10
30d	TGTAAACATCCCCGACTGGAAG	22	10
100i3	AACCCGTAGATCCGAAC	17	10
100i4	AACCCGTAGATCCGAACTTGTGT	23	10
125bi2	TCCCTGAGACCCTAAC	16	10
130bi3	CAGTGCAATGATGAAAGGGCATA	23	10
151-3pi11	CTAGACTGAAGCTCCTTGAGGAAA	24	10
181bi1	AACATTCATTGCTGTCGGTGGGTT	24	10
let-7ai7	TGAGGTAGTAGGGTGTATAGTT	22	10
let-7ii3	TGAGGTAGTAGTTTGTGCTGTTA	23	10

Table S3.6A
Table lists the miRNAs of neural stem cells based on their sequencing number from highest to lowest. IsomiR-9 (22 nts) ranked number 42 and was sequenced higher than some of the canonical miRNAs. Yellow highlighted miRNAs denote canonical miRNAs.

Human Embryonic Stem Cells

hsa-miR-	Sequence	Length	Seq.no.
302a*i1	TAAACGTGGATGTACTTGCTTT	22	1113
130a	CAGTGCAATGTTAAAAGGGCAT	22	888
423	TGAGGGGCAGAGAGCGAGACTTT	23	529
302a*i2	TAAACGTGGATGTACTTGCTT	21	491
182	TTTGGCAATGGTAGAACTCACACT	24	432
103	AGCAGCATTGTACAGGGCTATGA	23	393
340	TTATAAAGCAATGAGACTGATT	22	381
21i1	TAGCTTATCAGACTGATGTTGAC	23	324
151-3p	CTAGACTGAAGCTCCTTGAGG	21	290
130b	CAGTGCAATGATGAAAGGGCAT	22	257
320a	AAAAGCTGGGTTGAGAGGGCGA	22	221
378i1	ACTGGACTTGGAGTCAGAAGGC	22	214
182i1	TTTGGCAATGGTAGAACTCACACTGG	26	166
101i1	TACAGTACTGTGATAACTGAAG	22	165
151-3pi1	CTAGACTGAAGCTCCTTGAGGA	22	162
182i2	TTTGGCAATGGTAGAACTCACAC	23	148
183i1	ATGGCACTGGTAGAATTCACTG	22	133
302a*i3	CTTAAACGTGGATGTACTTGCTT	23	132
148a	TCAGTGCACTACAGAACTTTGT	22	110
302a*i4	CTTAAACGTGGATGTACTTGCT	22	109
103i1	AGCAGCATTGTACAGGGCTAT	21	104
25	CATTGCACTTGTCTCGGTCTGA	22	98
302a*	ACTTAAACGTGGATGTACTTGCT	23	96
182i3	TTTGGCAATGGTAGAACTCACACTG	25	95
302a*i5	TAAACGTGGATGTACTTGCTTTGA	24	86
378i2	ACTGGACTTGGAGTCAGAAGGCA	23	86
423i1	TGAGGGGCAGAGAGCGAGACTTTT	24	85
30e*i1	CTTTCAGTCGGATGTTTACAGT	22	82
93	CAAAGTGCTGTTCGTGCAGGTAG	23	81
21	TAGCTTATCAGACTGATGTTGA	22	80
101i2	GTACAGTACTGTGATAACTGAA	22	80
302a*i6	ACTTAAACGTGGATGTACTTGC	22	80
191	CAACGGAATCCCAAAAGCAGCTG	23	76
221*i1	ACCTGGCATACAATGTAGATTTCTGT	26	72
340i1	TTATAAAGCAATGAGACTGAT	21	72
183i2	ATGGCACTGGTAGAATTCACT	21	69
302a*i7	TAAACGTGGATGTACTTGCT	20	67
183	TATGGCACTGGTAGAATTCACT	22	66
335i1	TCAAGAGCAATAACGAAAAATG	22	65
151-3pi2	CTAGACTGAAGCTCCTTGAGGAA	23	64
302d	TAAGTGCTTCCATGTTTGAGTGT	23	63

140i1	ACCACAGGGTAGAACCACGGAC	22	62
320ai1	AAAAGCTGGGTTGAGAGGGCGAA	23	62
302b	TAAGTGCTTCCATGTTTTAGTAG	23	60
335i2	TCAAGAGCAATAACGAAAAAT	21	58
30e*	CTTTCAGTCGGATGTTTACAGC	22	55
302a	TAAGTGCTTCCATGTTTTGGTGA	23	54
320ai2	AAAAGCTGGGTTGAGAGGGCGAT	23	54
1323	TCAAAACTGAGGGGCATTTTCT	22	53
130ai1	CAGTGCAATGTTAAAAGGGCATA	23	52
302ci1	AAGTGCTTCCATGTTTCAGTGGT	23	51
30a*	CTTTCAGTCGGATGTTTGCAGC	22	50
151-3pi3	CTAGACTGAAGCTCCTTGAG	20	49
378	ACTGGACTTGGAGTCAGAAGG	21	47
183i3	TATGGCACTGGTAGAATTCACTG	23	45
532	CATGCCTTGAGTGTAGGACCGT	22	42
17	CAAAGTGCTTACAGTGCAGGTAG	23	39
26a	TTCAAGTAATCCAGGATAGGCT	22	38
30a*i1	CTTTCAGTCGGATGTTTGCAGT	22	38
106bi1	TAAAGTGCTGACAGTGCAGATA	22	38
423i2	TGAGGGGCAGAGAGCGAGACTT	22	38
28-3p	CACTAGATTGTGAGCTCCTGGA	22	37
367	AATTGCACTTTAGCAATGGTGA	22	34
20b	CAAAGTGCTCATAGTGCAGGTAG	23	33
106b	TAAAGTGCTGACAGTGCAGAT	21	33
335	TCAAGAGCAATAACGAAAAATGT	23	33
101i3	GTACAGTACTGTGATAACTGAAA	23	32
378i3	CTGGACTTGGAGTCAGAAGGC	21	32
31i1	AGGCAAGATGCTGGCATAGCTG	22	31
423i3	TGAGGGGCAGAGAGCGAGACT	21	31
29a	TAGCACCATCTGAAATCGGTTA	22	30
101	TACAGTACTGTGATAACTGAA	21	30
101i4	TACAGTACTGTGATAACTGAAT	22	29
185	TGGAGAGAAAGGCAGTTCCTGA	22	29
31	AGGCAAGATGCTGGCATAGCT	21	27
1270	CTGGAGATATGGAAGAGCTGTGT	23	27
103i2	AGCAGCATTGTACAGGGCTATG	22	26
130bi1	CAGTGCAATGATGAAAGGGCATT	23	26
302bi1	TAAGTGCTTCCATGTTTTAGTA	22	26
99bi1	CACCCGTAGAACCGACCTTGC	21	25
124i1	TAAGGCACGCGGTGAATGCCAA	22	25
141i1	TAACACTGTCTGGTAAAGATGGC	23	25
151-3pi4	CTAGACTGAAGCTCCTTGAGGT	22	25
191i1	CAACGGAATCCCAAAAGCAGCTGT	24	25
302a*i8	TAAACGTGGATGTACTTGCTTTG	23	24

320ai3	AAAAGCTGGGTTGAGAGGGCGT	22	24
374ai1	TTATAATACAACCTGATAAGT	21	24
374b	ATATAATACAACCTGCTAAGTG	22	24
25*i1	AGGCGGAGACTTGGGCAATTGCT	23	23
101i5	GTACAGTACTGTGATAACTGA	21	23
191i2	CAACGGAATCCCAAAAGCAGCT	22	23
302a*i9	TAAACGTGGATGTACTTGCTTTGAAACT	28	23
708	AAGGAGCTTACAATCTAGCTGGG	23	23
30ai1	TGTAAACATCCTCGACTGGAAGC	23	22
30ai2	TGTAAACATCCTCGACTGGAAGCT	24	22
30di1	TGTAAACATCCCCGACTGGAAGCT	24	22
99b	CACCCGTAGAACCGACCTTGCG	22	22
302a*i10	TAAACGTGGATGTACTTGCTTTA	23	22
222i1	AGCTACATCTGGCTACTGGGTCT	23	21
378i4	ACTGGACTTGGAGTCAGAAGGCAT	24	21
744	TGCGGGGCTAGGGCTAACAGCA	22	21
101i6	TACAGTACTGTGATAACTGAAA	22	20
130ai2	CAGTGCAATGTTAAAAGGGCATT	23	20
182i4	TTTGGCAATGGTAGAACTCACACTGGT	27	20
183i4	ATGGCACTGGTAGAATTCACTGT	23	20
219-2-3p	AGAATTGTGGCTGGACATCTGT	22	20
302a*i11	TAAACGTGGATGTACTTGCTTTGAAAC	27	20
let-7a	TGAGGTAGTAGGTTGTATAGTT	22	20
16	TAGCAGCACGTAAATATTGGCG	22	19
17i1	CAAAGTGCTTACAGTGCAGGT	21	19
31i2	AGGCAAGATGCTGGCATAGCTGT	23	19
92b	TATTGCACTCGTCCCGGCCTCC	22	19
378i5	ACTGGACTTGGAGTCAGAAG	20	19
423i4	TGAGGGGCAGAGAGCGAGACTTTA	24	19
92a	TATTGCACTTGTCCCGGCCTGT	22	18
130bi2	CAGTGCAATGATGAAAGGGCATA	23	18
151-3pi5	CTAGACTGAAGCTCCTTGAGGAAA	24	18
302ai1	TAAGTGCTTCCATGTTTTGGTG	22	18
302d*i1	ACTTTAACATGGAGGCACTTGCT	23	18
221*i2	ACCTGGCATACAATGTAGATTTCT	24	17
21i2	TAGCTTATCAGACTGATGTTGACA	24	16
30c-2*	CTGGGAGAAGGCTGTTTACTCT	22	16
148ai1	TCAGTGCACTACAGAACTTTGTC	23	16
302ai2	AAGTGCTTCCATGTTTTGGTGA	22	16
302a*i12	TAAACGTGGATGGACTTGCTTT	22	16
130bi3	CAGTGCAATGATGAAAGGGCA	21	15
193a	AACTGGCCTACAAAGTCCCAGT	22	15
221	AGCTACATTGTCTGCTGGGTTTC	23	15
302a*i13	TTAAACGTGGATGTACTTGCTT	22	15

302d*	ACTTTAACATGGAGGCACTTGC	22	15
340i2	TTATAAAGCAATGAGACTGA	20	15
30e*i2	CTTTCAGTCGGATGTTTACAG	21	14
130ai3	CAGTGCAATGTTAAAAGGGC	20	14
28-5pi1	AAGGAGCTCACAGTCTATTGA	21	13
151-3pi6	CTAGACTGAAGCTCCTTGAGA	21	13
182i5	TTTGGCAATGGTAGAACTCACA	22	13
302d*i2	ACTTTAACATGGAGGCACTTG	21	13
503i1	TAGCAGCGGGAACAGTTCTGAAA	23	13
1275i1	GTGGGGGAGAGGCTGT	16	13
92b*i1	AGGGACGGGACGCGGTGCAGTGTT	24	12
128	TCACAGTGAACCGGTCTCTTT	21	12
130ai4	CAGTGCAATGTTAAAAGGGCA	21	12
151	TCGAGGAGCTCACAGTCTAGT	21	12
182i6	TTTGGCAATGGTAGAACTCACACTGGA	27	12
302a*i14	TAAACGTGGATGTACTTGCTTTGAA	25	12
302bi2	TAAGTGCTTCCATGTTTTAGT	21	12
302bi3	TAAGTGCTTCCATGTTTTAGTAGT	24	12
302di1	AAGTGCTTCCATGTTTGAGTGT	22	12
363	AATTGCACGGTATCCATCTGTA	22	12
378i6	ACTGGACTTGGAGTCAGAAGGCAA	24	12
16i1	TAGCAGCACGTAAATATTGGCGT	23	11
24	TGGCTCAGTTCAGCAGGAACAG	22	11
103i3	AGCAGCATTGTACAGGGCTATGAT	24	11
182i7	TTTGGCAATGGTAGAACTCACACTGT	26	11
183i5	ATGGCACTGGTAGAATTCACTGA	23	11
191i3	CAACGGAATCCCAAAAGCAGCTGA	24	11
302ci2	AAGTGCTTCCATGTTTCAGTGG	22	11
367i1	AATTGCACTTTAGCAATGGT	20	11
378i7	CTGGACTTGGAGTCAGAAGGCA	22	11
548fi1	AAAACTGTAATTACTTTTGGAC	22	11
708*i1	AACTAGACTGTGAGCTTCTAGA	22	11
769	TGAGACCTCTGGGTTCTGAGCT	22	11
19b	TGTGCAAATCCATGCAAAACTGA	23	10
21i3	TAGCTTATCAGACTGATGTTGAA	23	10
25*i2	AGGCGGAGACTTGGGCAATTGC	22	10
28-3pi1	CACTAGATTGTGAGCTCCTGGAA	23	10
92b*i2	AGGGACGGGACGCGGTGCAGTGT	23	10
106b*	CCGCACTGTGGGTACTTGCTGC	22	10
107	AGCAGCATTGTACAGGGCTATCA	23	10
210	CTGTGCGTGTGACAGCGGCTGA	22	10
221*i3	ACCTGGCATACAATGTAGATTTC	23	10
320ai4	AAAAGCTGGGTTGAGAGGGCGAAT	24	10
1260	ATCCCACCGCTGCCACCA	18	10

Table S3.6B
Table lists the miRNAs of embryonic stem cells based on their sequencing number from highest to lowest. IsomiR-302a (22 nts) ranked number 118 and was sequenced higher than some of the canonical miRNAs. Yellow highlighted miRNAs denote canonical miRNAs.

No	Predicted targets unique to isomiR-9	Note
1	ADRA2C	Adrenal receptor
2	AVPR1A	Restricted, not brain
3	BACE2	Ubiquitous
4	C20ORF77	/
5	CACNG3	Brain/ Kidney tumour
6	CGI-09	/
7	CPT1A	Widespread
8	CSN3	Connective tissue
9	CYP2U1	/
10	DDX50	/
11	GALNACT-2	Heart
12	GFRA2	/
13	GJA3	Kidney tumour
14	GPR177	/
15	IFRD2	Ubiquitous
16	IQCH	Restricted
17	IRX3	/
18	KIAA0523	Glioma
19	LOC133308	/
20	LOC374395	/
21	LYSMD2	/
22	MGC4655	Widespread
23	N4BP3	Restricted, include brain
24	NDST3	Leukaemia
25	PPP2R2D	Widespread
26	PWWP2B	Widespread
27	RAB31	Widespread
28	RFT1	Head and neck tumour
29	RUTBC3	Widespread
30	SEP15/	Widespread
31	SF3A3	Splicing factor/ Ubiquitous
32	SLC25A26	Widespread
33	SMPD4	/
34	SNX9	Widespread
35	SSU72	Ubiquitous
36	SYNE2	Widespread
37	TCF7L2	Widespread
38	THAP4	Widespread

39	UBE2J2	Widespread
40	UBQLN3	Brain/ testis/ germ cell tumour
41	VCP	Ubiquitous
42	XLKD1	Widespread
43	ZCCHC8	Widespread
44	ZNF364	Widespread

Table S3.7 Table below lists all the unique targets of isomiR-9.

miR-	ES	NS	MS
9	–	870	–
9*	–	1299	–
10a	–	14	35
10a*	–	–	2
16	30	150	112
16*	–	–	9
20a	–	10	3
20a*	–	–	1
21	430	3127	880
21*	–	–	3
22	–	–	429
22*	–	–	6
23b	–	13	75
23b*	–	–	1
25	98	139	3
25*	33	24	–
27b	–	140	3
27b*	–	–	1
28-5p	–	–	41
28-3p	47	52	32
29a	30	546	179
29a*	–	–	4
30a	44	110	27
30a*	88	119	20
30b	–	–	1
30b*	–	–	2
30e	–	215	5
30e*	151	–	8
92a	18	34	–
92a-1*	6	–	–

miR-	ES	NS	MS
92b	25	885	–
92b*	22	188	–
106b	71	88	18
106b*	15	–	–
125a-5p	–	–	3
125a-3p	–	–	3
125b	–	273	97
125b*	–	–	12
140-5p	–	–	3
140-3p	62	157	9
151-5p	17	57	2686
151-3p	663	761	25
154	–	–	1
154*	–	–	1
186	–	11	4
186*	–	–	1
193a-5p	–	–	2
193a-3p	15	–	107
199b-5p	–	–	3
199-3p	–	–	20
221	15	–	36
221*	99	129	1
222	21	35	50
222*	–	–	1
302a	88	–	–
302a*	2306	–	–
302b	115	–	–
302b*	5	–	–
302d	75	–	–
302d*	46	–	–

miR-	ES	NS	MS
335	156	101	8
335*	–	–	10
379	–	21	6
379*	–	–	16
411	–	–	1
411*	–	–	26
423-5p	702	2361	10
423-3p	–	–	5
424	–	–	72
424*	–	–	3
425	–	–	8
425*	–	–	3
493	–	–	–
493*	–	–	119
501-5p	–	–	1
501-3p	–	–	1
654-5p	–	–	5
654-3p	–	–	4
708	23	–	2
708*	11	–	–
769-5p	11	29	–
769-3p	–	–	1
let-7d	–	183	1323
let-7d*	–	–	1
let-7i	–	137	40
let-7i*	–	–	1

Table S5.1
Table lists the total number of sequencing results in hESCs, NSCs and MSCs. Deep sequencing was performed by Elcie Chan.

1. Common elements in "302a/b/c/d", "302a*", "302b*/d*" and "367":
BCL11A
ZFHX4

2. Common elements in "302a/b/c/d", "302a*" and "367":
FNDC3B
PPP1R9A
ZNF148

3. Common elements in "302a/b/c/d", "302a*" and "302b*/d*":
INTS6
KLHL28
MYBL1
PLEKHA3
PURA
PURB
TRPS1

4. Common elements in "302a/b/c/d" and "302a*":
ATP2B2
BAHD1
BCL11B
C11orf30
C16orf72
CNOT6
E2F7
ELAVL2
ESR1
FAM13C1
FBXO11
KREMEN1
MBNL1
NPAS3
NR2F2
PCDHA1
PCDHA10
PCDHA12
PCDHA13
PCDHA2
PCDHA3
PCDHA4
PCDHA5
PCDHA6
PCDHA7
PCDHA8

PCDHAC1
PCDHAC2
PLAG1
RAB7A
RORB
SP3
TARDBP
TFAP4
YTHDF3
ZNF436

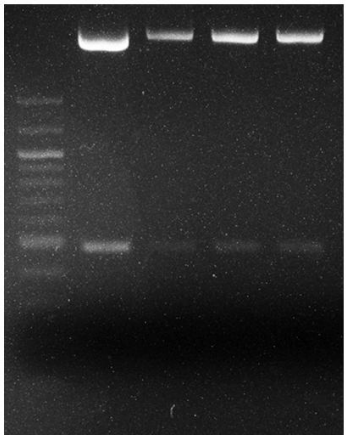
5. Common elements in "302a*", "302b*/d*" and "367":
APRIN
BMPR2
C18orf25
GRM7
LUZP1
NOVA1
STAG2
TNPO1
TOB1

6. Common elements in "302a*" and "302b*/d*":
BMP2
BRD1
CREBBP
CROP
CUL4B
DAZAP1
DOLPP1
ETV6
GNAZ
GPM6B
INTS2
RYBP
SAMD4B
SOX6
TAOK3
ZIC1
ZMIZ1

Table S5.2
List of predicted targets of miR-302 cluster that are common between members.

pGEM-T-DNMT3B mutant UTR digested by NheI and FseI

100 C1 C2 C3 C4

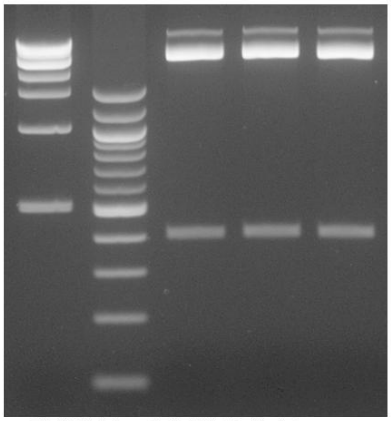


PG-DNMT3B Mutant 3'UTR Clone 2

```
00000055 gcagagccacctgactcttgcaggggtagcctgaggtgccgcctccttgt 00000104
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859514 gcagagccacctgactcttgcaggggtagcctgaggtgccgcctccttgt 30859563

00000105 gcacaaatcagacctggctgcttggagcagcctaacacggtgctcatttt 00000154
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859564 gcacaaatcagacctggctgcttggagcagcctaacacggtgctcatttt 30859613

00000155 ttcttctcctaaaactttaaaacttgaagtaggtagcaacgtggctttttt 00000204
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859614 ttcttctcctaaaactttaaaacttgaagtaggtagcaacgtggctttttt 30859663

00000205 ttttttcccttcctgggtctaccactcagagaaaacaatggccaagatgca 00000254
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859664 ttttttcccttcctgggtctaccactcagagaaaacaatggctaagatacc 30859713

00000255 accaccacagtgccgacagctctccaatactcaggttaatgctgaaaaat 00000304
>>>>>>>> |  |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859714 aaaaccacagtgccgacagctctccaatactcaggttaatgctgaaaaat 30859763

00000305 catccaagacagttattgcaagagtttaatttttgaaaactggctactgc 00000354
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859764 catccaagacagttattgcaagagtttaatttttgaaaactggctactgc 30859813

00000355 tctgtgtttacagacgtgtgcagttgtaggcatgtagctacaggacattt 00000404
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859814 tctgtgtttacagacgtgtgcagttgtaggcatgtagctacaggacattt 30859863

00000405 ttaagggcccaggatcgtttttttcccagggcaagcagaagagaaaatgtt 00000454
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859864 ttaagggcccaggatcgtttttttcccagggcaagcagaagagaaaatgtt 30859913

00000455 gtatatgtctttttacccggcacattccccttgcctaaatacaagggctgg 00000504
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859914 gtatatgtctttttacccggcacattccccttgcctaaatacaagggctgg 30859963

00000505 agtctgcacgggacctattag 00000525
>>>>>>>> ||||||||||||||||||| >>>>>>>>
30859964 agtctgcacgggacctattag 30859984
```

pGEM-T-Lefty1 3'UTR digested by NheI and FseI

1kb 100 C1 C2 C3

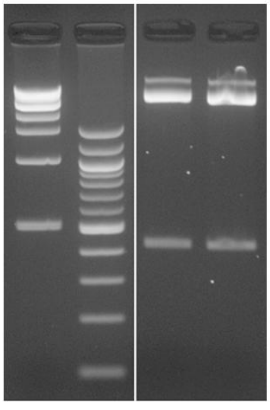


PG-lefty1-3UTR C1

```
000000109 ctggattggggatgcacaacccacacttaaccccaggtcttaggtccaga 000000158
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222380762 ctggattggggatgcacaacccacacttaaccccaggtcttaggtccaga 222380811

000000159 gtggtggtggaggactttagccccagatccagtgacaggacaagtaaacaa 000000208
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222380812 gtggtggtggaggactttagccccagatccagtgacaggacaagtaaacaa 222380861

000000209 tgacacattgggcttttctgccctcaggttacagtatctccacatgtaagt 000000258
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222380862 tgacacattgggcttttctgccctcaggttacagtatctccacatgtaagt 222380911

000000259 gcttagaatatagtgcagtgaataataagaatagagaaaactgagcaagg 000000308
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222380912 gcttagaatatagtgcagtgaataataagaatagagaaaactgagcaagg 222380961

000000309 gctctccagtggccaaagattctcattcctgagaagcaaaaattaggtga 000000358
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222380962 gctctccagtggccaaagattctcattcctgagaagcaaaaattaggtga 222381011

000000359 ggtaacttgtcagaggaagcaaattcagggctcactagagagcacagagc 000000408
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222381012 ggtaacttgtcagaggaagcaaattcagggctcactagagagcacagagc 222381061

000000409 atttgtccatcagcagttcagtcatcgccagctctcctggtaccctcgaa 000000458
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>>
222381062 atttgtccatcagcagttcagtcatcgccagctctcctggtaccctcgaa 222381111

000000459 cacttcagaaacacacacaagtcaagtccctcgatggctacgctag 000000504
>>>>>>>>> |||||||||||||||||||||||||||||||||||||||||| |||| >>>>>>>>>
222381112 cacttcagaaacacacacaagtcaagtccctcgatggctacactag 222381157
```

pGEM-T –PTEN 3'UTR digested by NheI and FseI

1kb 100 C1 C2

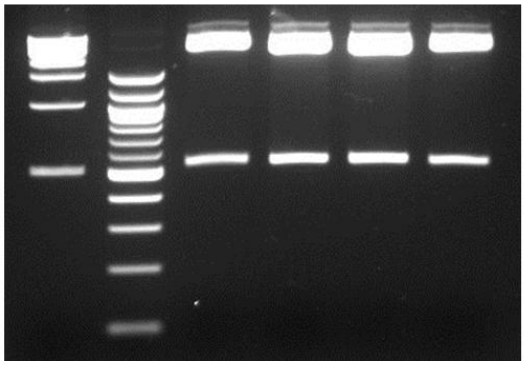


PG-PTEN-3UTR Clone 2

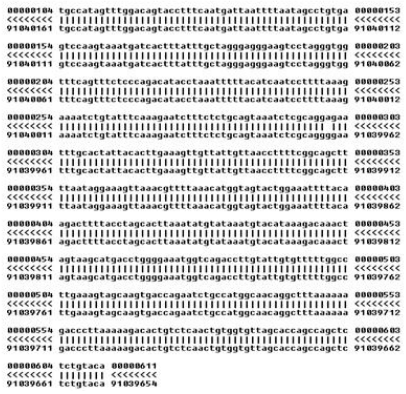
```
00000078 gtagggtacaagtttaatgtttagttctagaaatttttgtgcaatatgttc 00000127
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718031 gtagggtacaagtttaatgtttagttctagaaatttttgtgcaatatgttc 89718080

00000128 ataacgatggctgtggttgccacaaagtgcctcgtttaccttaaatact 00000177
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718081 ataacgatggctgtggttgccacaaagtgcctcgtttaccttaaatact 89718130

00000178 gttaatgtgtcatgcatgcagatggaaggggtggaactgtgcactaaagt 00000227
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718131 gttaatgtgtcatgcatgcagatggaaggggtggaactgtgcactaaagt 89718180

00000228 gggggctttaactgtagtatttggcagagttgccttctacctgccagttc 00000277
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718181 gggggctttaactgtagtatttggcagagttgccttctacctgccagttc 89718230

00000278 aaaagttcaacctgttttcatatagaatatatatactaaaaaatttcagt 00000327
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718231 aaaagttcaacctgttttcatatagaatatatatactaaaaaatttcagt 89718280

00000328 ctgttaaacagccttaattccgattcagcctcttcagatactcttgtgctg 00000377
>>>>>>>> |||||||||||||||||| |||||||||||||||||||||||||||||| >>>>>>>>
89718281 ctgttaaacagccttaatcctgattcagcctcttcagatactcttgtgctg 89718330

00000378 tgcagcagtggctctgtgtgtaaatgctatgcactgaggatacacaaaaa 00000427
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718331 tgcagcagtggctctgtgtgtaaatgctatgcactgaggatacacaaaaa 89718380

00000428 taccaatatgatgtgtacaggataatgcctcatcccaatcagatgtccat 00000477
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
89718381 taccaatatgatgtgtacaggataatgcctcatcccaatcagatgtccat 89718430

00000478 ttgtta 00000483
>>>>>>>> |||||| >>>>>>>>
89718431 ttgtta 89718436
```

pGEM-T-BTG1-3'UTR
digested by NheI and FseI

1kb 100bp C1 C2 C3 C4

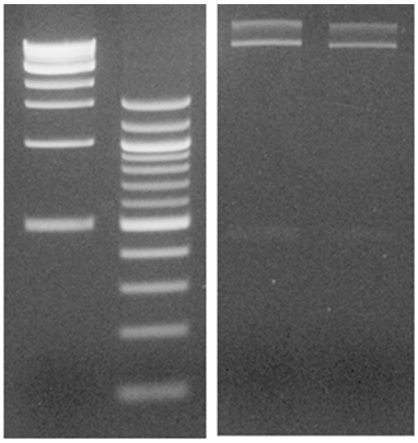


PG-BTG1-3UTR Clone1



pGEM-T-PTEN Mutant 3'UTR
digested by NheI and FseI

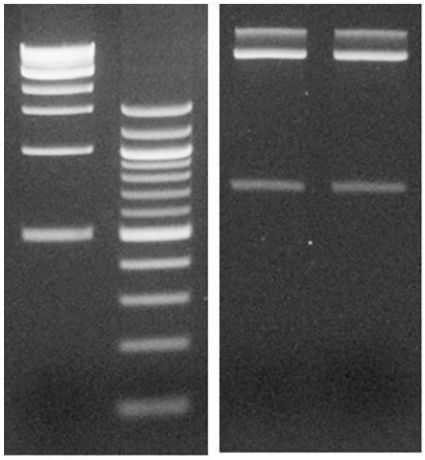
1kb 100 C1 C2



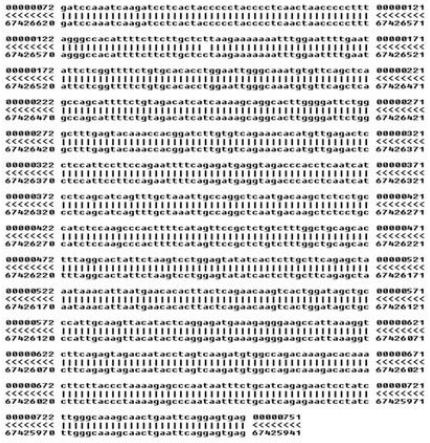
PG-PTEN Mutant 3'UTR Clone 1

pGEM-T-CDH1- 3'UTR
digested by NheI and FseI

1kb 100bp C1 C2

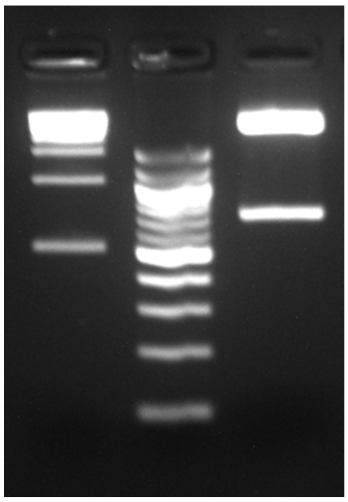


PG-CDH1 3'UTR Clone 1



pGEM-T-CDH1 Mutant-UTR
digested by NheI/Fse I

1kb 100

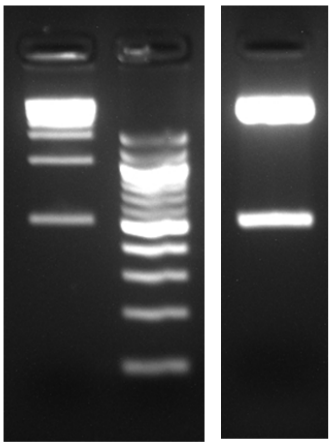


PG-CDH1 Mutant 3'UTR Clone 3

```
00000054 ctcactcctgaattcagttgctttgcccaagataggagttctctgatgca 00000103
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868448 ctcactcctgaattcagttgctttgcccaagataggagttctctgatgca 68868489

00000104 gaaattattgggctcttttagggtaagaagtttgtgtctttgtctggcca 00000153
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868490 gaaattattgggctcttttagggtaagaagtttgtgtctttgtctggcca 68868539

00000154 tatcttgactaggtattgtctactctgaagacctttaatggcttccctct 00000203
>>>>>>>>  |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868540 catcttgactaggtattgtctactctgaagacctttaatggcttccctct 68868589

00000204 ttcatctcctgagtatgtaacttgcaatgggcagctatccagtgacttgt 00000253
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868590 ttcatctcctgagtatgtaacttgcaatgggcagctatccagtgacttgt 68868639

00000254 tctgagtaagtgtgttcattaatgtttatttagctctgaagcaagagtga 00000303
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868640 tctgagtaagtgtgttcattaatgtttatttagctctgaagcaagagtga 68868689

00000304 tacactccaggacttagaaatagtgcctaaagtgctgcagacgtatgcaga 00000353
>>>>>>>> || |||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868690 tatactccaggacttagaatagtgcctaaagtgctgcagccaaagacaga 68868739

00000354 gcggaactatgaaaagtgggcttggagatggcaggagagcttgtcattga 00000403
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868740 gcggaactatgaaaagtgggcttggagatggcaggagagcttgtcattga 68868789

00000404 gcctggcaatttagcaaactgatgctgaggatgattgaggtgggtctacc 00000453
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868790 gcctggcaatttagcaaactgatgctgaggatgattgaggtgggtctacc 68868839

00000454 tcatctctgaaaatttctggaaggaatggagggagtctcaacatgtgtttct 00000503
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868840 tcatctctgaaaatttctggaaggaatggagggagtctcaacatgtgtttct 68868889

00000504 gacacaagatccgtggtttgtactcaaagcccagaatccccaagtgcctg 00000553
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868890 gacacaagatccgtggtttgtactcaaagcccagaatccccaagtgcctg 68868939

00000554 cttttgatgatgtctacagaaaatgctggctgagctgaacacatttgccc 00000603
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868940 cttttgatgatgtctacagaaaatgctggctgagctgaacacatttgccc 68868989

00000604 aattccaggtgtgcacagaaaaccgagaatattcaaaatttccaaatttttt 00000653
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68868990 aattccaggtgtgcacagaaaaccgagaatattcaaaatttccaaatttttt 68869039

00000654 ttcttaagagcaagaagaaaaatgtggcccttaaaggggggttagttgaggg. 00000702
>>>>>>>> |||||| |||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
68869040 ttcttaggagcaagaagaaaaatgtggcccttaaaggggggttagttgagggg 68869089

00000703 taggggggtagtgaggatcttgatttggatc 00000732
>>>>>>>> |||||||||||||||||||||||||||||| >>>>>>>>
68869090 taggggggtagtgaggatcttgatttggatc 68869119
```

pGEM-T DNMT3B-UTR
digested by NheI/FseI

1kb 100

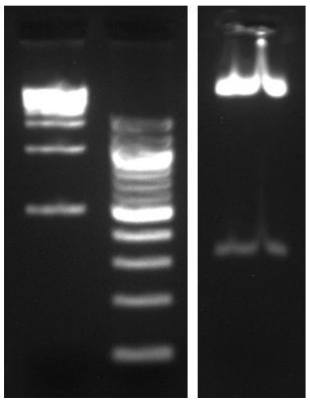


PG-DNMT3B 3UTR Clone 4

```
00000103 gcagagccacctgactcttgcaggggtagcctgaggtgccgcctccttgt 00000152
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859514 gcagagccacctgactcttgcaggggtagcctgaggtgccgcctccttgt 30859563

00000153 gcacaaatcagacctggctgcttggagcagcctaacacggtgctcatttt 00000202
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859564 gcacaaatcagacctggctgcttggagcagcctaacacggtgctcatttt 30859613

00000203 ttctttctcctaaaactttaaaacttgaagtaggtagcaacgtggcttttt 00000252
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859614 ttctttctcctaaaactttaaaacttgaagtaggtagcaacgtggcttttt 30859663

00000253 ttttttcccttcctgggtctaccactcagagaaacaatggctaagatacc 00000302
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859664 ttttttcccttcctgggtctaccactcagagaaacaatggctaagatacc 30859713

00000303 aaaaccacagtgccgacagctctccaatactcaggttaatgctgaaaaat 00000352
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859714 aaaaccacagtgccgacagctctccaatactcaggttaatgctgaaaaat 30859763

00000353 catccaagacagttattgcaagagtttaatttttgaaaactggctactgc 00000402
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859764 catccaagacagttattgcaagagtttaatttttgaaaactggctactgc 30859813

00000403 tctgtgtttacagacgtgtgcagttgtaggcatgtagctacaggacattt 00000452
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859814 tctgtgtttacagacgtgtgcagttgtaggcatgtagctacaggacattt 30859863

00000453 ttaagggcccaggatcgtttttttcccagggcaagcagaagagaaaatgtt 00000502
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859864 ttaagggcccaggatcgtttttttcccagggcaagcagaagagaaaatgtt 30859913

00000503 gtatatgtctttttacccggcacatttccccttgcctaaatacaagggctgg 00000552
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
30859914 gtatatgtctttttacccggcacatttccccttgcctaaatacaagggctgg 30859963

00000553 agtctgcacgggacctattag 00000573
>>>>>>>> ||||||||||||||||||| >>>>>>>>
30859964 agtctgcacgggacctattag 30859984
```

pGEM-T Rock1-UTR
digested by NheI/FseI

1kb 100



PG-Rock1 3UTR Clone 3

```
00000045 gtagaaggttgcaccaacataaaaaggaaatatggcaatacatccatgat 00000094
<<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
16784200 gtagaaggttgcaccaacataaaaaggaaatatggcaatacatccatgat 16784151

00000095 gttttccagttaacataggaattaccagataaatactgttaaactcttgt 00000144
<<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
16784150 gttttccagttaacataggaattaccagataaatactgttaaactcttgt 16784101

00000145 ccagtaacaagagttgattcatatggacagtatgatttattgtttatttt 00000194
<<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
16784100 ccagtaacaagagttgattcatatggacagtatgatttattgtttatttt 16784051

00000195 tttaaccaaatacctcctcagtaatttataatggctttgcagtaatgtgt 00000244
<<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
16784050 tttaaccaaatacctcctcagtaatttataatggctttgcagtaatgtgt 16784001

00000245 atcagataagaagcactggaaaaccgatcgtctctaggatgatatgcatg 00000294
<<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
16784000 atcagataagaagcactggaaaaccgatcgtctctaggatgatatgcatg 16783951

00000295 tttcaagtggtattgaaagccgcactgatggatatgt 00000331
<<<<<<<< |||||||||||||||||||||||||||||||||||| <<<<<<<<
16783950 tttcaagtggtattgaaagccgcactgatggatatgt 16783914
```

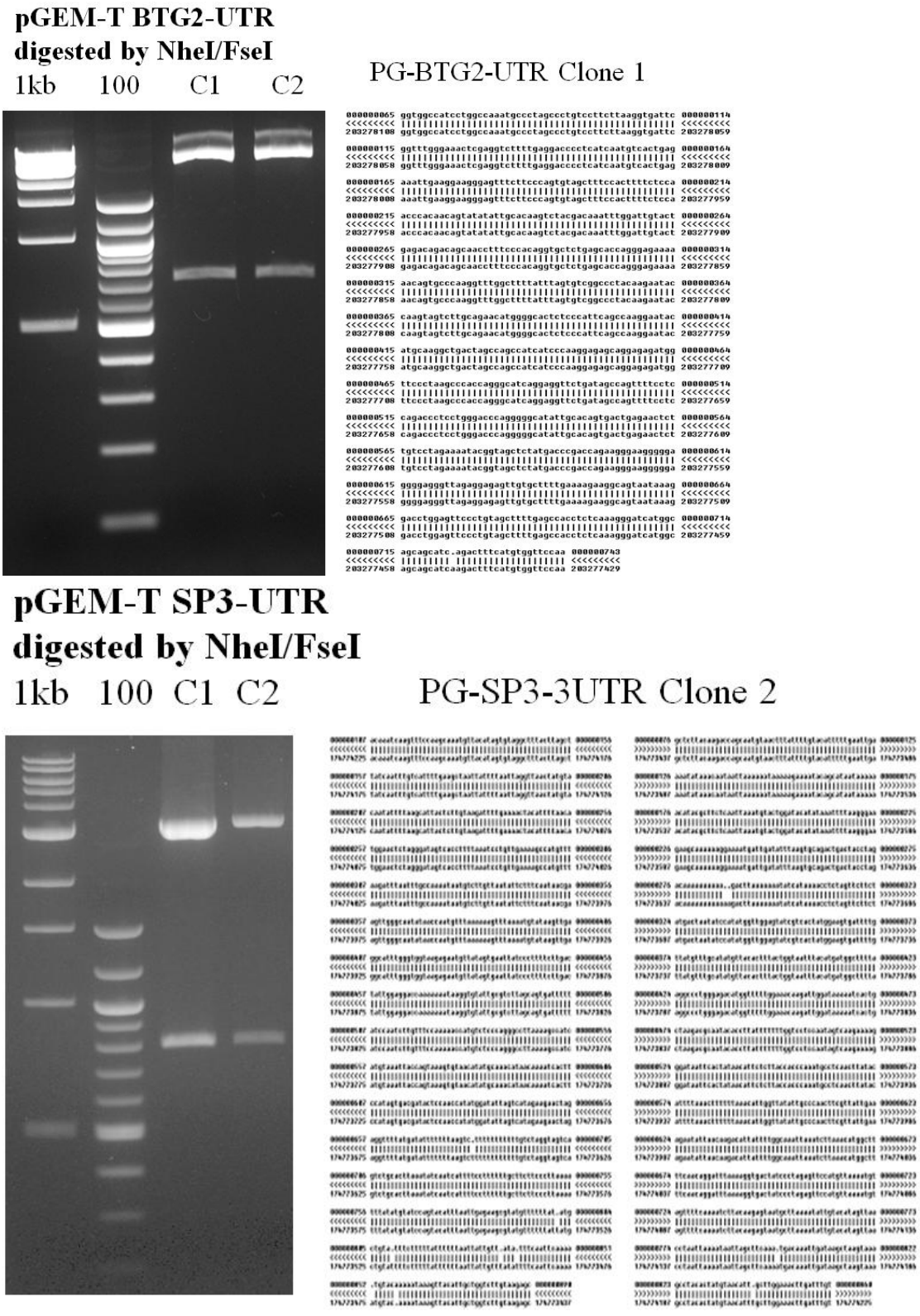
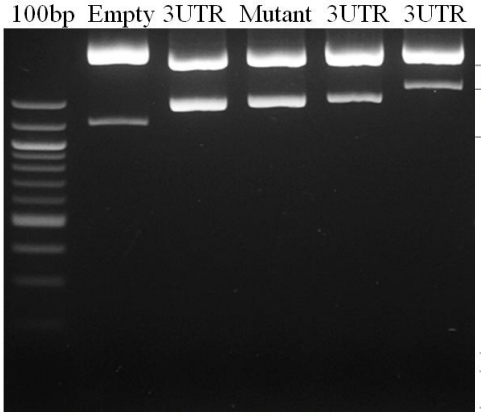


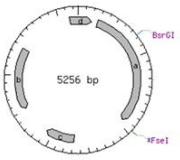
Figure S2.1
Gel images of digested pGEM-T with 3' UTR and their sequencing results. Segment of the 3' UTR were amplified by PCR and ligated into pGEM-T easy vector and sequenced. Sequencing results were analysed using human blat search (USCS genome bioinformatics).

pGL3-UTRs digested by BsrGI and FseI

Control PTEN PTEN Lefty1 CDH1
100bp Empty 3UTR Mutant 3UTR 3UTR

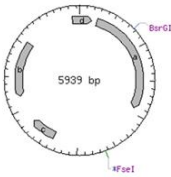


pGL3-Control vector



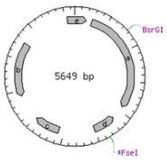
#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	1954-770	4073
2	BsrGI-FseI	771-1953	1183

pGL3-CDH1 3'UTR



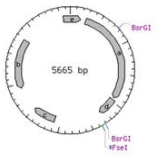
#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2637-770	4073
2	BsrGI-FseI	771-2636	1866

pGL3-Lefty1 3'UTR



#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2347-770	4073
2	BsrGI-FseI	771-2346	1576

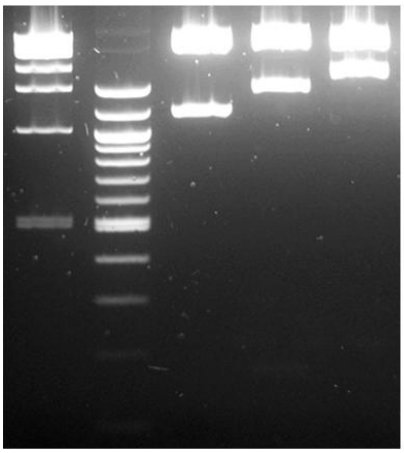
pGL3-PTEN 3'UTR



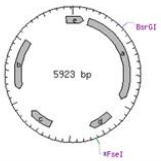
#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2363-770	4073
2	BsrGI-BsrGI	771-2314	1544
3	BsrGI-FseI	2315-2362	48

pGL3-UTRs digested by BsrGI and FseI

1kb 100 Control BTG1 BTG2
Empty 3UTR 3UTR

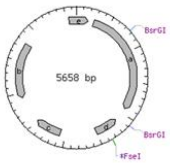


pGL3-BTG2 3'UTR



#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2621-770	4073
2	BsrGI-FseI	771-2620	1850

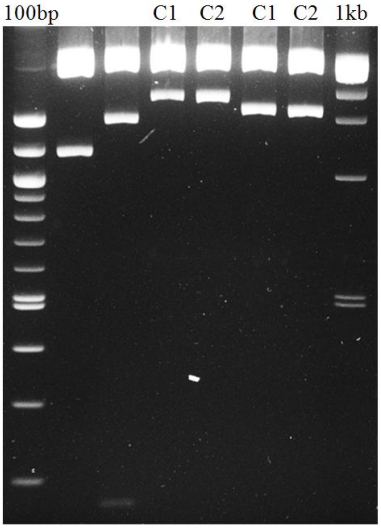
pGL3-SP3 (Site 1) 3'UTR



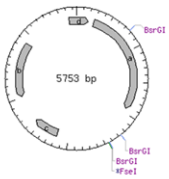
#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2356-770	4073
2	BsrGI-BsrGI	771-2079	1309
3	BsrGI-FseI	2080-2355	276

pGL3-UTRs digested by BsrGI and FseI

Empty BTG1 CDH1 DNMT3B
3UTR Mutant 3UTR
100bp C1 C2 C1 C2 1kb

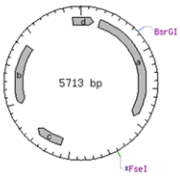


pGL3-BTG1 3'UTR

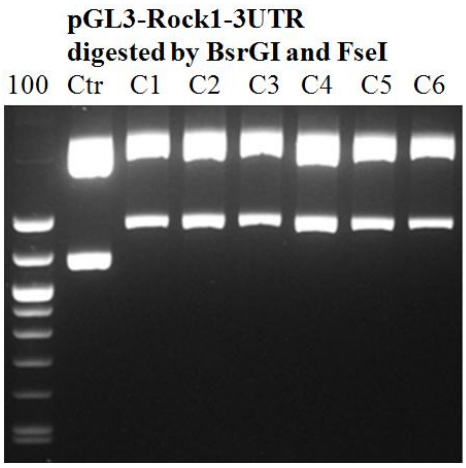


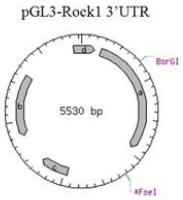
#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2451-770	4073
2	BsrGI-BsrGI	771-2267	1497
3	BsrGI-BsrGI	2268-2437	170
4	BsrGI-FseI	2438-2450	13

pGL3-DNMT3B 3'UTR

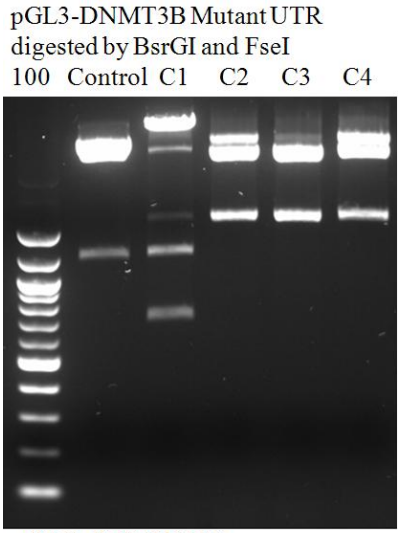


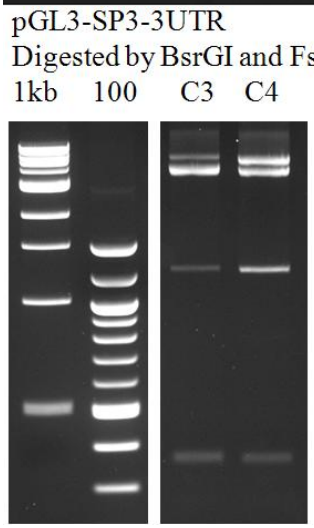
#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2411-770	4073
2	BsrGI-FseI	771-2410	1640

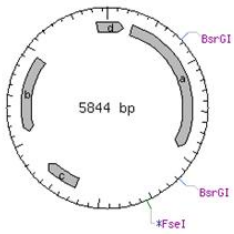




#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2228-770	4073
2	BsrGI-FseI	771-2227	1457







#	Ends	Coordinates	Length (bp)
1	FseI-BsrGI	2542-770	4073
2	BsrGI-BsrGI	771-2187	1417
3	BsrGI-FseI	2188-2541	354

Figure S2.2
Gel images of digested pGL3 reporter vectors.
pGL3 miRNA reporter vector constructs with inserts were validated by BsrGI and FseI digestion. Vector maps digested by the enzymes were generated by NEB cutter.

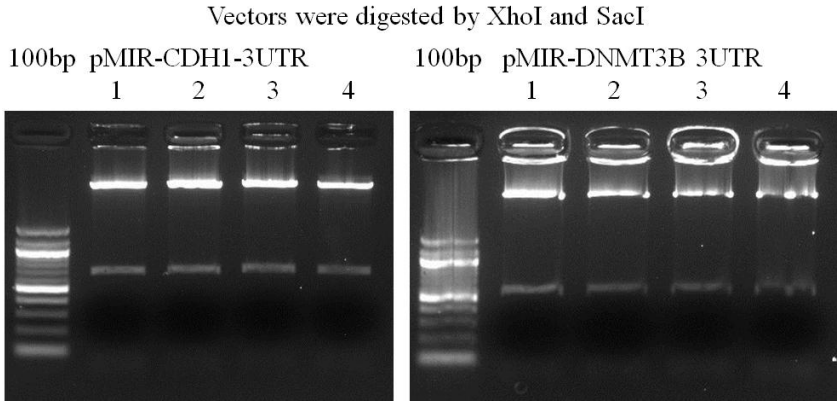


Figure S2.3
pMIR-miRNA reporter vectors were digested by XhoI and SacI to validate the presence of inserts.

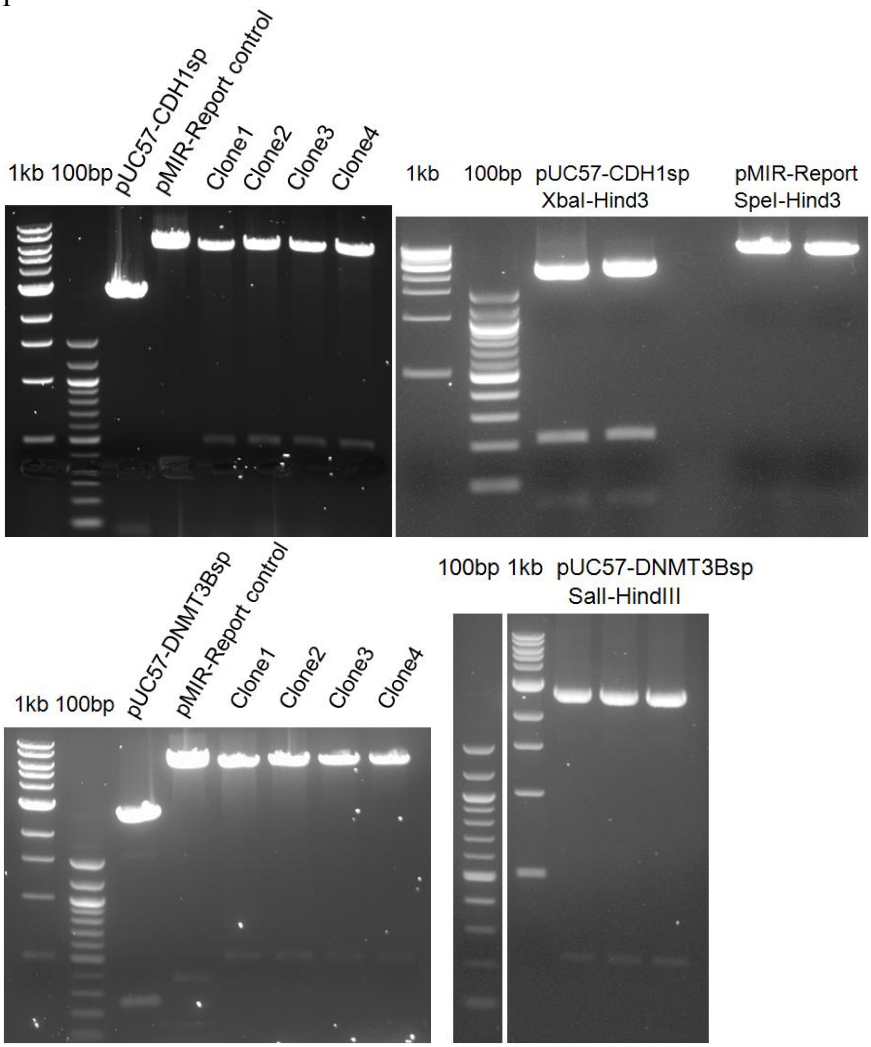


Figure S2.4
pMIR-Report-miR9 and -isomiR9 sponges were digested by ClaI and HindIII to validate the presence of inserts.

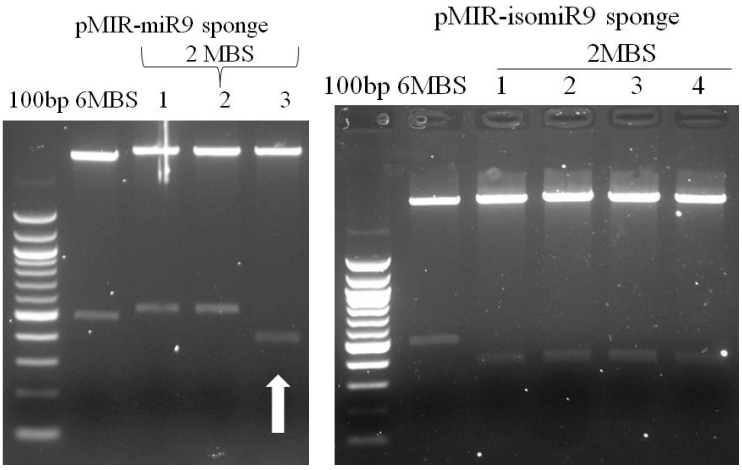


Figure S2.5
Gel image to validate the successful removal of 4 of the 6 MBS in sponge constructs.

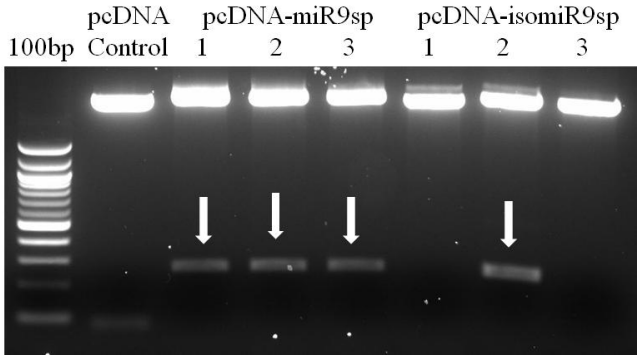


Figure S2.6
Gel image to look for the presence of inserts in pcDNA expression vectors.

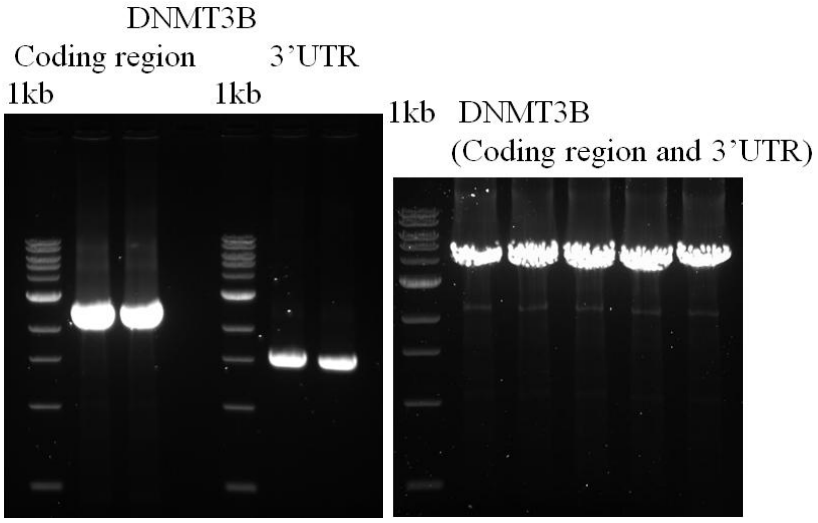


Figure S2.7
Gel image of PCR products of coding region and 3'UTR of DNMT3B.

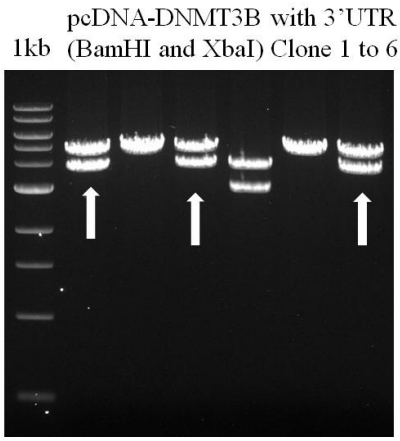


Figure S2.8
pcDNA-DNMT3B clone 1 to 6 were digested by BamHI and XbaI to look for clones that have the insert. Inserts were present in clone 1, 3 and 6.

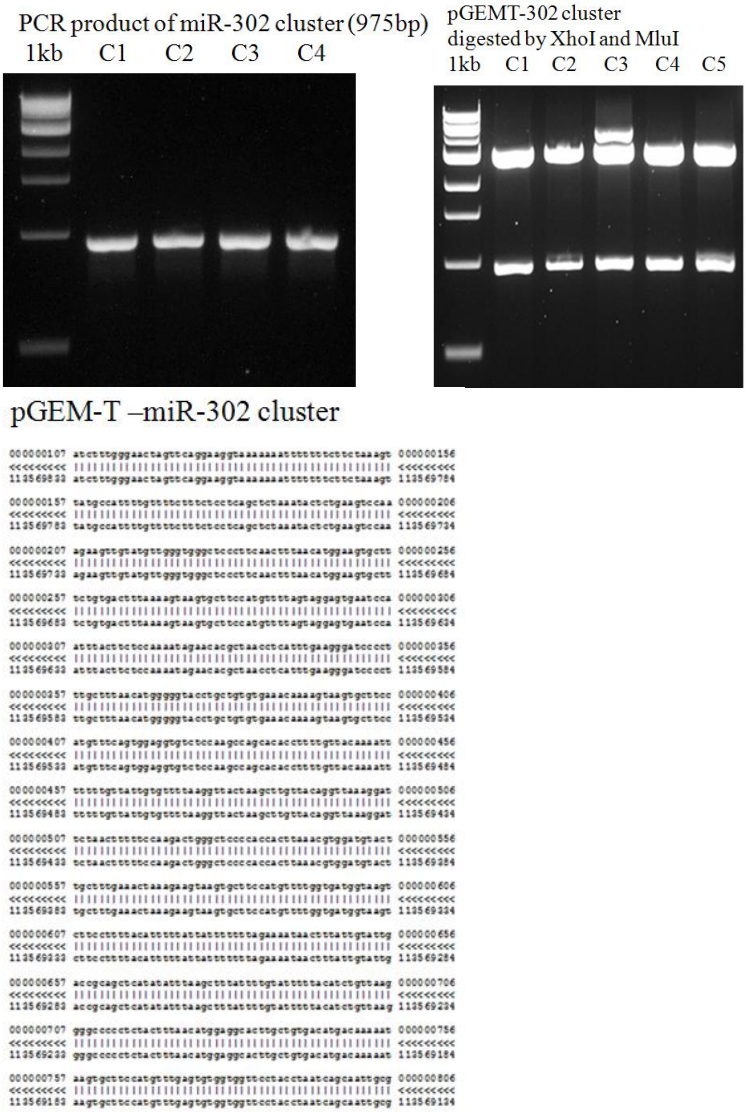


Figure S2.9
Gel image of PCR product of miR-302 cluster and validation of ligation into pGEM-T easy vector and its sequencing result.

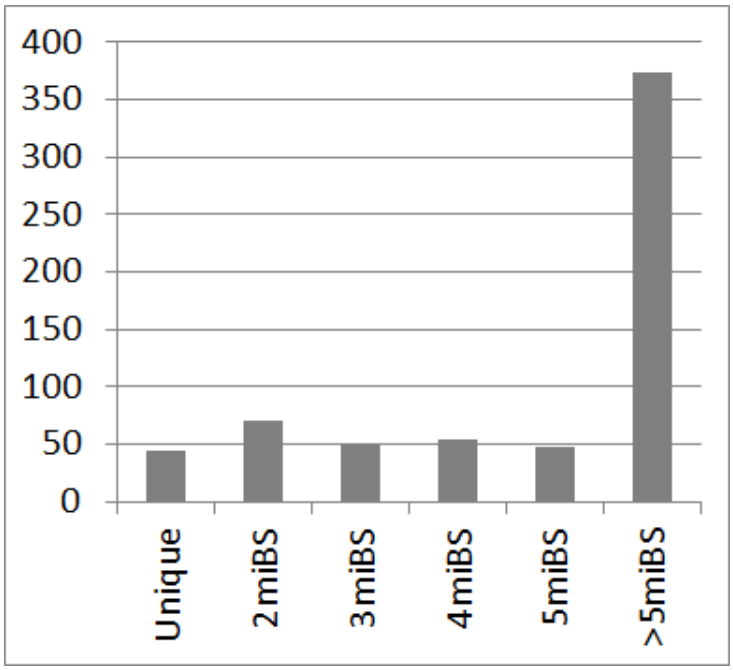


Figure S3.1
Figure illustrates the number of miRNA binding sites (miBS) in the genes that were predicted targets of isomiR-9. There are 44 predicted target genes that are solely target by isomiR-9, and were not predicted target of any canonical miRNAs.

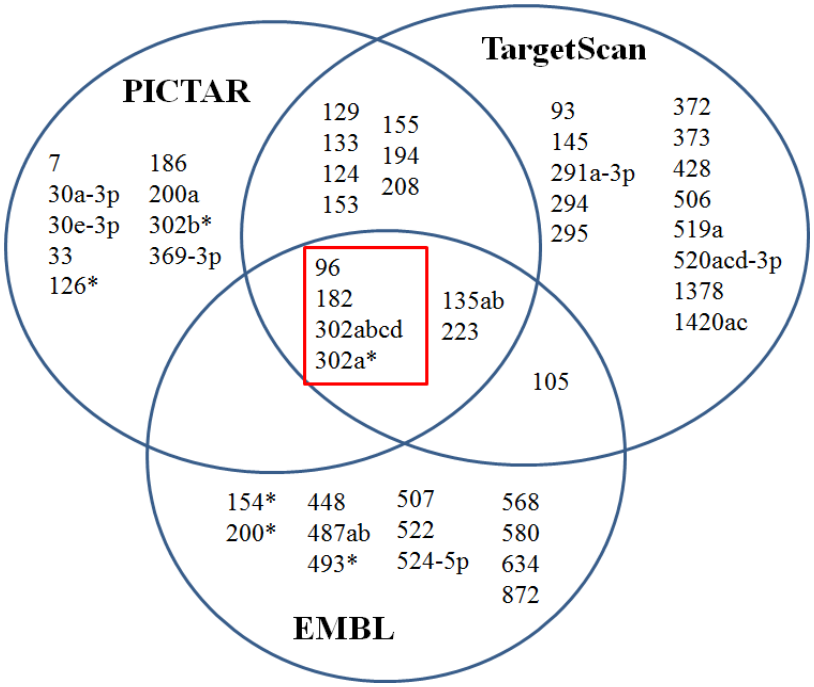


Figure S5.1
MiRNA target prediction of Sp3 transcription factor by 3 independent prediction databases, namely MicroCosm, Pictar and Targetscan.

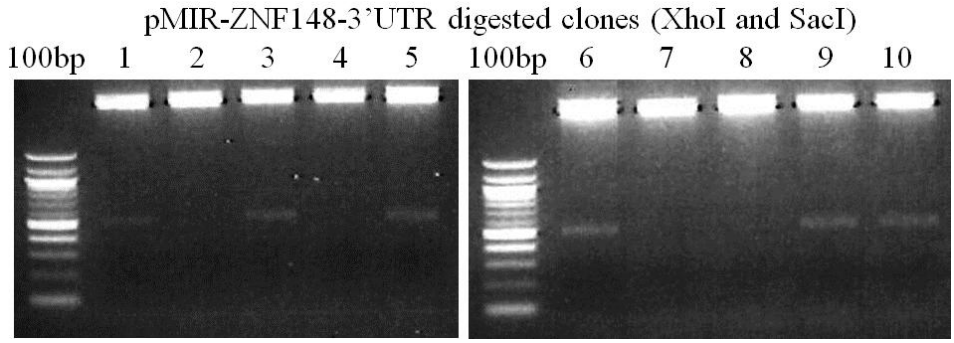


Figure S5.2 Gel image of digested pMIR-ZNF148-3' UTR.
pMIR-ZNF148-UTR reporter was validated by XhoI and SacI digestions which released a 500bp DNA fragment of ZNF148 3'UTR. Clones that were successfully ligated with ZNF148 -3'UTR include 1, 3, 5, 6, 9 and 10.

```
TGCCTTTTAC CCTTCTGGAG GAGAACACGA ATCTTTGGGA ACTAGTTCAG  113569814
GAAGGTAAAA AAATTTTTTT CTTCTAAAGT TATGCCATTT TGTTTTCTTT  113569764
CTCCTCAGCT CTAAATACTC TGAAGTCCAA AGAAGTTGTA TGTTGGGTGG  113569714
GCTCCCTTCA ACTTTAACAT GGAAGTGCTT TCTGTGACTT TAAAAGTAAG  113569664 miR-302b
TGCTTCCATG TTTTAGTAGG AGTGAATCCA ATTTACTTCT CCAAAATAGA  113569614
ACACGCTAAC CTCATTTGAA GGGATCCCCT TTGCTTTAAC ATGGGGGTAC  113569564 miR-302c
CTGCTGTGTG AAACAAAAGT AAGTGCTTCC ATGTTTCAGT GGAGGTGTCT  113569514
CCAAGCCAGC ACACCTTTTG TTACAAAATT TTTTTGTTAT TGTGTTTTAA  113569464
GGTTACTAAG CTTGTTACAG GTTAAAGGAT TCTAACTTTT TCCAAGACTG  113569414
GGCTCCCCAC CACTTAAACG TGGATGTACT TGCTTTGAAA CTAAAGAAGT  113569364 miR-302a
AAGTGCTTCC ATGTTTTGGT GATGGTAAGT CTTCCTTTTA CATTTTTATT  113569314
ATTTTTTTAG AAAATAACTT TATTGTATTG ACCGCAGCTC ATATATTTAA  113569264
GCTTTATTTT GTATTTTTAC ATCTGTTAAG GGGCCCCCTC TACTTTAACA  113569214 miR-302d
TGGAGGCACT TGCTGTGACA TGACAAAAAT AAGTGCTTCC ATGTTTGAGT  113569164
GTGGTGGTTC CTACCTAATC AGCAATTGCG TTAACGCCCA CACTGTGTGC  113569114
AGTTCTTGGC TACAGGCCAT TACTGTTGCT AATATGCAAC TCTGTTGAAT  113569064 miR-367
ATAAATTGGA ATTGCACTTT AGCAATGGTG ATGGATTGTT AAGCCAATGA  113569014
CAGAATTTAA ACCACAGACT TACTTTGATA GCACTCTTAA TGGTATAACT  113568964
TCTTCTCCCA TTTTATGTCT CTCTTTATGT TTTTTCTTAT GTTTCCTTTT  113568914
TGTTTTCAAG AGAGAGCTAT CTTTTAGATC TCCAGTATCC TTTTCCTCTT  113568864
```

Figure S5.3
MiR-302 cluster human genome DNA sequence located in chromosome 4. Red texts represent sequence of the members of miR-302 cluster gene, namely miR-302b, miR-302c, miR-302c, miR-302d and miR-367.

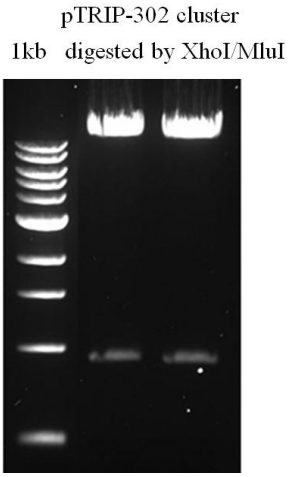


Figure S5.4 Gel image of digested pTRIPz-302 cluster.
pTRIPz-302 cluster was digested by XhoI and MluI that released the miR-302 cluster consisting of a 975bp DNA fragment. This image validated the successful ligation of miR-302 cluster in pTRIPZ lentiviral vector.

hsa-miR-302b-5p and hsa-miR-302c-5p

```
  UUUAACAUGGGGGUACCUGCUG       >hsa-miR-302c-5p MIMAT0000716
ACUUUAACAUGGAGGCACUUGC         >hsa-miR-302d-5p MIMAT0004685
ACUUUAACAUGGAAGUGCUUUC         >hsa-miR-302b-5p MIMAT0000714
```

hsa-miR-302c-5p MIMAT0000716

hsa-miR-302c-5p	hsa-miR-302c-3p	Count	RPM
.....GCUUUAACAUGGGGGUACCUGC.....		1	1.05
.......UUUAACAUGGGGGUACCUGCUG.....		3	3.15
.......UUUAACAUGGGGGUACCUGCU.....		1	1.77
........UUAACAUGGGGGUACCUGCUG.....		1	1.05
.....	AGUAAGUGCUUCCAUGUUUCAGU.....	2	2.1
.....	GUAAGUGCUUCCAUGUUUCAGU.....	5	5.26
.....	GUAAGUGCUUCCAUGUUUCAGUGG...	4	4.2
.....	UAAGUGCUUCCAUGUUUCAGUGG...	4951	5.27e+03
.....	UAAGUGCUUCCAUGUUUCAGU.....	3066	3.26e+03
.....	UAAGUGCUUCCAUGUUUCAG.....	184	201
.....	UAAGUGCUUCCAUGUUUCAGUGGA..	92	96.7
.....	UAAGUGCUUCCAUGUUUCAGUG....	68	75.1
.....	UAAGUGCUUCCAUGUUUCA.....	27	34.1
.....	UAAGUGCUUCCAUGUUU.....	5	8.13
.....	UAAGUGCUUCCAUGUUUC.....	3	4.59
.....	AAGUGCUUCCAUGUUUCAGUGG...	20	23.2
.....	AAGUGCUUCCAUGUUUCAGUGGA..	3	4.59
.....	AAGUGCUUCCAUGUUUCAGUG....	1	1.05
.....	AGUGCUUCCAUGUUUCAGUGG...	12	15.5

CCUUUGCUUUAACAUGGGGGUACCUGCUGUGUGAAACAAAAGUAAGUGCUUCCAUGUUUCAGUGGAGG

hsa-miR-302b-5p MIMAT0000714

hsa-miR-302b-5p	hsa-miR-302b-3p	Count	RPM
.....ACUUUAACAUGGAAGUGCUUUC.....		1	2.21
.....	AGUAAGUGCUUCCAUGUUUUAGUAG....	2	2.63
.....	AGUAAGUGCUUCCAUGUUUUAGU.....	1	1.31
.....	GUAAGUGCUUCCAUGUUUUAGUAG....	4	5.26
.....	GUAAGUGCUUCCAUGUUUUAG.....	1	1.31
.....	UAAGUGCUUCCAUGUUUUAGUAG....	28247	3.72e+04
.....	UAAGUGCUUCCAUGUUUUAGUA.....	3221	4.28e+03
.....	UAAGUGCUUCCAUGUUUUAGU.....	2174	3.13e+03
.....	UAAGUGCUUCCAUGUUUUAG.....	194	292
.....	UAAGUGCUUCCAUGUUUUA.....	21	34.8
.....	UAAGUGCUUCCAUGUUUU.....	17	34
.....	UAAGUGCUUCCAUGUUU.....	5	10.2
.....	UAAGUGCUUCCAUGUUUUAGUAGG...	4	5.26
.....	AAGUGCUUCCAUGUUUUAGUAG....	3	3.94
.....	AGUGCUUCCAUGUUUUAGUAG....	1	1.31
.....	UGCUUCCAUGUUUUAGUAG....	1	1.31

GCUCCCUUCAACUUUAACAUGGAAGUGCUUUCUGUGACUUUAAAAGUAAGUGCUUCCAUGUUUUAGUAGGAGU

hsa-miR-302d-5p MIMAT0004685

hsa-miR-302d-5p	hsa-miR-302d-3p	Count	RPM
.....ACUUUAACAUGGAGGCACUUGC.....		1	1.77
.....	AAAUAAGUGCUUCCAUGUUUGAGUG...	1	17.6
.....	AUAAGUGCUUCCAUGUUUGAGUGU..	5	5.26
.....	UAAGUGCUUCCAUGUUUGAGUGU..	36731	3.91e+04
.....	UAAGUGCUUCCAUGUUUGAGU....	520	596
.....	UAAGUGCUUCCAUGUUUGAGUG...	350	378
.....	UAAGUGCUUCCAUGUUUGAG.....	50	58.3
.....	UAAGUGCUUCCAUGUUUGA.....	7	9.52
.....	UAAGUGCUUCCAUGUUU.....	5	8.13
.....	UAAGUGCUUCCAUGUUUG.....	3	3.15
.....	AAGUGCUUCCAUGUUUGAGUGU..	19	20.7
.....	AGUGCUUCCAUGUUUGAGUGU..	40	47.1
.....	GUGCUUCCAUGUUUGAGUGU..	1	1.05
.....	UGCUUCCAUGUUUGAGUGU..	3	3.15
.....	GCUUCCAUGUUUGAGUGU..	1	1.05

CCUCUACUUUAACAUGGAGGCACUUGCUGUGACAUGACAAAAAUAAGUGCUUCCAUGUUUGAGUGUGG

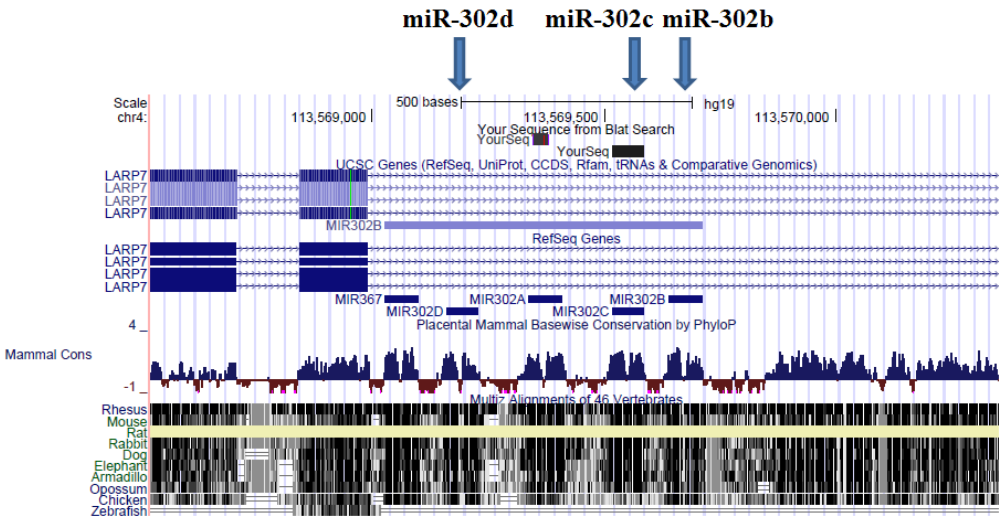


Figure S6.1 The dominant mature form of hsa-miR-302b-5p and hsa-miR-302c-5p represent the isomiRs of each others.

hsa-miR-518a-3p, hsa-miR-518f-3p and hsa-miR-518e-3p

```
GAAAGCGCUUCCCUUUGCUGGA          >hsa-miR-518a-3p MIMAT0002863
GAAAGCGCUUCUCUUUAGAGG           >hsa-miR-518f-3p MIMAT0002842
 AAAGCGCUUCCCUUCAGAGUG          >hsa-miR-518e-3p MIMAT0002861
```

```
>hsa-miR-518a-3p MIMAT0002863
GAAAGCGCUUCCCUUUGCUGGA
```

```
          hsa-miR-518a-5p                          hsa-miR-518a-3p                          Count      RPM
.............CUGCAAAGGGAAGCCCUUUCUG.....................................................    2          1.5
.............UGCAAAGGGAAGCCCUUUCUG.....................................................    2          1.5
.....................................................AAAGCGCUUCCCUUUGCUGGAU.............    3          0.325
UCUCAAGCUGUGACUGCAAAGGGAAGCCCUUUCUGUUGUCUGAAAGAAGAGAAAGCGCUUCCCUUUGCUGGAUUACGGUUUGAGA
```

```
>hsa-miR-518f-3p MIMAT0002842
GAAAGCGCUUCUCUUUAGAGG
```

```
          hsa-miR-518f-5p                          hsa-miR-518f-3p                          Count      RPM
.....................................................GAAAGCGCUUCUCUUUAGAGGA.............    3          1.13
UCUCAUGCUGUGACCCUCUAGAGGGAAGCACUUUCUCUUGUCUAAAAGAAAAGAAAGCGCUUCUCUUUAGAGGAUUACUCUUUGAGA
```

```
>hsa-miR-518e-3p MIMAT0002861
AAAGCGCUUCCCUUCAGAGUG
```

```
          hsa-miR-518e-5p                          hsa-miR-518e-3p                          Count      RPM
.............CUCUAGAGGGAAGCGCUUUCUG.....................................................    2          0.303
.....................................................AAAGAAAGCGCUUCCCUUCAGAGU..............    1          14.2
.....................................................GAAAGCGCUUCCCUUCAGAGUG.............    1          1.05
UCUCAGGCUGUGACCCUCUAGAGGGAAGCGCUUUCUGUUGGCUAAAAGAAAAGAAAGCGCUUCCCUUCAGAGUGUUAACGCUUUGAGA
```

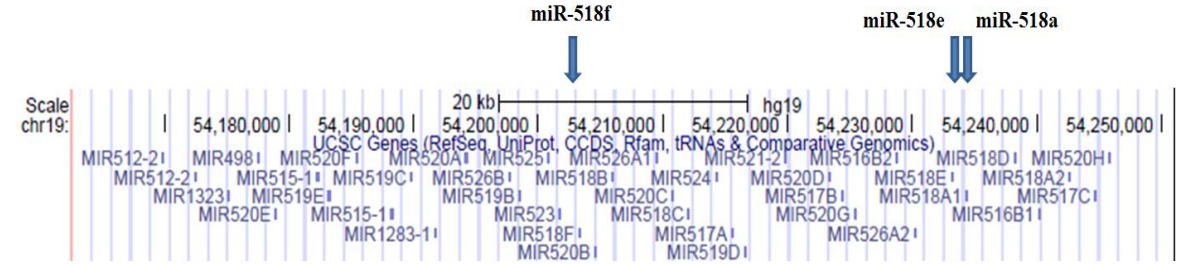


Figure S6.2 hsa-miR-518a-3p, hsa-miR-518f-3p and hsa-miR-518e-3p
Other examples where the dominant mature form represent the isomiRs of each others.