UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Examination of the Piwi/piRNA pathway in somatic stem cells of the mouse

Alexandra L. McCorkindale

A thesis in complete fulfilment of the requirements for the degree of

Doctor of Philosophy

Victor Chang Cardiac Research Institute

St Vincent's Clinical School, Faculty of Medicine

August, 2014

Supervisor: A/Prof. Catherine M. Suter, PhD.

Co-supervisors: Dr. Jennifer E. Cropley, PhD., Dr. Romaric Bouveret, PhD.

"What drives scientists is this tremendous intellectual adventure – pushing the boundaries of knowledge, walking down a track that nobody has walked before, not knowing what's around the corner, and then seeing a landscape so extraordinarily beautiful and complex, being part of the community that is driving the boundaries of knowledge and giving us insight into the amazing process of life." – Suzanne Cory

i. Abstract

The Piwi proteins are a clade of the Argonaute family specific to animals that bind small RNAs called Piwi-interacting RNAs (piRNAs). The Piwi/piRNA system is indispensable in germ cells, where it has dual roles in germline stem cell self-renewal and the epigenetic silencing of retrotransposons. Although Piwi/piRNA activity is generally assumed to be restricted to the germline, there is mounting evidence that Piwi genes and proteins also have a somatic function. In primitive metazoans, Piwi orthologues appear to be important for self-renewal of somatic stem cells, and Piwi homologues are expressed in multiple mammalian tissues as well as in a variety of cancers.

I hypothesised that the Piwi/piRNA system is important for mammalian somatic stem cell self-renewal, in particular, in maintaining pluripotency in embryonic stem cells. I also hypothesised that the Piwi/piRNA pathway is involved in conferring the stem cell-like phenotype of certain cancer cells. To test these hypotheses, I investigated the expression and function of the murine Piwi homologue, Miwi2, in embryonic stem (ES) cells and a panel of breast cancer cell lines.

In this thesis, I show that Miwi2 is expressed in the cytoplasm of murine ES cells, and that its expression diminishes along with ES cell differentiation; however, Miwi2 does not appear to be critical for ES cell self-renewal and pluripotency. I also show that Miwi2 is highly expressed in metastatic 4T1 murine breast cancer cells, and that ectopic *Miwi2* expression in syngenic, non-metastatic 4TO7 cells causes gene and small RNA expression changes in pathways relevant to metastasis. In ES cells, I find that Miwi2 associates with repeat-derived piRNAs and piRNAs produced from large intergenic clusters, as well from the 3'UTR of genes, including the master regulators of pluripotency, *Oct4, Sox2* and *Nanog*. In 4T1 cells, Miwi2 associates with piRNAs from 3'UTR clusters, as well with 5' tRNA fragments that are linked to translational control. The function of Miwi2 in ES cells and cancer biology remains unclear, but my data indicate that Miwi2 and its associated piRNAs have a role in the regulation of gene expression in somatic cells.

ii. Acknowledgments

First and foremost, I would like to give thanks to my supervisor, A/Prof. Catherine Suter, our fearless leader. From the second we accidentally started chatting at the VCCRI Open Day, to the animated phone call where she excitedly agreed to host me for my PhD, and throughout my time in her vibrant lab, I have wondered just *how* I was so fortunate for the Gods to have given me such a wonderful supervisor. She gave me intellectual freedom when I craved it, and has been right beside me when I needed it most: the times I lacked direction and felt lost, and through every step of the writing process. She has taught me how to think critically and to be a rigorous scientist, and I thank her for introducing me to William Strunk and E. B. White at a most crucial moment of my candidature. I am grateful for her shining personality, friendship, and for her love of celebrating successes with French champagne.

Secondly, I would like to thank Dr. Jennifer Cropley for her day-to-day leadership and expertise. In addition, her careful eye over each (yes, each) chapter in this thesis ensured that every analysis was relevant, no punctuation mark was out of place, and every sentence was perfectly constructed. As far as great scientists go, she is among the best, and I am extremely lucky to have had her as a mentor, role model, and friend throughout my PhD.

Next, it remains to thank Dr. Romaric Bouveret for sharing his expertise in ES cell culture and for his invaluable advice along the way. I am also grateful to all members of the Suter Laboratory, especially to Paul Young for his bioinformatics genius, and to Dr. Simon Keam for getting the ball rolling on the somatic Piwi project and for making me laugh every day.

I will be forever appreciative for the expertise, guidance, and friendship of every single person at the VCCRI during these critical years of my scientific career. I can quite hardly believe how so many brilliant minds can exist at all, let alone under one roof. The never-ending words encouragement that I received from my peers and my colleagues during my time at the VCCRI are akin to those of the late Thomas Edison:

> *"I have not failed. I've just found 10,000 ways that don't work"* – Thomas A. Edison

There are, however, several people that should be acknowledged individually. Firstly, Dr. Callum Smits and Dr. Chu Kong ("The Oracle") Liew for their expertise in all things protein-related. I would also like to thank Dr. Gavin Chapman for sharing his expertise in confocal microscopy. Thank you to Dr. Nicole Schonrock for performing the polybrene experiment, Dr. Alastair Stewart for his assistance with generating the Miwi2 homology model, and Peter

Schofield for sharing his expertise in hybridoma culture and antibody purification. In addition, thank you to Christopher Brownlee (aka. “The Aria Whisperer”) for his much-appreciated assistance with flow cytometry at the earliest stages of my PhD, even faced with my tears, frustration, and complete incompetence. I would also like to thank Alexis Bosman and Gustavo Alencastro (Tam Laboratory, CMRI) for their assistance with lentiviral transduction. I would also like to acknowledge funding from the National Health and Medical Research Council of Australia (NHMRC #1025210).

Surviving a PhD is also reliant upon a solid support network outside of the laboratory, and I could not have a stronger one. Thank you to my most excellent female support network (in particular, Annabel, Pip, Kat, and Kate) for the delicious meals, countless glasses of wine, dancing, inspirational quotes, and cuddles. A special thanks goes to my gloriously-intelligent friend Annabel, who painstakingly proof-read far too much of this thesis. Thank you to the Rapha girls and their shiny, two-wheeled companions for getting me out of bed at 5am and keeping me sane (and fit) during the writing process. Thank you to all of my other wonderful friends for beers and laughs along the way, and for pretending to understand what I do.

Thank you to my beloved Mum and Dad for their unconditional love and support always, for their friendship, for feeding me, for introducing me to cycling, for taking me skiing, for buying me blueberries, for supplying me with coffee, for hugs when I needed them, and most importantly, for stocking my wine fridge full of Barossa Valley shiraz. What “PhD student life”?

Finally, I would like to thank Oliver, the wonderful man whom I met along the way and who invited me into his crazy life. From our third date, where he accompanied me on a visit to the lab for a poorly-timed media change, during the hopeless times where I thought the end would never come, to now, as I conclude this chapter of my life and plan for my next, his encouragement has been unfaltering. I will be eternally grateful for his support, friendship, and companionship, and can’t wait to see what our next adventure has in store for us.

“Somewhere, something incredible is waiting to be known” - Carl Sagan.

iii. Certificate of originality

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed .....

Date .....

iv. Table of Contents

v. List of Figures

vi. List of Tables

vii. Table of Units and Abbreviations

Abbrev.	Full description
bp	base pairs
kb	kilobase pairs
nt	nucleotides
nm	nanometres
mm	millimetres
µM	micromolar
mM	millimolar
M	molar
µl	microlitres
ml	millilitres
ng	nanograms
µg	micrograms
mg	milligrams
g	grams
E	embryonic day
d	days
h	hours
min	minutes
s	seconds
× g	times gravity
°C	degrees Celcius
rad	radians
V	volts
5’	5’ end of nucleic acid
3’	3’ end of nucleic acid
1°	primary
2°	secondary
A	adenine
T	thymine
C	cytosine
G	guanine
U	uracil
piRNAs	piwi-interacting RNAs
RNP	ribonucleoprotein
RNAi	RNA interference
dsRNA	double-stranded RNA
ssRNA	single-stranded RNA
shRNA	short hairpin RNA
PCR	polymerase chain reaction
RT-PCR	reverse transcription PCR
qPCR	quantitative PCR
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger RNA
ORF	open reading frame
DTT	dithiothreitol
IAP	intracisternal A-particle
ncRNA	non-coding RNA
PCR	polymerase chain reaction
RNase	ribonuclease
DNase	deoxyribonuclease
UV	ultraviolet
GSC	germline stem cell
PGC	primordial germ cell
IPA	Ingenuity pathway analysis
IP	immunoprecipitation
EtOH	ethanol
RT	room temperature
DMEM	Dulbecco’s modified Eagle’s Medium

viii. List of publications and presentations

Publication: Simon P Keam, Paul E Young, **Alexandra L McCorkindale**, Thurston HY Dang, Jennifer L Clancy, David T Humphreys, Thomas Preiss, Gyorgy Hutvagner, David IK Martin, Jennifer E Cropley, and Catherine M Suter. 2014. The human Piwi protein, Hiwi2, associates with tRNA-derived piRNAs in somatic cells. *Nucleic Acids Research*. Epub ahead of print.

Publication: Alexandra L McCorkindale, Paul E Young, Jennifer E Cropley, and Catherine M Suter. Miwi2 associates with 3'UTR-derived piRNAs in mouse embryonic stem cells and cancer cells. Manuscript in preparation.

Oral presentation: "Investigating expression and function of a Piwi protein in stem cells and cancer", Paul Korner Lecture Series, Victor Chang Cardiac Research Institute. 2013.

Poster presentation: "Piwi has an unexpected role in stem cell self-renewal and potency in mammalian development, homeostasis and disease", 14th International EMBL PhD Symposium on Networks in the Life Sciences, Heidelberg, Germany. 2012.

Poster presentation: "Piwi: could the guardian of the genome act as the villain in cancer?" 7th International Heinrich F. C. Behr Symposium on Stem Cells and Cancer, Heidelberg, Germany. 2012.

Oral presentation: "A story about Piwi: why a stem cell biologist should care about the testes", Paul Korner Lecture Series, Victor Chang Cardiac Research Institute. 2012.

Oral presentation: "Piwi: an epigenetic regulator with an unexpected role in stem cell self-renewal and potency", 4th Australian Epigenetics conference, Adelaide, Australia. 2012.

Poster presentation: "Investigating the role of the murine Piwi/piRNA system in regulation and maintenance of the undifferentiated state", 33rd Annual Lorne Genome Conference, Lorne, Australia. 2012.

Poster presentation: "Investigating the role of the Piwi/piRNA pathway in regulation and maintenance of the pluripotent state", Stem Cell Programming and Reprogramming conference, Lisbon, Portugal. 2011.

Oral presentation: "Small RNAs: from silencing to stem cells", Australian Epigenetics Alliance meeting, Victor Chang Cardiac Research Institute. 2011.

Oral presentation: "The Piwi/piRNA pathway: an unexpected role in potency", Paul Korner Lecture Series, Victor Chang Cardiac Research Institute. 2011.

Poster presentation: "Investigating the epigenetic role of the Piwi/piRNA system in differentiation", Non-coding RNA, Epigenetics and the Environment conference, London, UK. 2011.

1. Introduction

Argonaute and Piwi proteins

The Piwi proteins are members of the Argonaute family with conserved protein structure encompassing four domains: N-terminal, Piwi-Argonaute-Zwille (PAZ), Mid, and Piwi (Meister, 2013). These are RNA-binding proteins known to be involved in germline stem cell function and transposon silencing in *Drosophila melanogaster,* mice, and humans; and stem cell maintenance in primitive organisms (reviewed in Juliano et al., 2011).

The Argonaute proteins

The Argonaute proteins – in complex with small regulatory RNAs – mediate gene silencing via direct endonucleolytic cleavage of target transcripts or via translational repression (Valencia-Sanchez et al., 2006). These proteins are composed of several domains, including a variable N-terminal domain and two cooperative RNA binding domains: PAZ and Mid (Figure 1-1). The core of the C-terminal PIWI domain has ribonuclease H activity which mediates cleavage of the PAZ/Mid-bound RNA transcript (Song et al., 2004). Argonaute proteins are highly conserved, but there is extraordinary variety in the number of Argonautes among species: *S. pombe* has only one*;* mice and humans have seven; *Drosophila* have five; *Arabidopsis* have ten; and *Caenorhabditis elegans* has 27 (Ghildiyal and Zamore, 2009, Pratt and MacRae, 2009, Su et al., 2009). In some species with multiple homologues, Argonaute proteins display some functional redundancy, for example, expression of either of the *C. elegans* Argonaute proteins, ALG-1 and ALG-2, is sufficient to compensate for the loss of the other (Grishok et al., 2001). Argonaute proteins are more extensively reviewed in (Pratt and MacRae, 2009) and (Meister, 2013).

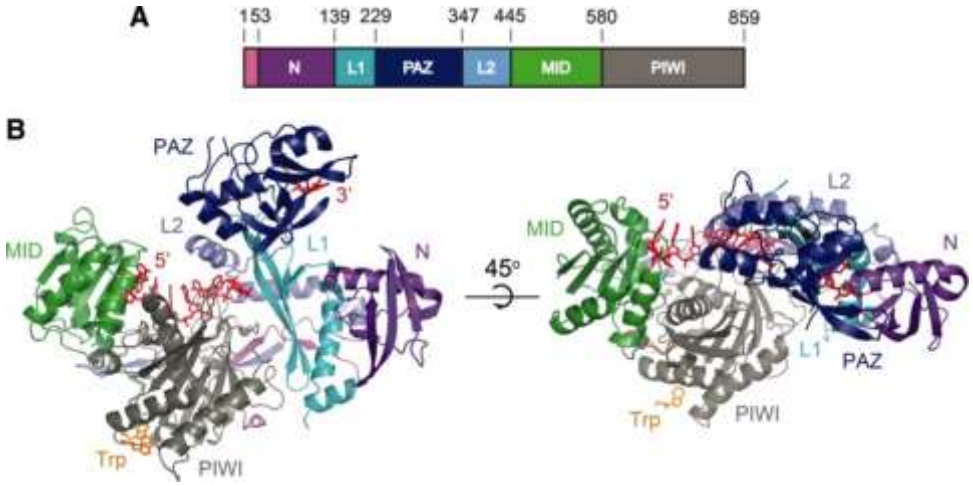


Figure 1-1: Structure of human Argonaute, Ago2, in complex with RNA

(A) Schematic of the human Ago2 domain structure. (B) Pseudo-3D representation of Ago2 with the N-terminal (purple), PAZ (navy), MID (green), and PIWI (gray) domains and linkers L1 (teal) and L2 (blue). A 21nt generic guide RNA is bound to the Argonaute protein, 5' and 3' ends indicated (red). Tryptophan molecules (Trp, orange) bind to tandem hydrophobic pockets in the PIWI domain. Obtained from (Schirle and MacRae, 2012) and used with permission from The American Association for the Advancement of Science.

Piwi is an Argonaute protein and is highly conserved in metazoans

The first report of a Piwi gene was in the *Drosophila* ovary, where disruption of *Piwi* (P-element induced wimpy testis) via P-element insertion abolished germline stem cell division (Lin and Spradling, 1997). Since then, Piwi genes have been described in multiple animals (refer to Table 1-1), although they have not been reported in plants. The Piwi proteins are members of the Argonaute family described above as they are RNA-binding proteins containing PAZ, Mid, and Piwi domains. However, they are distinguished from the classic Argonaute proteins in many organisms by their unique association with a 26-32nt small RNA population, the piwi-interacting RNAs (piRNAs) (Saito et al., 2006, Vagin et al., 2006). For some time, available evidence showed that expression of members of the Piwi protein family was restricted to the germline in animals; however, reports of somatic Piwi expression are emerging. Piwi proteins are highly conserved; they are expressed in ciliates and slime moulds and in all metazoans: from primitive sponges to mammals, including mice and humans (Table 1-1). Piwi homologues in *Drosophila,* mice, and humans are listed in Table 1-2. This thesis will focus on the murine Piwi proteins: Miwi, Mili, and Miwi2. There is no available crystal structure for mammalian Piwi proteins; the archaeal *Archaeoglobus fulgidis* Piwi is the only Piwi protein for which the structure has been determined (Figure 1-2).

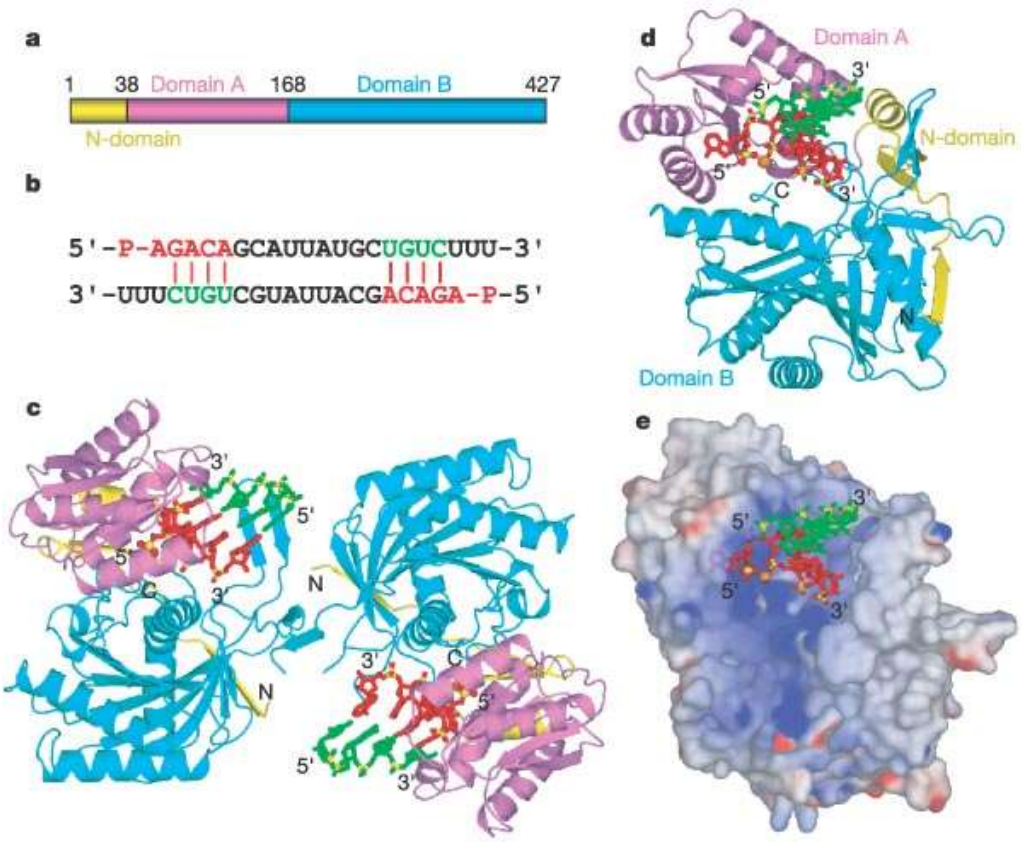


Figure 1-2: Structure of the *Archaeoglobus fulgidis* Piwi protein

(a) The *A. fulgidus* Piwi protein consists of an N-domain (1–37, yellow), domain A (38–167, magenta), and domain B (168–427, cyan). (b) The sequence and pairing alignment of the 21-mer RNA. The red and green segments are observed in the crystal structure of the complex, whereas those in black are disordered. (c) The relative alignments of the two Piwi proteins together with their bound RNAs within the asymmetric unit. (d) Ribbon and (e) electrostatic views of the complex. The RNA is shown in a stick representation. Obtained from (Ma et al., 2005) and used with permission from Nature Publishing Group.

Table 1-1: Piwi genes are highly conserved through metazoans.

Table adapted from (Ross et al., 2014).

Phylum	Common	Species	Described *Piwi* genes	Protein localisation	Reference
Porifera	Sponge	*Ephydatia fluviatilis*	*EfPiwiA, B*	Archeocytes	(Funayama et al., 2010)
Cnidaria	Jellyfish	*Clytia hemisphaerica*	*Piwi*	Somatic stem cells of the tentacle bulb	(Denker et al., 2008)
	Jellyfish	*Podocaryne carnea*	*Cniwi*	Somatic stem cells of the tentacle bulb	(Seipel et al., 2004)
Ctenophora	Jellyfish	*Pleurobrachia pileus*	*PpiPiwi1, 2*	Actively dividing somatic cells, germline	(Alie et al., 2011)
Platyhel-minthes	Planaria	*Schmidtea mediterranea*	*Smedwi1, 2, 3*	Neoblasts (totipotent mitotic cells)	(Reddien et al., 2005, Palakodeti et al., 2008)
	Flatworm	*M. lignano*	*Macpiwi*	Neoblasts (totipotent mitotic cells)	(De Mulder et al., 2009)
Mollusca	Sea slug	*Aplysia californica*	*Piwi*	Nervous system, heart, and germline	(Rajasethupathy et al., 2012)
Arthropoda	Fruit fly	*Drosophila melanogaster*	*Piwi, Aub, Ago3*	Gonads, brain, salivary gland	(Cox et al., 1998, Brower-Toland et al., 2007, Harris and Macdonald, 2001, Li et al., 2009)
Chordata	Sea squirt (ascidian)	*Botrylloides leachi Botryllus schlosseri*	*BlPiwi*	Stem cells	(Rinkevich et al., 2010, Rinkevich et al., 2013)
	Frog	*Xenopus tropicalis Xenopus laevis*	*Xiwil1a/b, Xili, Xiwi2*	mRNA detected in somatic and germline cells, protein detected in germline only	(Wilczynska et al., 2009)
	Mouse	*Mus musculus*	*Miwi, Mili, Miwi2*	Male germ line, diverse cancers	(Deng and Lin, 2002, Kuramochi-Miyagawa et al., 2004, Carmell et al., 2007, Lee et al., 2010)
	Human	*Homo sapiens*	*Hiwi, Hili, Hiwi2*	Male germ line, diverse cancers	(Sharma et al., 2001, Qiao et al., 2002, Lee et al., 2010, Zhao et al., 2012, Taubert et al., 2007, Liu et al., 2006)

Table 1-2: Piwi protein homologues in *Drosophila*, mice and humans

HUGO gene symbol	*Drosophila*	Mouse	Human
Piwil1	Piwi	Miwi	Hiwi
Piwil2	Aub	Mili	Hili
Piwil4	Ago3	Miwi2	Hiwi2

The murine Piwi protein family

In the mouse there are three Piwi proteins: Miwi, Mili, and Miwi2. Each of these display distinct patterns of expression in the testis. For example, Mili expression is restricted to perinuclear cytoplasmic granules in male germ cells, whereas Miwi2 is found predominately in the male germ cell nucleus, with a lesser fraction co-localising with Mili in the cytoplasm (Aravin and Bourc'his, 2008). Murine spermatogenesis can be divided into two stages of development: the pre-meiotic (pre-pachytene) phase, and the post-meiotic (pachytene) phase. The three murine Piwi proteins are also expressed during discrete stages of germline development (Figure 1-3): expression of Mili appears *in utero* and persists throughout pre-pachytene and pachytene spermatogenesis, whereas Miwi expression peaks during the post-meiotic pachytene stage. Miwi2 is most abundant in a small developmental window around the time of birth, concurrent with germline *de novo* methylation of transposable elements (Aravin et al., 2008) but is still expressed in testis from 10 day-old mice and is also detectable in the adult testis (Carmell et al., 2007). In the testis, Mili- and Miwi2-bound piRNAs are predominately pre-pachytene and differ in their size and genomic origin to the Miwi-bound pachytene piRNAs. Different spatio-temporal patterns of expression of the Piwi proteins – along with distinct bound piRNA subsets – implies that they may have individual functions.

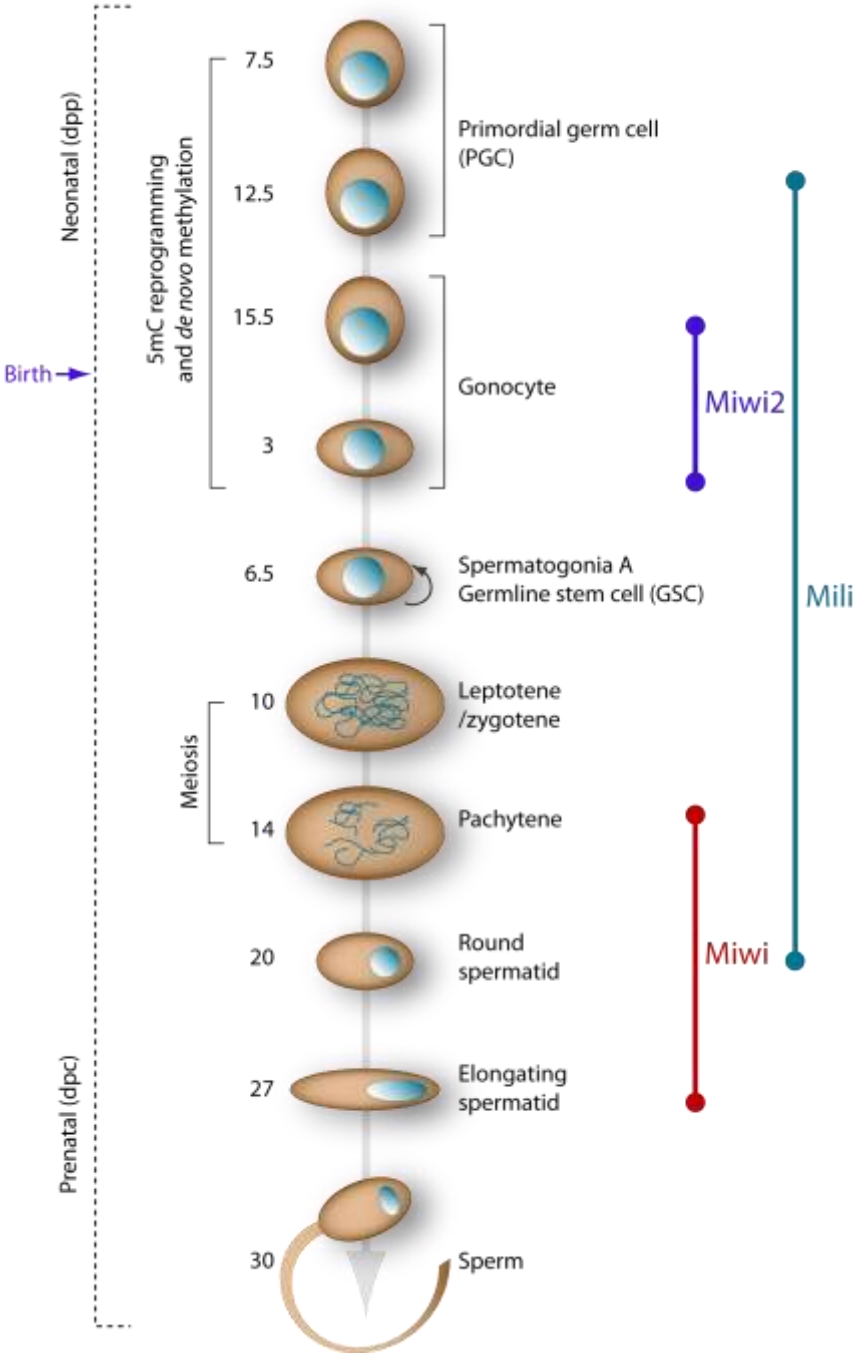


Figure 1-3: Spatio-temporal expression of murine Piwi proteins during development

Schematic depicting specific stages of murine spermatogenesis with previously reported expression timing of murine Piwi proteins Miwi, Mili, and Miwi2. Adapted from (Aravin et al., 2008) and (Thomson and Lin, 2009).

Noncoding regulatory RNAs

Small noncoding RNA populations are complex and varied

In 2012, the Encyclopaedia of DNA Elements (ENCODE) consortium reported that approximately 80% of the genome transcribes noncoding RNAs (Rosenbloom et al., 2012). The first described small regulatory RNA was the lin-4 microRNA (miRNA) (Lee et al., 1993), followed by the discovery of small interfering RNA (siRNA) silencing pathways in *C. elegans* (Fire et al., 1998). Reports of a second miRNA, let-7, emerged shortly after the discovery of RNA interference (RNAi) (Reinhart et al., 2000), and it was soon evident that siRNAs and miRNAs shared common biogenesis pathways (Grishok et al., 2001). miRNAs are perhaps the best characterised small noncoding RNAs: 21-25nt single-stranded RNA molecules processed from larger, double-stranded precursors by the enzymes Drosha and Dicer in animals (reviewed in (Ghildiyal and Zamore, 2009)). siRNAs and miRNAs are bound by the RNA-binding Argonaute proteins, which are guided by their small RNA counterparts to their targets by sequence homology, typically resulting in gene silencing (Ghildiyal and Zamore, 2009). miRNAs are derived from a number of multifunctional precursor molecules: protein-coding genes and long non-coding transcripts (reviewed in (Maute et al., 2014)), small nucleolar RNA (snoRNA) (Ender et al., 2008), or transfer RNA (tRNA) (Maute et al., 2013). Since the discovery of miRNAs and siRNAs, many other types of small RNA have been described with a variety of regulatory functions; Table 1-3 lists a summary of eukaryotic small noncoding RNAs and their current known or postulated functions. For an exhaustive list of all noncoding RNA (including long noncoding RNA) refer to (Ghildiyal and Zamore, 2009), (Kawaji et al., 2008) and (Fatica and Bozzoni, 2014).

Table 1-3: List of several types of small regulatory RNAs and their putative functions

ncRNA	Name	Length (nt)	Source	Putative functions	References
miRNA	MicroRNAs	19 – 24	Widespread encoded loci	mRNA silencing, inhibition of translation, promotion of gene expression	(Lee et al., 1993, Reinhart et al., 2000, Williams et al., 2008, Vasudevan, 2012)
siRNA	Small interfering RNAs	20 – 24	Endogenous /exogenous sources	Cleavage of target mRNA	(Fire et al., 1998, Riedmann and Schwentner, 2010)
piRNA	Piwi-interacting RNAs	26 – 31	Widespread loci incl. repeat elements, intergenic and genic regions	Transposon transcript cleavage, epigenetic silencing, epigenetic inheritance	(Brennecke et al., 2007, Kuramochi-Miyagawa et al., 2008, Aravin et al., 2008, Ashe et al., 2012)
hcRNA	Heterochromatic RNAs	24	Endogenous dsRNA precursors	Formation of heterochromatin	(Djikeng et al., 2001, Reinhart and Bartel, 2002)
siteRNA	Small intergenic transposable elements RNA	23-29	Remnant repeat elements within intronic gene regions in *Xenopus*	Possible transcriptional inhibition of host gene via recruitment of repressive histone modifications	(Harding et al., 2014)
snRNA	Small nuclear (spliceosomal) RNA	~150	Each snRNA is encoded at multiple loci throughout the genome	RNA–RNA and RNA–protein interactions in the assembly and function of canonical spliceosomes	(Butcher and Brow, 2005)
snoRNA	Small nucleolar RNAs	60 – 300	Intronic gene regions	RNA biogenesis and ribosomal RNA modifications	(Kiss-Laszlo et al., 1996, Ni et al., 1997, King et al., 2003)
tiRNA	Transcription initation RNAs	17 – 18	Downstream of transcription start sites	Possible transcriptional regulation	(Taft et al., 2009)
tRF	tRNA-derived fragments	13-30	Full-length tRNA processed by ELAC2 or Dicer	Inhibition of translation, inhibition of prostate cancer cell proliferation	(Kawaji et al., 2008, Cole et al., 2009, Lee et al., 2009)

Piwi-interacting RNAs

piRNAs are a complex and diverse population with unique biogenesis

Small (24-29nt) RNAs mapping to repetitive sequence elements were originally identified during a small RNA profiling study in the *Drosophila* embryo and testis where they were termed repeat-associated small interfering RNAs, or rasiRNAs (Aravin et al., 2003). These rasiRNAs were later shown to interact with Piwi proteins, hence were re-classified as a subset of the Piwi-interacting RNAs, or piRNAs (Aravin et al., 2006). Since then, hundreds of thousands of individual piRNA sequences have been submitted to the online piRNA database, piRNABank, from multiple species, including *Drosophila*, human, mouse, rat, zebrafish, and platypus (Sai Lakshmi and Agrawal, 2008). Specific piRNA sequences tend not to be conserved between species, but next-generation sequencing of piRNAs from humans, mice, and rats reveals synteny of these clusters (Girard et al., 2006).

Like miRNAs, piRNAs associate with proteins belonging to the Argonaute family, although the biogenesis pathways and functions of miRNAs and piRNAs are unique. Many piRNAs appear to originate from large and sometimes overlapping clusters in intergenic and gene-poor regions of the genome (Lau et al., 2006). Clusters that exhibit strand specificity are transcribed in a single direction; others are encoded on both strands and are transcribed bi-directionally, starting from the centre of the cluster and extending downstream on each strand (Brennecke et al., 2007). In *Drosophila*, the expression of different types of clusters appears to be cell-specific: uni-directional clusters are preferentially transcribed in the somatic follicle cells, with bi-directional clusters in the germline-derived nurse cells of the ovary. In *Drosophila,* it appears that piRNA production from uni-strand and dual-strand clusters proceeds via distinct mechanisms (Klattenhoff et al., 2009, Pane et al., 2011, Zhang et al., 2012, Mohn et al., 2014, Zhang et al., 2014). It is thought that any single cluster encoding hundreds of piRNAs is controlled by a single upstream promoter specific to that cluster, at least in *Drosophila* (Brennecke et al., 2007).

The biogenesis of piRNAs was first described in *Drosophila* and is postulated to occur via two mechanisms: i) primary biogenesis and ii) secondary amplification (Castaneda et al., 2011). In *Drosophila*, primary piRNA biogenesis occurs via the interaction of the RNA helicase Armitage (Arm), a Tudor domain-containing protein (Yb), the putative exonuclease Zucchini (Zuc), and Piwi (Brennecke et al., 2007, Olivieri et al., 2010). Long piRNA precursors are transcribed from piRNA

clusters and are processed by Arm-Yb-Zuc into primary piRNAs, which are bound by Piwi and used as guides to mediate exogenous transposon transcript cleavage by the Piwi slicer domain. In the *Drosophila* germline, secondary piRNAs are generated via a unique pathway known as "ping pong" biogenesis (Brennecke et al., 2007). Aubergine (Aub) – in complex with a primary piRNA – binds repeat element mRNA, cleaving the transcript 10 nucleotides upstream to generate the 5' end of a secondary piRNA. Primary piRNAs display a conserved bias for a 5' uracil (1U) (Saito et al., 2006, Girard et al., 2006, Aravin et al., 2006, Lau et al., 2006), hence secondary piRNAs exhibit a bias for adenine 10 nucleotides downstream of the 5' end (10A), the complementary binding site of the 5'U of the primary piRNA; this sequence feature is unique to ping pong biogenesis in germ cells (Brennecke et al., 2007). In *Drosophila, s*econdary piRNAs complex with Ago3, which in turn cleaves primary piRNA precursors to generate new primary piRNAs, forming a heterotypic amplification loop. Following piRNA loading onto the Piwi protein, bases from the 3' ends are trimmed (Kawaoka et al., 2011, Vourekas et al., 2012) and are 2'-O-methylated for stability (Saito et al., 2007).

Murine piRNAs in pre-pachytene male germ cells also exhibit hallmarks of ping pong biogenesis (1U and 10A bias); it was originally thought that this occurred via an interaction between Mili and Miwi2 in a heterotypic amplification loop (Aravin et al., 2007b, Aravin et al., 2008). However, it has been shown that once loaded with a secondary piRNA, Miwi2 localises to the nucleus to facilitate epigenetic silencing, hence is unlikely to participate in piRNA amplification (Carmell et al., 2007). In fact, it has been shown that only Mili is required for a homotypic amplification loop; this process is required for silencing of LINE1 retrotransposon transcripts (De Fazio et al., 2011). Furthermore, the expression of Miwi2-bound piRNAs is dependent on the presence of Mili (Aravin et al., 2008). In the mouse, primary piRNAs are produced mainly from uni-strand clusters (Li et al., 2013). The currently-accepted model of pre-pachytene piRNA biogenesis in male mouse germ cells is depicted in Figure 1-4.

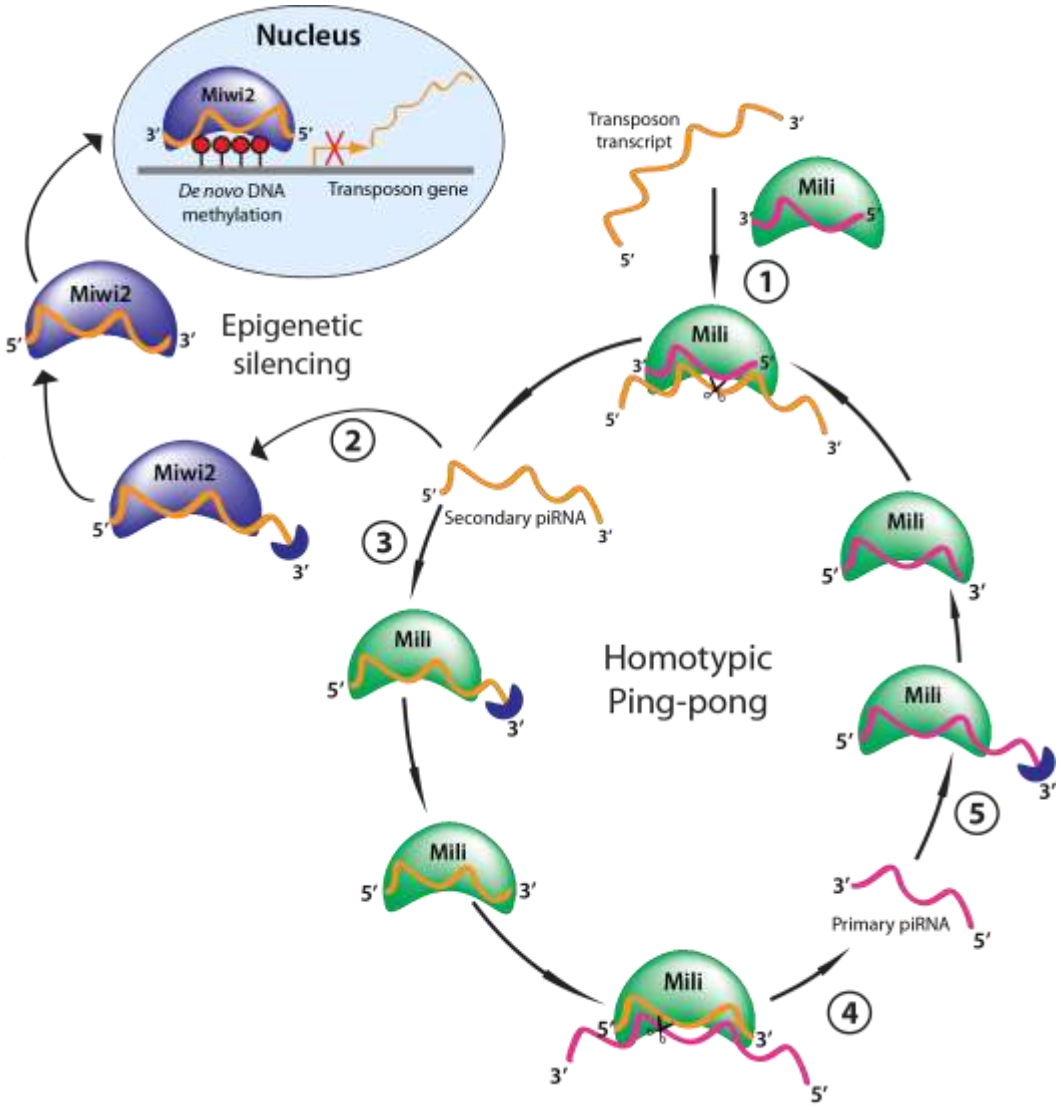


Figure 1-4: piRNA ping pong biogenesis in the cytoplasm of murine germ cells

(1) Mili uses a bound primary piRNA (magenta) as a sequence-specific guide to bind transposon transcripts, which is cleaved by the Mili slicer domain, resulting in a secondary piRNA (orange). (2) After binding a secondary piRNA, Miwi2 localises to the nucleus to perform a *de novo* methylation transposon silencing function. (3) Other secondary piRNAs are bound by Mili and 3' ends trimmed by a 3' exonuclease. Mili uses the secondary piRNA as a guide to bind complementary nascent transposon transcripts. (4) Cleavage of the transposon transcript by Mili slicer domain results in primary piRNAs, which are bound by Mili. Mili-bound primary piRNAs are subjected to 3' end trimming by a 3' exonuclease and the ping pong piRNA amplification cycle repeats. Adapted from (Ishizu et al., 2012).

Certain Piwi-bound piRNA subsets are generated independently of ping pong biogenesis. For example, *Drosophila* Piwi – but not Aub and Ago3 – is expressed in the somatic cells of the ovary where it appears to have a function quite distinct from that in germ cells. Somatic piRNAs bound to Piwi are generated independently of ping pong amplification and target a different subset of transposable elements to that targeted by Piwi-bound germline piRNAs (Pelisson et al., 2007, Malone et al., 2009). Ping pong-independent piRNAs have also been shown to originate from the 3'UTR of protein-coding genes in the murine testis; these tend to exhibit a bias for 1U, but not for 10A (Robine et al., 2009). 3'UTR piRNAs associate with Mili in the pre-pachytene mouse testis and with Mili and Miwi in the pachytene testis and are produced via primary processing involving transcription from large genomic clusters (Robine et al., 2009; Beyret et al., 2012).

The different *Drosophila* Piwi proteins bind mostly unique piRNA subsets: most Piwi- and Aub-bound piRNAs are antisense to repeat elements, where most piRNAs associating with Ago3 are sense; there is also a difference in piRNA size profiles associated with particular Piwi proteins (Brennecke et al., 2007). Evidence for Piwi protein piRNA subset specificity in the mouse comes from differential size preferences: Miwi-bound pachytene piRNAs tend to be 30-31nt long (Girard et al., 2006); Mili binds 26-28nt piRNAs (Aravin et al., 2006), and Miwi2 mostly associates with 27-29nt piRNAs (Aravin et al., 2008). The Mili- and Miwi2-bound pre-pachytene piRNAs tend to map to transposons and repeat elements (Aravin et al., 2008), whereas the subset of Miwi-bound pachytene piRNAs are derived from intergenic regions devoid of repeat sequences (Vourekas et al., 2012). In the adult testis, Miwi also directly binds translationally-inactive spermiogenic mRNAs without the presence of a guide piRNA (Vourekas et al., 2012). The function of Miwi-bound RNA is, as yet, unclear. There is some evidence of Piwi redundancy in *Drosophila*, where *roo* rasiRNAs associate both with Piwi and Aub, suggesting that either Piwi or Aub could potentially compensate for the loss of the other without detrimental effects to the cell (Vagin et al., 2006). At the time of writing, no such evidence for Piwi protein redundancy had been reported in mammals. Mice with knockout of all three Piwi proteins do not display overt defects in development apart from male infertility, suggesting that mice with mutations in a single Piwi protein do not exhibit redundancy by the other Piwi family members to maintain somatic integrity (Nolde et al., 2013).

The biogenesis and function of piRNAs in *Caenorhabditis elegans* is different to *Drosophila* and mammals. There are two Piwi homologues in *C. elegans*, PRG-1 and PRG-2, although PRG-2 does not appear to have a function (Batista et al., 2008, Das et al., 2008, Bagijn et al., 2012). There is no

evidence for ping pong amplification of piRNAs (Das et al., 2008). PRG-1 binds 21nt piRNAs called 21U-RNAs, which are classified into two types: motif-dependent and motif-independent piRNAs (Weick et al., 2014). Motif-dependent piRNAs are encoded by two large clusters on Chromosome IV and are interspersed with protein-coding genes; these piRNAs map to intergenic regions and introns (Ruby et al., 2006, Batista et al., 2008, Bagijn et al., 2012). They exhibit an octomeric sequence motif known as the "Ruby motif" 40bp upstream of the 5' end of the piRNA sequence; this is recognised by specific transcription factors for transcription of 26nt piRNA precursors (Ruby et al., 2006, Gu et al., 2012, Weick et al., 2014). At the time of writing, only one factor essential for motif-dependent piRNA biogenesis had been described (Weick et al., 2014). Motif-independent piRNAs have been recently described and map to regions outside of the two main piRNAs clusters; they also lack the upstream Ruby motif although little is known about the factors involved in their biogenesis (Gu et al., 2012, Weick et al., 2014). Both types of *C. elegans* piRNAs indirectly silence target transcripts via activation of a secondary endogenous siRNA (endosiRNA) pathway (Bagijn et al., 2012, Weick et al., 2014).

Germline functions of the Piwi/piRNA pathway during development

A role in germline stem cell self-renewal in Drosophila and C. elegans

Oogenesis in *Drosophila* provides a simple, yet powerful model for the study of germline development. Unipotent germline stem cells (GSCs) within the developing ovary undergo asymmetric cell division, giving rise to both a self-renewing daughter stem cell and a differentiated cystoblast (Lin and Spradling, 1997). This cystoblast is enveloped by an epithelial monolayer of follicle cells, and together, these cells bud off to form an egg chamber; multiple budding events form a string of egg chambers at progressive stages of development (Margolis and Spradling, 1995). These oogenic events are disrupted upon mutation of *Piwi*: Lin and Spradling (Lin and Spradling, 1997) observed that mutation of *Piwi* eradicated egg chamber budding, inducing female sterility. Defects in spermatogenesis are also observed in male *Piwi*-null flies (Cox et al., 1998, Cox et al., 2000) and in mutants of the *C. elegans* Piwi-like gene *Prg-1* (Wang and Reinke, 2008, Batista et al., 2008, Das et al., 2008). These gametogenic defects in *Drosophila* and *C. elegans* can be attributed to germline stem cell differentiation without self-renewal (Cox et al., 1998) and provided the first insights into function of the Piwi proteins.

Piwi proteins are required for murine spermatogenesis

Murine spermatogenesis consists of multiple phases: the mitotic phase involves self-renewal of spermatogonia to yield primary spermatocytes, which then undergo meiosis to generate haploid cells, followed by spermiogenesis to yield mature spermatozoa. Homozygous knockout of any of the murine Piwi proteins results in sterility in males, attributed to arrests at distinct stages of spermatogenesis. *Mili*-null germ cells predominately arrest as pre-meiotic spermatogonia A, with the exception of several cells that form spermatocytes that are blocked at the zygotene or early pachytene phase of Meiosis I (Kuramochi-Miyagawa et al., 2004). *Miwi2* mutants display spermatogenic arrest as self-renewing spermatogonia A to some extent, but predominately during the leptotene stage of Meiosis I (Carmell et al., 2007). In contrast, *Miwi*-knockout mouse sterility is attributed to halted development during spermiogenesis at the post-meiotic round spermatid stage (Deng and Lin, 2002). Table 1-4 lists several of the Piwi-null phenotypes observed in higher organisms.

Table 1-4: Phenotypic effects of Piwi knockout in higher eukaryotes.

Adapted from (Juliano et al., 2011).

Species	Piwi protein	Null phenotype	References
D. melanogaster	Piwi	Male, female sterility	(Cox et al., 1998, Cox et al., 2000)
	Aub	Male, female sterility	(Schupbach and Wieschaus, 1991, Harris and Macdonald, 2001, Wilson et al., 1996)
	Ago	Male sterility Female semi-fertility	(Li et al., 2009)
M. musculus	Miwi	Male sterility	(Deng and Lin, 2002)
	Mili	Male sterility	(Kuramochi-Miyagawa et al., 2004)
	Miwi2	Male sterility	(Carmell et al., 2007)
	Miwi/Mili/Miwi2	Male sterility; No apparent defects in haematopoiesis	(Nolde et al., 2013)
C. elegans	Prg-1	Sterile at 25°C	(Wang and Reinke, 2008)
	Prg-2	No reported defects in development	(Wang and Reinke, 2008)

Piwi-mediated transposon silencing: transcript cleavage and DNA methylation

Despite its prominent role in germline stem cell maintenance, the most widely studied function of Piwi proteins is the silencing of transposable elements in the animal germline. Such function for the Piwi/piRNA pathway was originally described in *Drosophila,* where *Piwi* mutation results in the mobilisation and accumulation of LTR transposon transcripts in the male germline (Kalmykova et al., 2005). As part of the ping pong amplification loop pathway, piRNAs are used as sequence guides to target endogenous and exogenous transposon transcripts for cleavage by effector Piwi proteins: Aubergine in *Drosophila,* and Mili in the mouse (Brennecke et al., 2007, Aravin et al., 2008). Piwi/piRNA-mediated defence against transposons also occurs at the level of transcription. The pathway is required for stable, epigenetic silencing of the transposon DNA. In the mouse, knockout of *Mili* or *Miwi2* results in a global loss of cytosine methylation at transposon sequences and a phenotype reminiscent of mice deficient for *Dmnt3L*, a DNA methyltransferase involved in *de novo* DNA methylation (Aravin et al., 2007b, Kuramochi-Miyagawa et al., 2004, Carmell et al.,

2007). At the *de novo* DNA methylation stage of spermatogenesis in mice, the majority of Miwi2-bound piRNAs and around half of those associated with Mili map to LTR and LINE transposons (Aravin and Bourc'his, 2008). It is thought that these small RNAs act – via effector Piwi proteins – as specificity guides for deposition of repressive epigenetic marks. It has also recently been shown that Miwi2 is required in male germ cells for transcriptional repression of LINE1 – but not LTR – retrotransposons via tri-methylation at Histone 3 Lysine 9 (H3K9) (Pezic et al., 2014).

The concept of RNA-mediated epigenetic silencing has precedent in plants: RNA-directed *de novo* DNA methylation was first described by Wassenegger and colleagues (Wassenegger et al., 1994) and was later shown to be mediated by small RNAs (Mette et al., 2000). Furthermore, RNAi-mediated silencing and H3K9 methylation of heterochromatic repeat sequences in centromeric regions is well established in the fission yeast *Saccharomyces pombe* (Volpe et al., 2002). Transposon repression is also seen in the cnidarian *Hydra*, which expresses two Piwi proteins, Hywi and Hyli (Lim et al., 2013). There is also evidence that piRNAs are transmitted through the female germline to maintain stable transposon silencing in the developing progeny of *Drosophila* (Shpiz and Kalmykova, 2009). Moreover, Piwi is required for transposon silencing in *C. elegans* and that silencing - triggered by piRNAs - can be inherited and maintained for over 20 generations (Das et al., 2008, Ashe et al., 2012).

Piwi function is dependent on subcellular localisation in germ cells

Depending on subcellular location, Piwi proteins can function quite differently, particularly in regards to its role in GSC maintenance versus transposon suppression. This is best demonstrated by *Drosophila* $Piwi^{Nt}$ mutants, where *Piwi* lacks a nuclear localisation signal, thus restricting expression to the germ cell cytoplasm (Klenov et al., 2011). In $Piwi^{Nt}$ mutants, *Piwi*-mediated silencing functions are abolished, with severe transposon de-repression reminiscent of *Piwi*-null flies (Cox et al., 1998, Cox et al., 2000). However, unlike in *Piwi*-null *Drosophila*, the loss of nuclear localisation of Piwi protein in $Piwi^{Nt}$ mutants has little effect on GSC maintenance (Klenov et al., 2011). In these mutants, oogenesis is not disrupted: there is a normal number of budding egg chambers within a developing ovary, consistent with normal asymmetric GSC division. Another example is the behaviour of Miwi2 in Mili-null mouse testes: Miwi2 – normally localised to the nucleus – relocates to the cytoplasm (Figure 1-5) and no longer binds piRNAs; whether it is still functional is unknown (Aravin et al., 2008). Based on these observations, it is likely that each of the Piwi proteins have more than one function, linked to subcellular localisation.

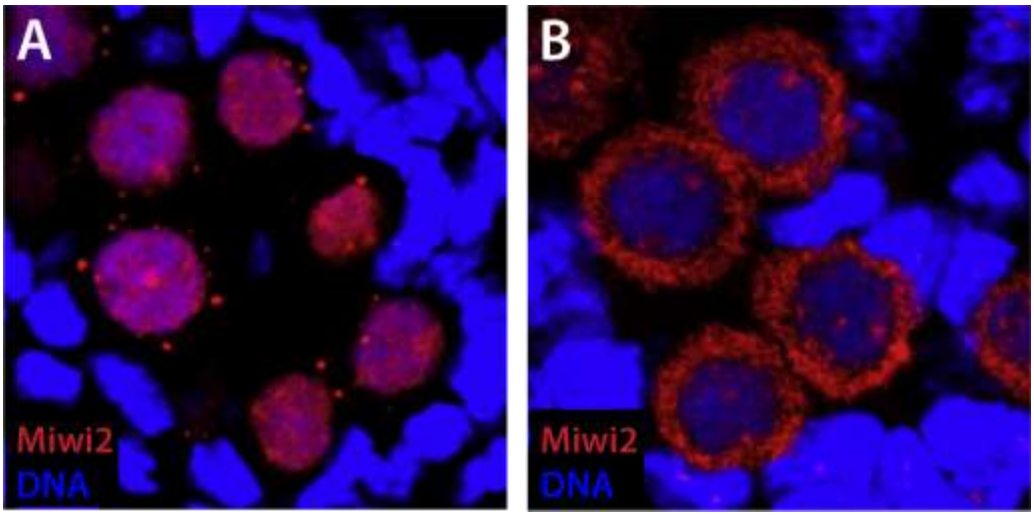


Figure 1-5: Miwi2 protein expression in wild-type and Mili-null mouse testis

Miwi2 expression in (A) wild-type and (B) Mili-knockout mice. Miwi2 expression is shown in red, DNA in blue. Obtained from (Aravin et al., 2008) and used with permission from Elsevier.

Somatic functions of the Piwi/piRNA pathway

Piwi is involved in epigenetic gene regulation outside of the germline

While the Piwi/piRNA pathway is known almost solely for its role in the germline in complex animals, it is also expressed and functional in the somatic cells of more primitive organisms, such as *Schmidtea mediterranea, Pleurobrachia pileus, Ephydatia fluviatilis*, *Botrylloides leachi, Podocoryne carnea,* and *Aplysia* (Reddien et al., 2005, Alie et al., 2011, Funayama et al., 2010, Rinkevich et al., 2010, Seipel et al., 2004, Rajasethupathy et al., 2012). The activity of the Piwi system in these primitive organisms demonstrates that this is an ancient pathway; primitive organisms thus have great potential to yield insight into Piwi gene and protein function. For example, in the sea slug, *Aplysia*, Piwi is abundant in neurons, where it is restricted to the nucleus and binds a diverse population of 28nt piRNAs (Rajasethupathy et al., 2012). This group showed that a subset of *Aplysia* piRNAs mediates cytosine methylation of the promoter region of *Creb2*, a gene integral for the regulation of neural plasticity and a repressor of memory. This gene silencing function in *Aplysia* is reminiscent of Piwi protein-mediated adaptive response to silence transposons in the mammalian germline (Aravin et al., 2007a). The function of Piwi orthologues in other organisms in the maintenance of stem-like somatic cells will be discussed.

An ancient role in stem cell-mediated growth and regeneration in non-mammalian organisms

There is mounting evidence from primitive organisms that points to a role for Piwi genes and proteins in maintenance of somatic stem cell self-renewal. Piwi orthologues in the planarian *Schmidtea mediterranea* have been implicated in the regulation of somatic stem-like cells (Reddien et al., 2005). Neoblasts are the only actively-dividing cells present in *S. mediterranea* and comprise approximately 30% of all cells; they are thought to be involved in regeneration following injury, as they are capable of giving rise to all somatic and germ cell types (Baguna, 1989). The Piwi homologues *Smedwi-1* and *Smedwi-2* are exclusively expressed by neoblasts and elimination of neoblasts via irradiation results in a loss of Piwi gene expression (Sanchez Alvarado et al., 2002, Reddien et al., 2005). In addition, RNAi inhibition of *Smedwi-2* and a more recently-discovered homologue, *Smedwi-3,* results in defects in regeneration post-injury reminiscent of irradiated animals lacking dividing neoblasts (Figure 1-6) (Reddien et al., 2005, Palakodeti et al., 2008). These data show that *Smedwi-2* and *-3* are required for maintenance of the stem-like properties of neoblasts in the planarian. Notably, RNAi against *Smedwi-1* did not impede regeneration, suggesting an independent function for this Piwi protein in *S. mediterranea*. pi-sized RNAs have

been identified in *S. mediterranea* with diverse genomic origin: repeats, genic and intergenic regions (Palakodeti et al., 2008).

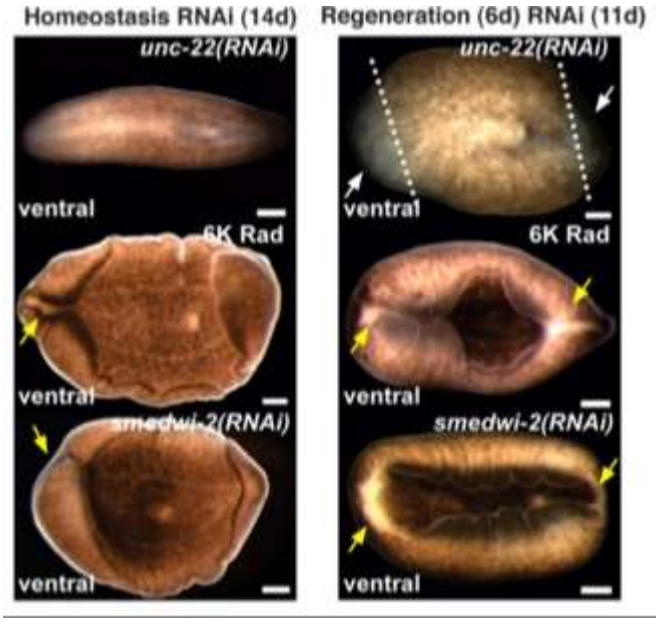


Figure 1-6: Defects in homeostasis and regeneration following *Smedwi-2* knockdown in *S.mediterranea*

Left panel: homeostasis after 14d. Right panel: regeneration after 11d following head and tail amputation (white dotted lines). Defective head regression indicated by yellow arrows; white arrows indicate regenerative cells (blastema). Animals were either i) treated with *C. elegans unc-22* negative control siRNA (top), ii) irradiated at 6000 rad (6K Rad), eradicating neoblasts and resulting in defects in regeneration (middle), or iii) treated with *Smedwi-2* RNAi (bottom). There are common defects in irradiated and *Smedwi-2* RNAi-treated animals consistent with neoblast dysfunction. Obtained from (Reddien et al., 2005). Used with permission from the American Association for the Advancement of Science (AAAS).

Although a direct association between *S. mediterranea* pi-sized RNAs and Smedwi proteins has yet to be established, they share several features with *bona fide* piRNAs: marked population diversity, specific size (32nt), and bias towards a 5'U. Palakodeti and colleagues observed that these piRNAs are mostly confined to the neoblasts. Notably, these pi-sized RNAs are present in immature, asexual *S. mediterranea* that lack testes and ovaries in similar proportions to those in sexually mature animals; furthermore, *Smedwi* genes do not appear to be expressed in the planarian's few presumptive germ cells at all (Palakodeti et al., 2008). Knockdown of *Nanos* – required for germ cell viability (Wang et al., 2007) – did not affect production of pi-sized RNAs, confirming that *Smedwi* genes are restricted to somatic cells (Reddien et al., 2005, Palakodeti et al., 2008).

Additional evidence for Piwi expression in somatic stem cells comes from the primitive comb jelly, *Pleurobrachia pileus*. This ctenophore has a number of unique anatomical traits, including both male and female gonads, a set of eight longitudinal comb rows, and a pair of predatory tentacles. Regeneration of the tentacles is an ongoing process as a result of damage and injury during predation upon planktonic crustaceans. Differentiated muscle cells and colloblasts, a type of adhesive predatory cell unique to ctenophora, are generated by a group of self-renewing, undifferentiated cells at the tentacle root (originally reported by (Hertwig, 1880), extensively described by (Franc, 1978)). A Piwi orthologue in *P. pileus, PpiPiwi1*, while strongly expressed throughout the male and female gonads, is also co-expressed in the undifferentiated muscle and colloblast progenitors of the tentacle root and the resident stem-like cells of the comb rows, and expression decreases with cell maturation (Alie et al., 2011).

Regenerative capabilities from undifferentiated tissue have long been recognised in the sponge *Ephydatia fluviatilis* (Wilson, 1907), presenting an ideal model for the investigation of stem cell self-renewal. The stem-like cells of the sponge, known as the archeocytes, have aggregative properties and are indispensable for regeneration from dissociated fragments (De Sutter, 1977). Proliferating archeocytes in *E. fluviatilis* express the Piwi orthologues *EfPiwiA* and *EfPiwiB*, and *EfPiwiA* expression declines upon maturation of the archeocytes (Funayama et al., 2010). However, expression is maintained in mature choanocytes, which are downstream self-renewing cells that undergo spermatogenesis, giving rise to male gametes (Paulus, 1986). Therefore, it is possible that *EfPiwiA* has a dual role in germ line development *and* archeocyte self-renewal (discussed in (Funayama, 2010)).

Colonial Botryllid ascidians are the only known chordates to undergo whole body regeneration (WBR). Isolated blood vessel fragments from *Botrylloides leachi* are capable of aggregating and forming regeneration niches 2-3 days post-separation (dps) from the body and undergo complete regeneration after 10-14 dps (Brown et al., 2009). It has been speculated that this process of WBR is mediated by a small population of stem-like cells circulating in the blood stream of *B. leachi* (Rinkevich et al., 1995). It has been demonstrated that the *B. leachi* Piwi orthologue, *Bl-Piwi*, is involved in WBR (Rinkevich et al., 2010). *Bl-Piwi* is not expressed in the mature blood cells or vascular epithelium at separation. But Rinkevich and colleagues showed that *Bl-Piwi* is *de novo* synthesised ~2 dps in cells that normally lie dormant within the vascular epithelium; the *Bl-Piwi*$^{+}$ cells invade regeneration niches, line the vascular epithelia, and proliferate, contributing to WBR. The number of aggregates of *Bl-Piwi*-expressing cells within regeneration niches peaks at ~10 dps, at which point *Bl-Piwi* levels begin to decline, concurrent with the expression of differentiation genes and formation of functional zooids (Rinkevich et al., 2010). This group also demonstrated that RNAi against *Bl-Piwi* arrests cellular proliferation and prevents WBR, with the formation of functional zooids completely abolished.

Piwi is also implicated in transdifferentiation and regeneration events in the cnidarian, *Podocoryne carnea.* The cnidarian Piwi orthologue, *Cniwi*, is expressed through the life cycle of *P. carnea* and is enriched in the developing egg and the gonads of the free-swimming adult medusa (Seipel et al., 2004). During normal development, expression of *Cniwi* is at its lowest at the late larval stage, concomitant with a reduction in proliferation and acquisition of terminal differentiation. Isolated, cultured striated muscle from *P. carnea* transdifferentiates *in vitro* to smooth muscle, performs DNA synthesis and proliferates, regenerating the organism (Schmid, 1975, Schmid and Alder, 1984, Alder and Schmid, 1987). During the first several hours of *in vitro* regeneration, isolated muscle fragments transiently express *Cniwi*. During the transdifferentiation phase (3h-3d) – a period of increased cellular proliferation – *Cniwi* expression steadily increases in proliferating cells until they begin to terminally differentiate into smooth muscle. Thus, it appears that upon reactivation in differentiated somatic cells, *Cniwi* acts as a stem cell gene to reintroduce self-renewal abilities during transdifferentiation and regeneration.

Mammalian stem cell potency

Stem cells are a diverse cell population with infinite potential for self-renewal (Amit et al., 2000). Stem cells possess the ability to asymmetrically divide, yielding i) an identical self-renewing daughter cell and ii) one committed to differentiation. Stem cells from different stages of development have different "potency", or differentiation potential. The fertilised zygote undergoes multiple rounds of cell division, resulting in a pre-implantation blastocyst (Chazaud et al., 2006, Niwa et al., 2005, Yamanaka et al., 2006). The zygote and inner cell mass (ICM) of the pre-implantation blastocyst contain totipotent cells, which are capable of forming all germ layers of the organism as well as the trophoblast - extra-embryonic tissues required to support embryonic development, including the placenta (Surani et al., 2007). Pluripotent cells are able to contribute to all three embryonic germ layers in the developing embryo: the ectoderm, endoderm and mesoderm (Odorico et al., 2001). The ICM of the post-implantation blastocyst contains pluripotent epiblast stem cells, which develop into founder germ cells (primordial germ cells) or progress along the somatic lineage (embryonic stem cells, or ES cells) (Surani et al., 2004). Multipotent stem cells retain a degree of plasticity within a designated lineage and are present in adult tissues within designated stem cell "niches" (Watt and Hogan, 2000, Li and Xie, 2005, Spradling et al., 2001, Lin, 2002). For example, self-renewing human $CD34^+$ haematopoietic stem cells differentiate to a range of lymphoid and myeloid progeny (Till and McCulloch, 1961, Till and McCulloch, 1964). Unipotent stem cells - such as germline stem cells, which only give rise to gametes – are completely lineage-restricted (Lehmann, 2012). The multiple states of stem cell potency are illustrated in Figure 1-7.

ES cells are distinguished at a molecular level from their differentiated counterparts by unique gene expression and epigenetic signatures (Jaenisch and Young, 2008). The most widely known and best studied ES cell-specific proteins are the homeodomain-containing transcription factors Nanog and Oct4 (Chambers et al., 2003, Chambers and Smith, 2004, Mitsui et al., 2003, Nichols et al., 1998) and the HMG box-containing transcription factor Sox2 which heterodimerises with Oct4. Oct4, Sox2, and Nanog co-occupy hundreds of gene promoters to drive expression of downstream transcription factors, mediating the balance between self-renewal, pluripotency and cell fate determination during tissue development (Boyer et al., 2005, Loh et al., 2006). In addition, each of these three "master" pluripotency transcription factors are encompassed in an auto-regulatory positive feedback loop whereby each factor binds to its own promoter and that of the other

master factors (Boyer et al., 2005, Loh et al., 2006). Oct4 is required for maintaining the pluripotent state in ES cells but Nanog and Sox2 are not (Nichols et al., 1998, Chambers et al., 2007, Masui et al., 2007).

Given that Piwi genes are required for stem cell function in primitive organisms such as *S. mediterranea* (Reddien et al., 2005) and that they are highly conserved, it is possible that the Piwi/piRNA pathway is also required for maintaining stem cells in mammals. A Piwi orthologue is required for whole body regeneration in *B. leachi* (Rinkevich et al., 2010), implying that Piwi genes and proteins are required at the earliest stages of development, perhaps in pluripotent stem cells. The validity of this hypothesis requires that Piwi genes and proteins are expressed in the somatic cells of higher eukaryotes; multiple studies have reported this phenomenon.

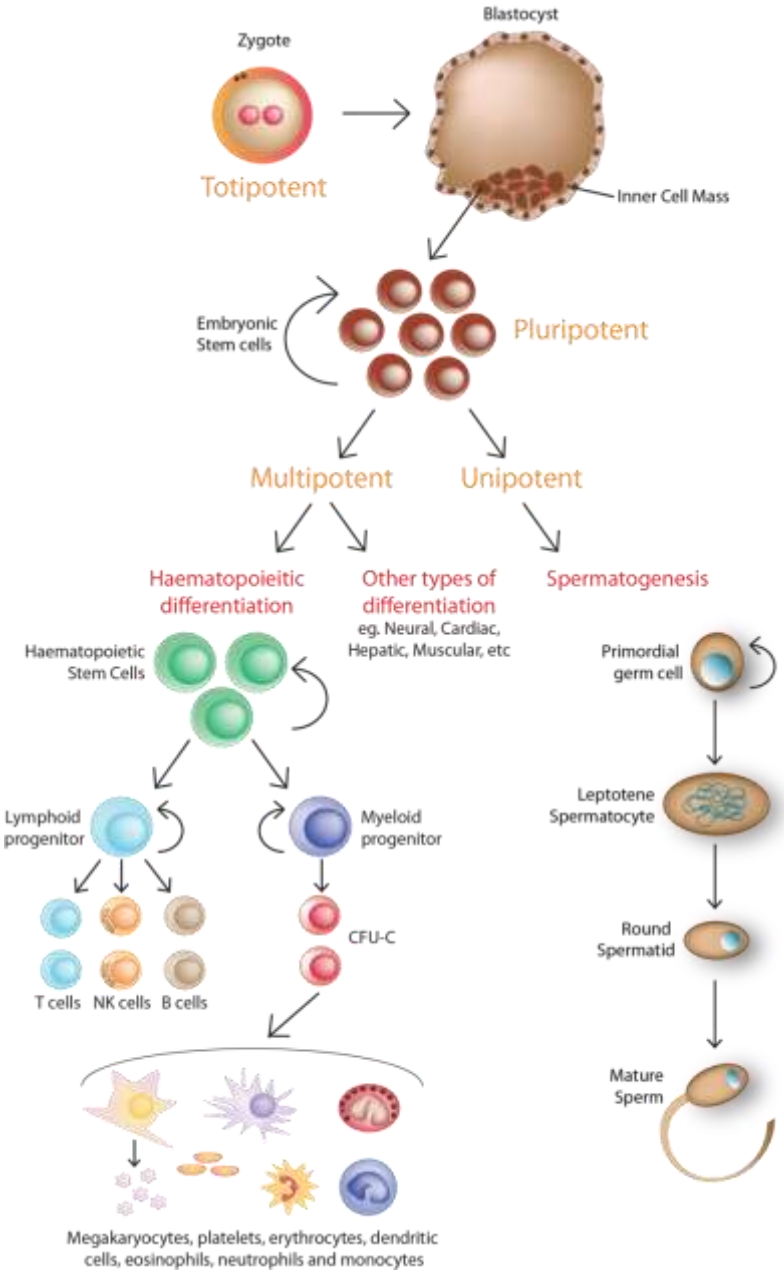


Figure 1-7: Mammalian stem cell types and their differentiated progeny

The zygote, formed from the fusion of gametes, is totipotent with the potential to form all embryonic and extra-embryonic tissues. Embryonic stem cells are contained within the inner cell mass of the blastocyst and are pluripotent, with the potential to form all three germ layers. Multipotent stem cells – such as haematopoietic stem cells, and lymphoid and myeloid progenitors – are able to form multiple differentiated cell types. Unipotent stem cells – such as primordial germ cells – are able to give rise to a single differentiated cell type.

Piwi genes and proteins are expressed in somatic cells of higher eukaryotes

Somatic Piwi gene expression has been described in multiple tissues of higher eukaryotes. For example, *Drosophila Piwi* – but not *Aub* or *Ago3* – is expressed in the somatic cells of the ovary where it appears to have a function distinct from that in germ cells. Somatic piRNAs bound to *Piwi* are generated independently of ping pong amplification and target a different subset of transposable elements to Piwi-bound germline piRNAs (Pelisson et al., 2007, Malone et al., 2009). *Piwi*, *Ago3*, and *Aub* are each expressed in specific regions of the *Drosophila* brain and their mutation results in increased brain transposon expression (Perrat et al., 2013). Prior to the discovery of piRNAs, *Hiwi* was reported to be expressed in human $CD34^+$ haematopoietic stem cells (Sharma et al., 2001). Notably, *Hiwi* is not expressed in $CD34^-$ bone marrow cells, which have low haematopoietic potency. Furthermore, *Hiwi* expression diminishes as the $CD34^+$ population differentiates. This is reminiscent of the pattern of Piwi gene and protein expression during progenitor cell differentiation in ancient metazoans. More recently, the presence of another Piwi orthologue, Mili (and its bound piRNAs), has been reported in adult murine mesenchymal stem cells (Wu et al., 2010). In addition, *Miwi2* has been reported to be expressed in haematopoietic stem cells, multipotent progenitor cells, and granulocyte-macrophage progenitors in the mouse bone marrow but not in committed haematopoietic cells, although it does not appear to be required for haematopoiesis (Nolde et al., 2013). Furthermore, Miwi is expressed in the mouse hippocampus, where it binds piRNAs (Lee et al., 2011). A large-scale interrogation of over 130 next-generation sequencing datasets from *Drosophila*, mouse, and rhesus macaque revealed that pi-like RNAs are present in multiple somatic tissues including *Drosophila* embryo and head, mouse pancreas and haematopoietic cells and macaque epididymis, seminal vesicles, and cerebral cortex (Yan et al., 2011). Notably, pi-like RNAs in the murine pancreas and macaque epididymis are present in similar abundance to that observed in germ cells in these animals. Piwi orthologue transcripts are also detectable in macaque tissues, and Miwi protein is present in the mouse pancreas. Furthermore, our own group has recently found that *Hiwi2* is expressed ubiquitously in human somatic tissues, including brain, colon, liver, heart, kidney, liver, spleen, placenta, and thymus, among other tissues (Sugimoto et al., 2007, Keam et al., 2014).

Expression of Piwi genes and proteins is required for somatic stem cell self-renewal and stem cell-mediated regeneration in primitive organisms, and, as outlined above, reports are emerging that suggest a somatic function for Piwi/piRNA in higher eukaryotes. It is currently unclear whether

Piwi genes and proteins maintain stem-like cells during development and homeostasis in mammals, although evidence is accumulating that they function in cancer.

Piwi/piRNA in cancer cells

Cancer cells exhibit stem-like qualities

Early cancer studies reported that tumours are phenotypically and functionally heterogeneous (Fidler and Kripke, 1977, Heppner, 1984, Nowell, 1986). Many types of cancer exhibit a hierarchy of undifferentiated and differentiated cells reminiscent of stem cell hierarchies in normal tissues; these include some germ lineage cancers (Kleinsmith and Pierce, 1964) and myeloid leukaemias (Fearon et al., 1986, Ogawa et al., 1970). Many have drawn parallels between stem cells and cancer cells, to the point where some have proposed the idea of "differentiation therapy" to treat aggressive cancers containing undifferentiated, tumorigenic cells (Sell, 2004).

Both multipotent stem and cancer cells reside within designated stem cell "niches": stem cells rely on precise signalling from their microenvironment to mediate a controlled balance between self-renewal and differentiation, although this balance is dysregulated in cancer and can lead to uncontrolled proliferation and tumorigenesis (Watt and Hogan, 2000, Li and Xie, 2005, Spradling et al., 2001, Lin, 2002, Li and Neaves, 2006, Malanchi et al., 2012). Furthermore, common pathways are active in both stem and cancer cells, such as Polycomb silencing. ES cells rely on Polycomb Group (PcG) proteins to epigenetically silence the expression of differentiation genes to maintain pluripotency (Ringrose and Paro, 2004) and in cancer, the promoters of PcG target genes are more likely to be silent and methylated than non-PcG targets (Widschwendter et al., 2007). Widschwendter et al suggest that cancerous cells may retain a stem cell memory and relinquish Polycomb silencing of critical tumour suppressor regions. In addition, Wnt signalling is required for development and stem cell-mediated regeneration in normal tissues (Korinek et al., 1998, Batlle et al., 2002) and in tumours such as colon cancer (van de Wetering et al., 2002). Abnormal Wnt signalling has also been implicated in epidermal cancer and leukaemia, amongst numerous other malignancies (Reya and Clevers, 2005). Furthermore, the Sonic Hedgehog (Shh) and Notch signalling pathways – which are required for stem cell self-renewal in normal mammalian tissues (Bhardwaj et al., 2001, Wechsler-Reya and Scott, 1999, Varnum-Finney et al., 2000) – are commonly dysregulated in cancers, including medulloblastoma (Wechsler-Reya and Scott, 2001),

basal cell carcinoma (Gailani and Bale, 1999), and T-cell leukaemia (Ellisen et al., 1991). These observations have led to the emergence of the "cancer stem cell" (CSC) hypothesis and the concept of a "tumour-initiating cell" (reviewed in (Medema, 2013)).

Cancer stem cells

There are two major models of tumorigenesis (Figure 1-8): the stochastic model proposes that all cells within a cancer can give rise a secondary tumour and that this is influenced by random intrinsic and extrinsic factors; the hierarchy model assumes that a unique subset of cells within a cancer bears the capacity for tumour initiation and thus are solely responsible for metastasis (Dick, 2009, Nguyen et al., 2012). The CSC hypothesis is based on the hierarchy model and proposes that the tumorigenicity and metastatic potential of cancer is directly correlated with the presence of a resident self-renewing stem-like cell population (reviewed in (Visvader and Lindeman, 2008)).

The first report of a cancer stem cell-like population was in 1940, where Milton D. Bosse reported the presence of differentiated muscle tissue – presumably resulting from aberrant differentiation of stem-like cells – in pulmonary metastases from testis teratocarcinoma (Bosse, 1940). Later, it was observed that teratocarcinomas indeed comprise a heterogeneous mix of differentiated tissue and undifferentiated stem-like cells (Jackson, 1941).

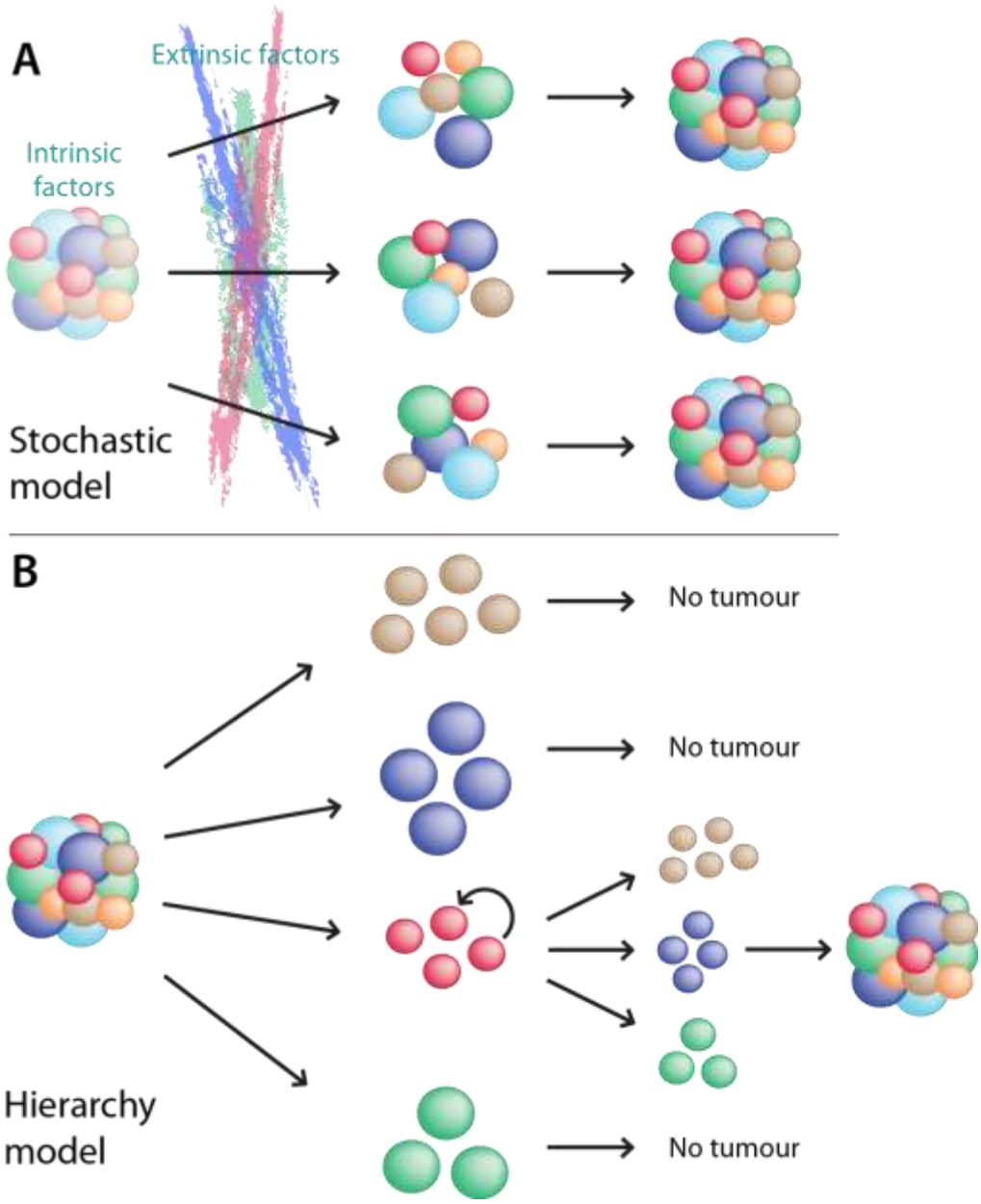


Figure 1-8: Two models of tumour formation: stochastic model and cellular hierarchy model

(A) The stochastic model assumes that all cells within a tumour are biologically equivalent, but the influence of stochastic intrinsic and extrinsic factors determines function; therefore, tumour-initiating potential can not be enriched. (B) The hierarchy model predicts that only a small subset of tumour cells is able to undergo asymmetric division and initiate tumorigenesis; tumour-initiating activity can therefore be enriched by cell sorting for intrinsic factors. In theory, the cycle of cell enrichment and tumour initiation can be propagated infinitely. Adapted from (Dick, 2009).

William Dameshek first proposed that leukaemia arises from a haematopoietic progenitor cell with multi-lineage differentiation potential (Dameshek, 1951). Later, the first conclusive evidence of cancer stem cells came from Fialkow et al: chronic myeloid leukaemia stem cells arise via a multi-step clonal process from normal blood HSCs, generating progeny with enriched proliferative potential (Fialkow et al., 1967). Over time, these out-compete regular haematopoietic cells in the blood, resulting in chronic disease (reviewed in (Sloma et al., 2010)). But how do these leukaemic cancer stem cells originate in the first place? One group has described a population of cells with the potential for both malignant and benign differentiation within a murine dendritic cell-like leukaemia (Chen et al., 2007). The cells, termed "precancerous stem cells" or pCSCs, have the ability to rapidly develop aggressive solid and leukaemic tumours following injection into immunodeficient mice. It is conceivable that these pCSCs are derived from normal stem cells present within the tissues of a developed adult. Stochastic environmental events could deleteriously alter stem cell fate, resulting in malignancy. This idea is supported by the fact that leukaemia stem cells share similar cell surface markers to normal haematopoietic stem cells (Bonnet and Dick, 1997, Bhatia et al., 1997).

Similar to that observed in leukaemia, hierarchic cell heterogeneity is also observed the human brain tumour glioblastoma. Singh and colleagues isolated putative brain tumour stem cells (BTSCs) from a range of primary tumours using the neural stem cell surface marker, CD133 (Singh et al., 2003). $CD133^+$ BTSCs express Nestin - another normal neural stem cell marker - and demonstrate the ability to proliferate, self-renew and differentiate *in vitro* to form a cell population that is phenotypically similar to the original primary tumour. Moreover, compared to low-grade gliomas, the $CD133^+$ BTSCs isolated from highly aggressive medulloblastoma samples were more likely to self-renew and repopulate a tumour-like mass of cells in culture (Singh et al., 2003).

BRCA1$^{-/-}$-associated basal-like breast cancers are thought to arise from a luminal epithelial progenitor population which is normally enriched in breast tissue with mutated *BRCA1* (Lim et al., 2009, Molyneux et al., 2010). This cell-type-of-origin is also postulated to be involved in tumour propagation, as the *BRCA1*$^{-/-}$ luminal progenitors exhibit spontaneous clonal growth *in vitro* without the addition of exogenous factors normally required for colony formation (Lim et al., 2009).

There is evidence for a hierarchical cellular structure within colon cancer with varying propensities for tumour initiation (O'Brien et al., 2007, Ricci-Vitiani et al., 2007). Putative colon cancer-initiating cells can be isolated on the basis of CD133, CD44, or CD166 cell surface marker expression; these act in antigen recognition and as growth factor receptors (O'Brien et al., 2007, Ricci-Vitiani et al., 2007, Dalerba et al., 2007). Other examples of putative cancer stem cell markers are the Inhibitor of DNA binding proteins (IDs), which are helix-loop-helix transcription factors that govern self-renewal and pluripotency in embryonic stem cells by promoting expression of BMP4 (Gray et al., 2008, Hollnagel et al., 1999, Ruzinova and Benezra, 2003). Furthermore, overexpression of ID1 in stem cells of the mouse brain increases their capacity for self-renewal and asymmetric division (Nam and Benezra, 2009). Expression of IDs has also been reported in multiple solid tumours including pancreatic, cervical, ovarian, prostate, breast, and colon (Kleeff et al., 1998, Schindl et al., 2001, Ouyang et al., 2002, Lin et al., 2000, Fong et al., 2003, Meteoglu et al., 2008, Gray et al., 2008). In particular, ID1 and ID3 have been reported to stabilise expression of p21 and together encourage self-renewal of human colon cancer-initiating cells (O'Brien et al., 2012). This has also been observed in mouse mammary carcinoma: ID1 cooperates with the oncogene Ras to promote the growth of highly metastatic tumours with high expression of p21 (Swarbrick et al., 2008).

Given that there are many commonalities between normal stem cell and putative cancer stem cell gene and protein expression, it is conceivable that there is at least one novel common factor involved in maintenance of the undifferentiated state in normal stem cells and tumour-initiating cells.

Piwi/piRNA function in cancer

There are emerging reports of Piwi gene and protein expression in solid tumours. For example, in *Drosophila*, *Piwi, Aub,* and other piRNA biogenesis genes are up-regulated in malignant larval brain tumours, and their loss dramatically inhibits tumour formation and invasion (Janic et al., 2010). In humans, *Hiwi* is expressed in seminomas (Qiao et al., 2002) and gastric cancer (Liu et al., 2006); *Hili* is found in breast cancer (Liu et al., 2010) and both *Hiwi* and *Hili* have been identified in epithelial ovarian cancer (Chen et al., 2013). Furthermore, *Hiwi* expression in human glioma, pancreatic cancer, oesophageal cancer, and soft-tissue sarcoma is correlated with poor prognosis (Sun et al., 2011, Grochola et al., 2008, He et al., 2009, Taubert et al., 2007). Hiwi, Hili, and Hiwi2 proteins are also enriched in human colon cancer relative to normal tissue (Li et al., 2010).

Furthermore, pi-sized RNAs have been identified in the HeLa human cervical cancer cell line (Lu et al., 2010), suggesting that the Piwi/piRNA pathway could be active in cancer cells.

There have been several mechanisms proposed for a function of Piwi/piRNAs in cancer. Siddiqi et al speculate that reactivation of this pathway in normal cells could lead to the acquisition of "stem-like" epigenetic characteristics often observed in cancer – such as aberrant hypermethylation at CpG islands (Esteller, 2007) – resulting in silencing of tumour suppressor genes (Siddiqi and Matushansky, 2012). In support of this theory, one group has shown that overexpression of *Hiwi2* in human embryonic kidney (HEK293FT) cells represses $p16^{Ink4a}$ – a gene commonly silenced in cancer – via di-methylation of H3K9 (Sugimoto et al., 2007).

Secondly, recent data from our laboratory suggests a function for Piwi/piRNA in the regulation of protein translation in cancer: *Hiwi2* is expressed in metastatic MDAMB231 (231) breast cancer cells and partners with piRNAs derived from tRNAs (Keam et al., 2014). These tRNA fragments – or tRFs – are highly conserved throughout evolution and have been implicated in sequence-dependent and -independent repression of translation (Gebetsberger et al., 2012, Haussecker et al., 2010, Sobala and Hutvagner, 2013). Furthermore, our laboratory has shown that Hiwi2 associates with actively-translating ribosomes, and complexes with proteins with a putative role in translation, including eEF1-alpha and the heat shock proteins (Keam et al., 2014). There is precedent for the proposed role of the Piwi-piRNA pathway in translational control in the mouse germline, by Miwi and Mili (Grivna et al., 2006). In addition, tRFs are found to be complexed with the Piwi orthologue, Twil2, in the protozoan *Tetrahymena thermophila* and pi-sized tRFs are found to complex with actively-translating ribosomes in the archaeon *Haloferax volcanii* (Couvillion et al., 2010, Gebetsberger et al., 2012). The association of Piwi proteins with tRNA fragments means that aberrant activation of Piwi genes and proteins in cancer cells could affect expression of other proteins via translational regulation in response to environmental stimuli to confer a growth advantage for a tumour.

The final mechanistic theory for a role of Piwi/piRNA in cancer is based on its indispensable role in somatic stem cell self-renewal in primitive organisms: that the pathway promotes self-renewal and potency of stem-like cells within a tumour. There is some evidence that human breast cancer aggression correlates with expression of *Hili* in the "stem-like" cell subset within a tumour (Lee et al., 2010). This group has shown that *Hili* is aberrantly expressed in certain human cancer cell lines

and cancerous tissues, including 231 breast cancer cells and ductal and lobular carcinoma tissues. Notably, expression of the transcript is enriched in the $CD44^{+}CD24^{-}$ cancer stem cell subset of 231 cells and *in vitro* culture of *Hili*-expressing cells results in spontaneous ES-like colony formation, a phenomenon not observed following culture of *Hili*-negative cells (Lee et al., 2010). The group reported perturbations in the cancer anti-apoptosis and proliferation pathway following transient siRNA-mediated ablation of *Hili* expression, including downregulation of *Stat3, Bcl-*X_L and *Cyclin D1*. It is clear that Piwi gene and protein expression and function in cancer – if it exists – is complex and a mechanism for Piwi/piRNA-mediated genetic and epigenetic aberrations in cancer remains to be characterised.

The function of the somatic Piwi/piRNA pathway in mammals is, to date, unclear. It is becoming increasingly evident that Piwi protein expression in somatic progenitor cells is not restricted to primitive metazoans. Considering its indispensable role in the stem cells of primitive organisms and in germline stem cell self-renewal in higher organisms, it is tempting to speculate that this pathway is also somehow involved in somatic stem cell function in higher eukaryotes. Germline stem cells share characteristics with somatic stem cells, therefore it is possible that there is conservation of underlying molecular mechanisms for processes such as self-renewal. It is conceivable that Piwi proteins and their bound piRNAs maintain stem-like self-renewal and potency during development and homeostasis, and even in deleterious stem-like function in disease-related processes, such as tumorigenesis and metastasis.

Hypothesis and Aims: a putative role for the Piwi/piRNA pathway in maintenance of stemness in mouse somatic stem cells and cancer

Piwi orthologues are highly conserved in metazoans and appear to have a role in stem cell maintenance during homeostasis and regeneration in primitive organisms (Reddien et al., 2005, Alie et al., 2011, Funayama et al., 2010, Rinkevich et al., 1995). Moreover, Piwi protein, Piwi-bound piRNA, and pi-like RNA expression have been reported in multiple somatic tissues in higher organisms such as *Drosophila,* mouse, rhesus macaque, and humans (Yan et al., 2011). Piwi orthologue expression is reported to be enriched in stem-like cells relative to differentiated cells in multiple species, and in some instances, ablation of the *Piwi* genes causes stem cell differentiation without self-renewal (reviewed in (Ross et al., 2014)). In addition, Piwi protein and transcript expression has been reported in a number of tumours and cancer cell lines in *Drosophila* and humans. Part A of my hypothesis is that one or more Piwi proteins are expressed in murine stem cells and that – along with their bound piRNAs – they function to maintain self-renewal and pluripotency. Furthermore, if required for the stem cell phenotype, I predict that murine Piwi gene and protein expression will decrease concurrently with cell differentiation (Figure 1-9A). Part B of my hypothesis predicts that murine Piwi proteins are expressed in a small subset of stem-like cells within a tumour and that these cells contribute to tumorigenesis and cancer metastasis (Figure 1-9B). In this thesis, I aim to address this hypothesis in the following manner:

i. Investigate somatic expression and cellular localisation of Miwi, Mili, and Miwi2 transcripts and proteins in mouse embryonic stem cells and somatic tissues
ii. Observe phenotypic changes, perturbations in gene expression and differentiation potential of mouse embryonic stem cells following murine Piwi gene knockdown
iii. Determine whether murine Piwi genes and proteins are expressed and are involved in the maintenance of a stem-like phenotype in mammary carcinoma cells
iv. Ascertain whether murine Piwi proteins bind piRNAs in mouse embryonic stem cells and cancer cells to provide an insight into somatic Piwi/piRNA pathway function

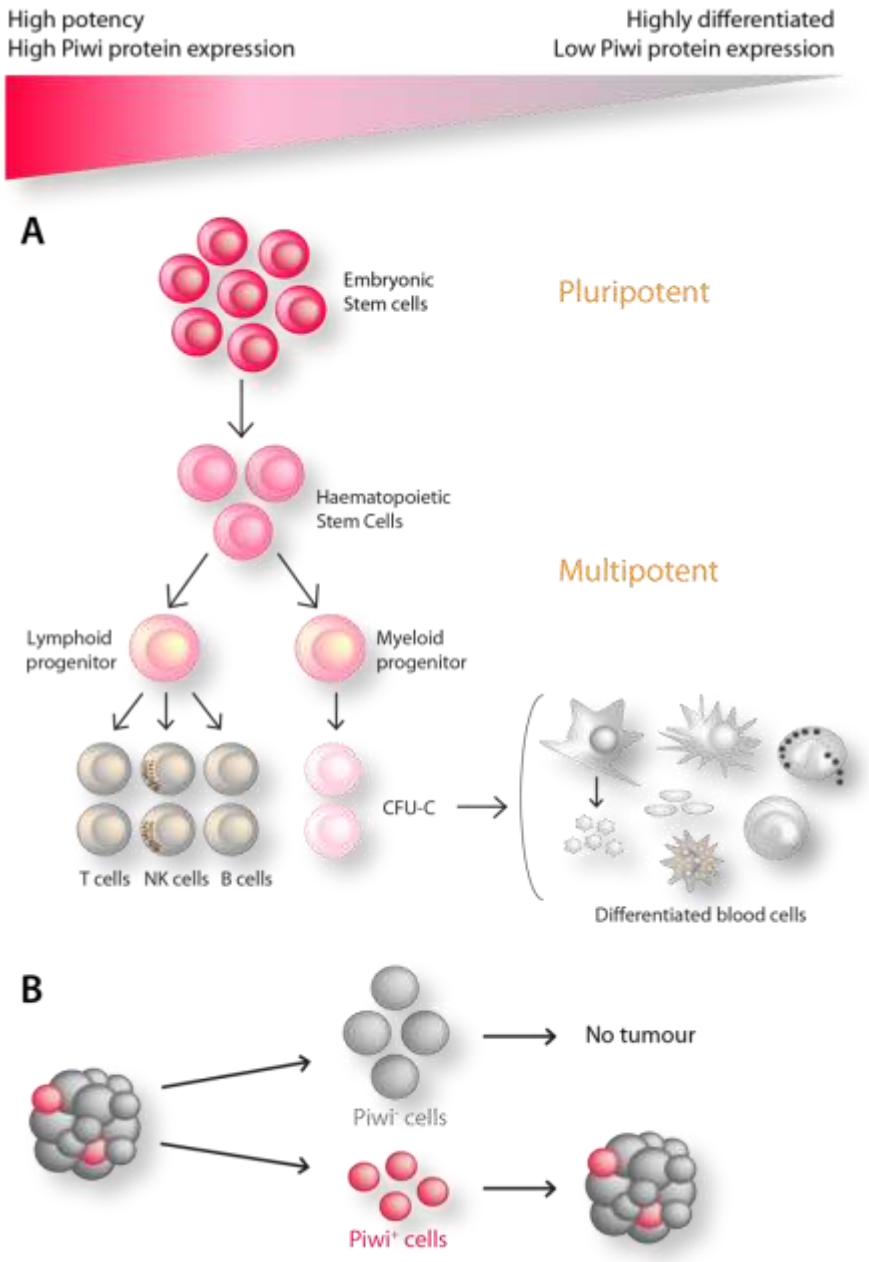


Figure 1-9: Hypothesis (Part A and Part B)

(A) Piwi gene and protein expression is correlated with cellular potency. Haematopoietic differentiation is shown as an example to illustrate this hypothesis. Pluripotent embryonic stem cells exhibit the highest level of Piwi protein expression followed by multipotent progenitors. Piwi genes and proteins are not expressed in differentiated cells. (B) Tumours contain a subset of Piwi gene- and protein-expressing stem-like cells with the potential to differentiate to form all cell types within a secondary tumour. Piwi-negative cells are unable to reconstitute a secondary tumour.

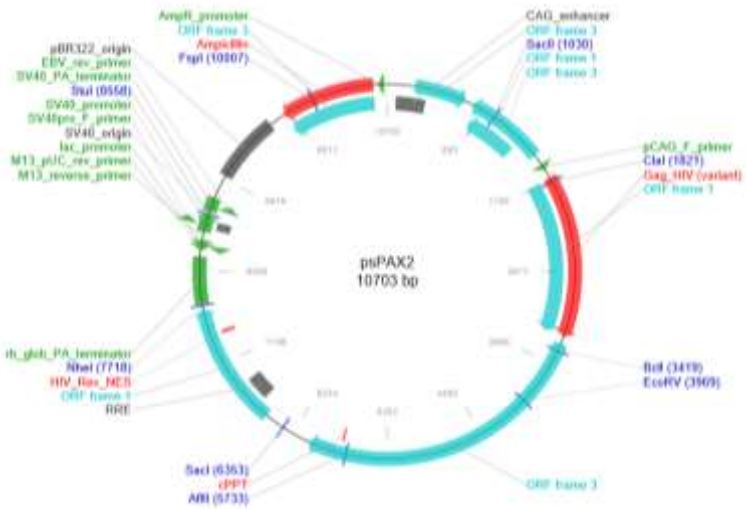
2. Materials and Methods

2.1 Chemicals used in this study (Sigma Aldrich)

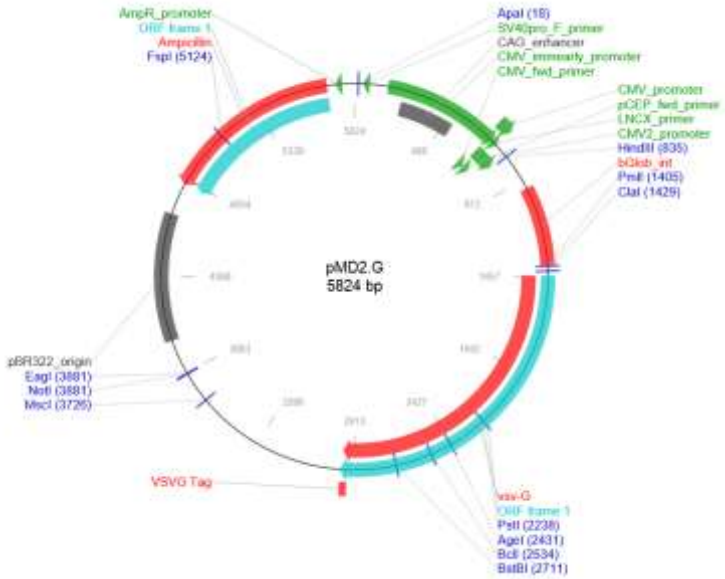
Name	Alias	Catalogue number
Beta-mercaptoethanol	β-me	M6250
Mitomycin C	MMC	M0503
Dimethyl Sulfoxide	DMSO	472301
Puromycin dihydrochloride	Puromycin	P8833
Butyric acid		B103500
Hexadimethrine Bromide	Polybrene	H9268
Glycine		G8898
HEPES		H3375
Sodium phosphate dibasic	Na_2HPO_4	S7907
Calcium chloride	$CaCl_2$	449709
HEPES		H3375
DL-Dithiothreitol	DTT	D9779
Potassium Chloride	KCl	P9541
Magnesium Chloride	$MgCl_2$	M8266
Ethylenediaminetetraacetic acid	EDTA	E6758
Trizma® hydrochloride	Tris-HCl	T3253
Triton™ X-100		T8787
Sodium Chloride	NaCl	S7653
1,2,3-Propanetriol	Glycerol	G5516
Sodium dodecyl sulfate	SDS	L4390
Nonidet™ P-40	NP-40	74385
Sodium deoxycholate		D6750

2.2 Vectors used in this study

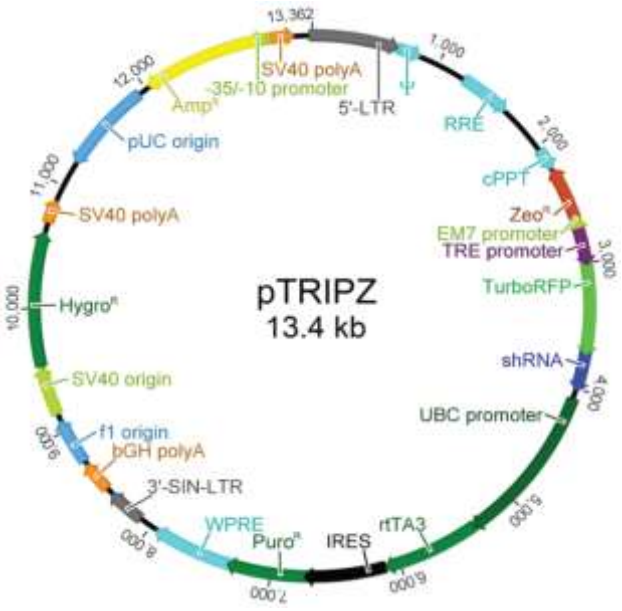
psPAX2 (Addgene, 12260)



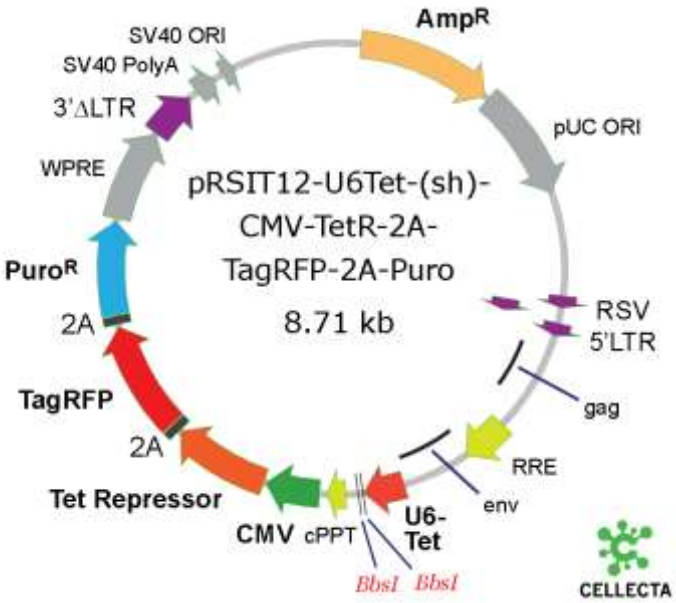
pMD2.G (Addgene, 12259)



pTRIPZ (Thermo Scientific Open Biosystems)

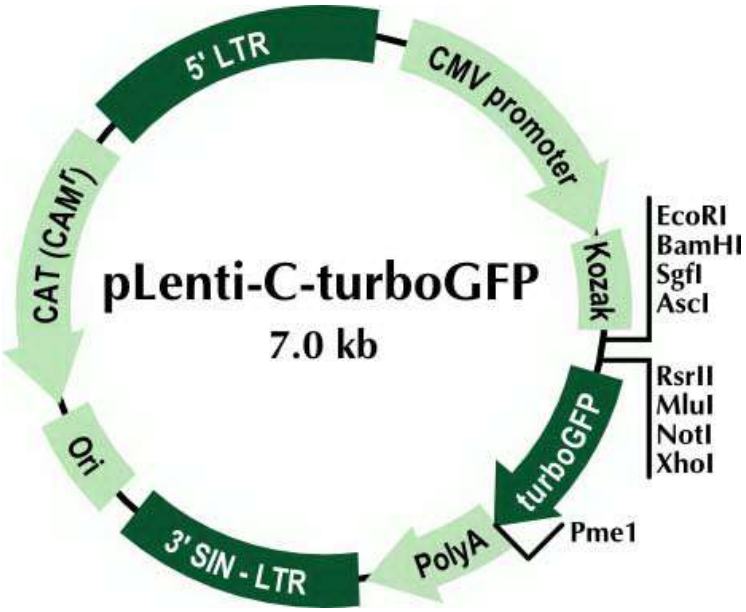


pRSIT (Cellecta, Inc.)



Nb. The vector used in this study contained TagGFP instead of TagRFP (vector map unavailable)

pLenti-C-turboGFP (OriGene, PS100065)



2.3 General molecular biology procedures

2.3.1 Determination of nucleic acid concentration and integrity

DNA and RNA concentration was estimated using the Nanodrop 2000 (Thermo Scientific) using absorbance at 260nm (A_{260}); an A_{260} absorbance of 1.0 was assumed to be equivalent to a dsDNA concentration of 50µg/ml and RNA concentration of 40µg/ml. Purity was measured via A_{260}/A_{280} and $A_{260}/_{230}$: pure RNA exhibits $A_{260}/_{280} \approx 2.0$, $A_{260}/A_{230} \approx 2.0$; pure DNA measures $A_{260}/A_{280} \approx 1.8$, $A_{260}/A_{230} \approx 2.0$. For very accurate measurements of nucleic acid concentration and integrity, I used the 2100 Bioanalyser High Sensitivity DNA Kit (Agilent, 5067-4626) or RNA 6000 Nano Kit (Agilent, 5067-1511).

2.3.2 Restriction digestion of plasmid DNA

For restriction digestion, all enzymes were obtained from New England Biolabs® (NEB) Incorporated. Optimal reaction conditions for each enzyme were calculated using the NEB Enzyme Finder or Double Digest Finder (https://www.neb.com/tools-and-resources). Reactions were heat-inactivated at 65°C or 80°C according to the manufacturer's directions.

2.3.3 Agarose gel electrophoresis, visualisation and gel extraction and ligation of plasmid DNA

Restriction-digested plasmid DNA fragments were separated and visualised on a horizontal agarose gel containing 1µg/ml Ethidium Bromide (1.5% agarose for fragments <1kb, 0.8% agarose for fragments >1kb). Fragments were visualised on a FotoPrep UV transilluminator (Fotodyne) and manually excised. DNA was extracted from solubilised gel slices using the QIAEXII Gel Extraction Kit (Qiagen, 20021). Linearised plasmid backbones and DNA inserts were ligated at a 1:1 molar ratio using the LigaFast™ Rapid DNA Ligation System (Promega, M8221) for 4h at room temperature.

2.3.4 Plasmid DNA preparation

MAX Efficiency® DH5α™ Competent *E. Coli* cells (Invitrogen, 18258-012) were transformed with approximately 1µg plasmid DNA using heat shock for 45-50sec at exactly 42°C. Cells were allowed to recover in LB medium ~1.5h at 37°C with shaking prior to plating onto LB agar containing either 100µg/ml Ampicillin or 25µg/ml Kanamycin and incubation overnight at 37°C. Resistant colonies were picked and cultures grown in LB medium plus 100µg/ml Ampicillin or 25µg/ml Kanamycin overnight at 37°C with shaking at 140rpm in an Innova® 42 Incubator Shaker (Eppendorf). After 12-16h incubation, bacteria were harvested via centrifugation at 4000xg at 4°C. Bacteria were lysed and plasmid DNA isolated using the QIAprep Spin Miniprep kit (Qiagen, 27104).

2.4 Tissue Culture and Cell Biology

Table 2-1: Cell lines used in this study

Cell line	Cell type	Details of origin	Reference
MEF	Embryonic fibroblast	Isolated from E14 mouse embryos	N/A
R1	Embryonic stem	Isolated from the inner cell mass of male 129X1 x 129S1 blastocyst	(Nagy et al., 1993)
W9.5	Embryonic stem	Isolated from the inner cell mass of male 129/SvJ blastocyst	Jackson Labs
HEK293FT	Embryonic kidney	Human embryonal kidney cells transformed with SV40 Large T antigen	(Naldini et al., 1996)
4T1	Breast cancer	Metastatic, thioguanine-resistant cells isolated from BALB/cfC3H mouse mammocarcinoma	(Aslakson and Miller, 1992)
66cl4	Breast cancer	Metastatic, thioguanine-, ouabain-resistant cells isolated from BALB/cfC3H mouse mammocarcinoma	(Aslakson and Miller, 1992)
4TO7	Breast cancer	Non-metastatic, thioguanine-, ouabain-resistant cells isolated from BALB/cfC3H mouse mammocarcinoma	(Aslakson and Miller, 1992)
168FARN	Breast cancer	Non-metastatic, diamino-, purine-, geneticin (G418)-resistant cells isolated from BALB/cfC3H mouse mammocarcinoma	(Aslakson and Miller, 1992)
67NR	Breast cancer	Non-metastatic, geneticin (G418)-resistant cells isolated from BALB/cfC3H mouse mammocarcinoma	(Aslakson and Miller, 1992)

2.4.1 Expansion and mitotic arrest of mouse embryonic fibroblast feeder cells

Mouse embryonic fibroblast cells (MEFs) were obtained at Passage 1 (Harvey Laboratory, Victor Chang Cardiac Research Institute) and plated onto 150mm tissue culture treated culture dishes (Corning, CLS430599). Cells were expanded in DMEM containing 10% Foetal Bovine Serum (FBS, GIBCO, cat #10099-141, lot #769369), 1mM Sodium Pyruvate (GIBCO, 11360-070**)**, 2mM GlutaMAX™ (GIBCO, 35050061), 10mM MEM Non-Essential Amino Acids Solution (GIBCO, 11140-

050), 0.1mM Beta-mercaptoethanol. At Passage 4, expansion media was replaced by DMEM + 10µg/ml Mitomycin C. Mitotic inactivation was conducted for 3h at 37°C.

2.4.2 Growth and maintenance of cell lines

Mitomycin C-treated MEFs were maintained in DMEM high glucose (GIBCO, 11965-092) with 10% FBS (GIBCO, cat #10099-141, lot #769369) for up to two weeks. R1 and W9.5 mouse embryonic stem cells were maintained in complete ES media: DMEM high glucose (GIBCO, 11965-092) with 15% ES cell batch-tested Foetal Bovine Serum (Thermo Scientific, cat #15-010-0500V, lot #E08045-500), 2mM GlutaMAX™ (GIBCO, 35050061), 10mM MEM Non-Essential Amino Acids Solution (GIBCO, 11140-050), 0.1mM Beta-mercaptoethanol and 1000U/ml ESGRO® Mouse Leukaemia Inhibitory Factor (LIF, Millipore, ESG1106). Complete ES media was replaced daily and cells passaged every 48h with 0.05% Trypsin-EDTA (GIBCO, 25300-054). At each passage, 1 x 10^5 ES cells were plated on a confluent layer of MEFs (5 x 10^5 cells/well) on tissue culture-treated 6 well plates (Corning® Costar®, CLS3516). All murine mammary carcinoma cell lines (4T1 panel) were maintained in MEM-alpha (GIBCO, 12571071) + 5% Foetal Bovine Serum (GIBCO, cat #10099-141, lot #769369) and passaged at 80-90% confluence. HEK293FT cells were maintained in DMEM high glucose (GIBCO, 11965-092) with 10% FBS (GIBCO, cat #10099-141, lot #769369) and passaged at 80% confluence.

2.4.3 Cryopreservation and recovery

Cells were harvested with 0.05% Trypsin-EDTA (GIBCO, 25300-054) and pelleted at 300xg, 3 min. Cells were resuspended in freeze media containing 90% FBS (GIBCO, cat #10099-141, lot #769369) and 10% Dimethyl Sulfoxide (DMSO). Cells were frozen at -80°C in 1.0ml aliquots in a Mr Frosty™ Freezing Container (Thermo Scientific, 5100-0001) and stored at -80°C up to one month or transferred to Liquid Nitrogen vapour phase for long-term storage.

2.4.4 In vitro differentiation of mouse embryonic stem cells

Pluripotent mouse ES cells were harvested with 0.05% Trypsin-EDTA (GIBCO, 25300-054), diluted in ES cell medium and pre-plated onto tissue culture-treated dishes 2 x 30 min to deplete MEFs from pluripotent culture. Cells were resuspended at 5 x 10^4 cells/ml in EB media (complete ES media with LIF omitted, see Section 2.4.2) and plated in 20µl hanging drops on the lid of a hydrophobic petri dish, which was flipped and place on a petri dish base containing 5ml PBS to

prevent drops from drying out. Hanging drops were incubated at 37°C, 5% CO_2 for 1-6 days. Embryoid bodies were harvested with p1000 pipette.

2.4.5 Antibiotic selection

Optimal antibiotic concentration was determined by kill curve in a 6 well plate: cells were plated at 2 x 10^5 cells/ well and exposed to puromycin at optimised concentrations. Cell viability was estimated daily via microscopic examination and the optimal concentration was selected based on the minimum which caused 100% cell death after 6-8 days. 24h post-transduction, cells were exposed to puromycin at the optimal concentration (1µg/ml for ES cells and 6µg/ml for 4T1 cells); antibiotic was refreshed every 1-2 days. When resistant clones were established (ie. no cell death occurring), puromycin concentration was halved to maintain resistance.

2.4.6 Transfection of cell lines

24h prior to transfection, HEK293FT cells were plated at 2 x 10^6 cells/100mm tissue-culture treated dish (Corning, CLS430167). The next day, 5µg plasmid DNA was combined with 14µl PLUS™ reagent (Invitrogen**,** 11514015) and made up to 600µl with OptiMEM® (Life Technologies, 31985070). 20µl Lipofectamine® Transfection Reagent (Life Technologies, 18324020) diluted in 580µl OptiMEM was added dropwise to DNA:PLUS and transfection complexes formed at room temperature. After 30 min, transfection complexes were made up to 6ml with OptiMEM. 70% confluent HEK293FT cells were washed with OptiMEM prior to dropwise addition of transfection complexes. After 6h incubation at 37°C, media transfection complexes were replaced with DMEM high glucose + 10% FBS.

2.4.7 Lentiviral packaging of plasmid DNA

I used the calcium chloride method of lentiviral packaging as outlined by (Kingston et al., 2003) and adapted by Alexis Bosman (Harvey Laboratory, VCCRI). In brief, HEK293FT cells were plated at 1.5 x 10^6 cells/10cm dish 24h prior to transfection. On the day of transfection, 10µg lentiviral vector was mixed with packaging plasmids: 10µg psPAX2 and 5µg pMD.2G. The plasmid mixture was diluted in 250µl MilliQ H_2O with 1M HEPES, pH 7.3 and added to 500µl 2X HBS solution (see Section 2.12). DNA:H_2O:HBS was added dropwise to 4X $CaCl_2$ solution (see Section 2.12) while vortexing at top speed and incubated at room temperature for exactly 20 min. Transfection complexes were overlaid, dropwise, onto ~70% confluent HEK293FT cells and incubated overnight

at 37°C, 5% CO_2. 16-18h post-transfection, transfection complexes were discarded and replaced with DMEM + 10% FBS (GIBCO, cat #10099-141, lot #769369) + 10mM Butyric Acid. 30h later, HEK293FT supernatant was harvested. Cell debris was removed by centrifugation at 1000xg, 10min. Virus-containing supernatant was filtered using a 0.45µm Minisart filter (Sartorius, 16555K) and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 (Millipore, UFC910008).

2.4.8 Lentiviral transduction and stable selection of cell lines

For transduction in 6 well plates, 250µl concentrated virus was diluted 1:4 in 750µl cell-specific media containing 1 x 10^4 target cells. Hexadimethrine Bromide (polybrene) was added at 8µg/ml. 1 ml virus:cell suspension was "spin-fected" via centrifugation at 850xg, 60min at room temperature. An additional 1ml media + 8µg/ml polybrene was overlaid onto transduced cells and incubated overnight at 32°C. The following day, transduction complexes were removed and replaced with cell-specific media; from this point transduced cells were maintained at 37°C.

2.4.9 Flow cytometry for transduced reporter fluorescence

Cells were harvested with 0.05% Trypsin-EDTA, washed twice with PBS and resuspended in ice cold FACS buffer (PBS + 1% FBS + 2mM EDTA). Cells were kept on ice until sorting. FACS was conducted using a BD FACS Influx: GFP^+ single cells were collected in 1ml cell-specific media.

2.5 Tissue collection and preparation

2.5.1 Tissue preparation for nucleic acid and protein extraction

Mice were euthanised via CO_2 inhalation. Tissues were harvested immediately, snap-frozen in liquid nitrogen and stored at -80°C until required. For RNA extraction, frozen tissues were homogenised in ice cold TRIzol® Reagent (Life Technologies, 15596-026) and RNA extracted according to Section 2.6.1. For protein extraction, frozen tissue was homogenised in ice cold RIPA buffer (see Section 2.12) plus cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001).

2.6 RNA and cDNA analyses

2.6.1 RNA extraction and reverse transcription

RNA was extracted using TRIzol® Reagent (Invitrogen, 15596-026) and Direct-Zol™ RNA MiniPrep (Zymo Research, R2050) according to the manufacturer's directions. In brief, cells/tissues in TRIzol are diluted 1:1 in 100% ethanol and loaded onto a Zymo-Spin™ IIC Column with filter. RNA bound to the column is washed x1 with Direct-Zol™ RNA PreWash and x2 with Direct-Zol™ RNA Wash Buffer and eluted with nuclease-free H_2O. 1µg RNA was reverse-transcribed to cDNA using QuantiTect Reverse Transcription Kit (Qiagen, 205313).

2.6.2 Selection of reference genes and primer design for RT-qPCR

Genevestigator (https://www.genevestigator.com/gv/biomed.jsp) was used to select a list of candidate reference genes based on i) typical expression level of target gene and ii) range of tissue types of interest. NCBI Primer BLAST (Ye et al., 2012) was used to generate putative primer pairs according to user-defined parameters. Primer oligos were synthesised by Integrated DNA Technologies. Primer pairs were tested empirically and final pairs selected from multiple candidates based on an amplification efficiency of ~2 (>1.9 and <2.1). All primers used in this study were designed for common PCR conditions.

2.6.3 qPCR sample setup, cycling conditions and analysis

qPCR master mix was prepared containing a final concentration of 1X Roche Lightcycler® SYBR Green I Master and 0.5µM each forward and reverse primers. 9µl master mix was dispensed into each well of a Lightcycler® 480 Multiwell Plate, 384 (Roche, 04729749001) using the EpMotion 5075 Liquid Handling Robot (Eppendorf) and 1µl cDNA was dispensed manually. qPCR was performed using the Lighcycler® 480 Real-Time PCR Instrument (Roche, 05015243001). Relative expression was determined for each sample via Comparative C_T analysis (Schmittgen and Livak, 2008).

2.6.4 Microarray

Labelling, hybridization and scanning was performed by the Ramaciotti Centre for Genomics, University of New South Wales. In brief, sense-strand cDNA was synthesised from total RNA using the Ambion® WT Expression Kit (Life Technologies), followed by fragmentation and labelling using

the Affymetrix GeneChip® WT Terminal Labelling Kit. Finally, the fragmented, labelled cDNA was hybridised to the GeneChip® Mouse Gene 2.0 ST Array, containing probes for >35,000 coding and non-coding transcripts, including >2000 long intergenic non-coding transcripts. Each sample was hybridised to an individual array.

Microarray-generated intensity files were pre-processed (including background adjustment normalisation and probe summarisation) using Robust Multi-Array Average (RMA) as part of the Affymetrix® Expression Console™ software package. Normalised data were annotated using ReannotateGCT, followed by differential gene expression analysis using the LimmaGP module (version 19.3) as part of the Gene Pattern software platform (Broad Institute, available at pwbc.garvan.unsw.edu.au/gp). This module makes use of the Limma algorithm, which combines linear models with an empirical Bayes, moderated t-statistic. This has more power than the Student's t-test or ANOVA for analysing microarray data.

2.7 Protein extraction and quantification for immunodetection and RNA-immunoprecipitation

2.7.1 Protein extraction from cultured cells and tissues

Adherent cells were washed twice with ice cold PBS and harvested by scraping in ice cold PBS. Cells were pelleted at 300xg, 5 min at 4°C and lysed in RIPA buffer (for western blot, see Section 2.12) or Polysome Lysis Buffer (for RNA-immunoprecipitation, see Section 2.12) containing cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001). Frozen tissues were homogenised in ice-cold RIPA buffer containing protease inhibitor.

2.7.2 Fractionation of cytoplasmic and nuclear cellular compartments

Cell pellets were washed in ice cold PBS and resuspended in ice cold Fractionation Buffer A (see Section 2.12) and allowed to swell on ice for 15 min. Cell membranes were lysed with the addition of 10% NP-40 (Pierce, 85124) and nuclear fraction pelleted via centrifugation at 13,000 x g, 4°C for 10 min. Supernatant (cytoplasmic fraction) was removed and stored at -80°C. Nuclear fraction was washed x2 with Fractionation Buffer A and lysed with Fractionation Buffer C (see Section 2.12). Genomic DNA was sheared by sonification with a 450 Digital Sonifier Cell Disruptor (Branson, 2601-04) twice for 5 sec at 10% amplitude. Debris was pelleted via centrifugation at max speed, 4°C for 15 min. Supernatant (nuclear fraction) was stored at -80°C.

2.7.3 Quantification of protein concentration

Protein concentration was estimated using Direct Detect™ Spectrometer (Millipore, DDHW00010-WW). In brief, 2µl sample is loaded directly onto a card-based hydrophilic polytetrafluoroethylene (PTFE) membrane; the instrument estimates protein concentration via quantification of amide bonds within protein chains in the sample.

2.8 Synthesis of monoclonal antibody panel

2.8.1 Generation of monoclonal antibodies

Monoclonal antibody panel was generated by Abmart™, Shanghai, China. Peptide antigens were developed prior to immunisation using Surface Epitope Antibody Library (SEAL™) technology. Sequences for multiple SEAL™ antigens and a number of potent adjuvants were carried on a DNA vector and, once transcribed and translated, were expressed on the surface of the resulting immunogenic recombinant protein. Following immunisation with the peptide antigen, mouse spleen cells were fused with immortalised myeloma cells. The fused hybridoma cells were injected into the peritoneal cavity of mice, resulting in tumours that secreted antibody-rich ascites fluid.

2.8.2 IgG affinity purification from hybridoma cell supernatant

Hybridoma cell lines were obtained from Abmart and initially expanded and maintained in RPMI 1640 (GIBCO, 11875-093) + 20% FBS (GIBCO, cat #10099-141, lot #769369) + 2mM GlutaMAX (GIBCO, 35050061) prior to weaning onto IgG-free CD hybridoma media (Invitrogen, 11279-023). Cells were expanded in a 250ml tissue-culture treated flask and allowed to overgrow and apoptose. Antibody-containing supernatant was filtered with a 0.22µm Vacuum Filtration System (Corning, CLS430767). IgG affinity purification was performed with Peter Schofield (Antibody Development Laboratory, Garvan Institute of Medical Research). Protein G-Sepharose was mixed 1:3 with unconjugated Sepharose and pre-equilibrated with PBS. Beads were added to hybridoma supernatant and incubated 15min at room temperature with agitation to allow Protein G to bind IgG. Sepharose beads-IgG were packed into a chromatography column (Bio-Rad Econo-Column, 737-1512). Column was washed x1 with PBS and IgG eluted with 100mM Glycine + 0.1M NaCl. Elution fraction was neutralised with 1/10th volume Tris-HCl, pH 8.0. Elution was repeated 4-5 times and each fraction tested for presence of IgG by measuring absorbance at 280nm (A_{280}) using a Nanodrop 2000 (Thermo Scientific). This assumes that a solution with an optical density of 1.4

contains 1mg/ml protein. Antibody-containing elution fractions were pooled and decanted into SnakeSkin Dialysis Tubing (10kDa cut-off, Thermo Scientific, 68100). Dialysis was performed twice in DPBS (GIBCO, 14040-133): once for 1h at room temperature and again overnight at 4°C. Dialysed fractions were filtered with a 0.22µm Minisart syringe filter (Sartorius, 16534K) and concentrated using Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 (Millipore, UFC910008).

2.8.3 ELISA

ELISA was performed with Peter Schofield (Antibody Development Laboratory, Garvan Institute of Medical Research). A Nunc F96 Maxisorp® ELISA plate (Affymetrix, 442404) was coated with 5µg/ml unlabelled mouse IgG and IgM antibody. The coated plate was blocked in 4% skim milk in PBS + 0.01% Tween (PBS-T), 1h at room temperature and washed x3 with PBS-T and x1 with PBS. 50µl hybridoma supernatant was added to each testing well and 50µl positive control IgG2a in skim milk + PBS-T to each control well and incubated from 1-3h at room temperature. After washing x3 with PBS-T and x1 with PBS, 2° antibody was added at 1:2000 in skim milk + PBS-T (SouthernBiotech, Goat polyclonal anti-mouse, HRP-conjugated) and incubated 1h at room temperature. Secondary antibodies were IgG isotype-specific: IgG1, IgG2a-c, IgG3 and IgM. After washing 3x with PBS-T and x1 with PBS, 50µl TMB substrate (3,3',5,5'-Tetramethylbenzidine substrate and hydrogen peroxide, Pierce, 34022) was added to each well and the reaction stopped with hydrochloric acid after colour change was apparent. Strong absorbance at 450nm was used to determine antibody isotype.

2.8.4 Generation of protein homology model and prediction of antibody binding site

The peptide sequence of the protein of interest was subjected to NCBI protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against all known protein structures derived from X-ray Crystallography, nuclear magnetic resonance spectrometry and electron microscopy contained within the Protein Data Bank (PDB). Then, the peptide sequence of interest and most homologous available protein structure were aligned using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and the alignment was mapped onto the homologous structure using CHAINSAW (Stein, 2008). The homology model was visualised in multiple orientations using Pymol (http://www.pymol.org/).

2.9 Immunodetection

2.9.1 Western Blotting

10µg protein in RIPA was diluted in Novex® LDS Sample Buffer and DTT, loaded onto a vertical gradient polyacrylamide gel (NuPAGE® Novex® 4-12% Bis-Tris), and electrophoresed at 200V, 35 min. Protein was transferred onto a nitrocellulose membrane for 7 min at room temperature using iBlot® Gel Transfer Device (Life Technologies, IB1001). Membrane was blocked >2h at room temperature in Blocking Buffer for Fluorescent Western Blotting (Rockland, MB-070). Membrane was incubated with primary antibody followed by fluorescent secondary antibody diluted in blocking buffer for varying conditions (see Table 2-2) with 3x 5 min washes with PBS + 0.1% Tween-20 after each. Membrane fluorescence was visualised at 680nm and 800nm with the LI-COR Odyssey Infrared Imaging System.

Table 2-2: Commercial antibodies used for western blot

1°/2°	Supplier / cat #	Target	Species / type	Dilution / concentration	Incubation time / temp
1°	Abcam ab21869	Anti-Miwi2	Rabbit polyclonal	1µg/ml	Overnight, 4°C
1°	Sigma F1804	Anti-FLAG	Mouse monoclonal	1µg/ml	1h, RT
1°	Abcam ab1790	Anti-H2B	Rabbit polyclonal	1µg/ml	30min, RT
1°	Santa Cruz	Anti-Tubulin	Mouse monoclonal	1µg/ml	1h, RT
2°	LI-COR	Anti-rabbit	Goat polyclonal	1:10,000	1h, RT
2°	LI-COR	Anti-mouse	Goat polyclonal	1:10,000	1h, RT

2.9.2 Immunofluorescence

Unless otherwise specified, all steps were performed at room temperature; all wash steps are performed with PBS containing 0.1mM $CaCl_2$ and 1mM $MgCl_2$. HCl-treated, sterile glass coverslips were placed into 6 well plates and gelatinised overnight at 37°C. Cells were plated at high density

and maintained overnight at 37°C in cell-specific media. For immunostaining, media was removed, cells were washed x2 and fixed with 4% paraformaldehyde (PFA) in PBS, 15 min on ice. 4% PFA was discarded and 150mM Glycine in PBS added to each well for 15 min. Cells were washed x2 and blocked with Blocking Solution (2% Bovine Serum Albumin and 0.1% Triton X-100 in PBS) for 30 min. Cells were incubated with 1° antibody diluted in Blocking Solution 4h to overnight at 4°C. Cells were washed x3, 10 min with agitation and incubated with 2° antibody diluted in Blocking Solution for 1h in the dark. Additional stains (see Table 2-3) were added for the final 5 min of 2° antibody incubation and cells were washed x3, 10min with agitation. Coverslips were mounted onto glass slides with ProLong® Gold Reagent (Life Technologies, P36930) and kept in the dark overnight.

Table 2-3: Commercial antibodies and stains used for immunofluorescence

1°/2°/Stain	Supplier/cat #	Target/Stain	Species / type	Dilution/conc
1°	Abcam cat# ab21869, lot #GR-1303231	Anti-Miwi2	Rabbit polyclonal	5µg/ml
2°	Molecular Probes, cat #A11012, lot #757099	Anti-rabbit IgG	Goat polyclonal conjugated to Alexa Fluor 594	1:500
2°	Jackson Labs	Anti-rabbit IgG	Donkey polyclonal conjugated to Cy3	1:500
Stain	Life Technologies, H3570	Hoechst 33342		1:10,000
Stain	Life Technologies, A22286	Phalloidin	Conjugated to Alexa Fluor 647	1:500

2.9.3 SSEA1 immunostaining for flow cytometry sorting

Adherent cells were harvested with 5mM EDTA in PBS at 37°C, 15min, pelleted at 300xg, 3 min and resuspended in ice cold FACS buffer (PBS + 1% FBS + 2mM EDTA). 5µl anti-SSEA1 Mouse IgM antibody conjugated to Allophycocyanin (APC, R&D Systems, FAB2155A) or control IgM-APC antibody was added per $1x10^5$ cells and incubated for 45 min on ice in the dark. After staining, sample was diluted in FACS buffer and centrifuged at 300xg, 3min. Supernatant was discarded and cells were resuspended in ice-cold FACS buffer and kept on ice until sorting. 7AAD-negative, single

SSEA1^{+} cells were sorted using a BD FACSAria™ Cell Sorter (BD Biosciences) and cells collected into cell-specific media.

2.10 Immunoprecipitation

2.10.1 Immunoprecipitation for optimisation of monoclonal panel

Unless otherwise specified, all incubation steps were performed at 4°C with rotation. 1mg cell lysate in RIPA buffer was pre-cleared with 30µl Protein G Dynabeads® (Invitrogen, 10003D) for 2h. In the meantime, a separate aliquot of 30µl Protein G Dynabeads was incubated with 5µg antibody in NT2 buffer for at least 30min. Pre-equilibrated, antibody-bound Protein G Dynabeads were added to pre-cleared lysate and immunocomplexes formed overnight. Bead:IgG:protein complexes were washed x3 in NT2 buffer and IgG:protein complexes were eluted from the beads and dissociated by resuspending in Novex® LDS Sample Buffer and DTT and heating at 95°C, 5 min.

2.10.2 Miwi2/RNA-immunoprecipitation

Unless otherwise specified, all incubation steps were performed at 4°C with rotation. >10mg cell lysate in Polysome Lysis Buffer was pre-cleared with 30µl Protein G Dynabeads® (Invitrogen, 10003D) for 2h. In the meantime, a separate aliquot of 30µl Protein G Dynabeads was incubated with 5µg antibody in NT2 Buffer for at least 30min. Pre-equilibrated, antibody-bound Protein G Dynabeads were added to pre-cleared lysate and immunocomplexes formed overnight. Bead:IgG:protein complexes were washed x6 in NT2 Buffer. For western blot validation of immunoprecipitation, IgG:protein complexes were eluted from 5% of the beads and dissociated by resuspending in Novex® LDS Sample Buffer and DTT and heating at 95°C, 5 min. For RNA isolation, 95% of the beads were resuspended in NT2 buffer + RNaseOUT (10777019) and incubated with 60µg Proteinase K (Millipore, 539480) for 2h at 55°C with agitation. The aqueous phase was double-extracted with Acid-Phenol:Chloroform, pH 7.5 (Ambion®, AM9720) and RNA precipitated with 75% ethanol, overnight at -20°C. RNA was pelleted at max speed, 4°C for 20 min, washed with 70% ethanol, air-dried and resuspended in 8µl nuclease-free H_2O for use in small RNA library preparation.

2.10.3 Protein mass spectrometry

Liquid Chromatography-Mass Spectrometry (LC-MS) was performed by the Bioanalytical Mass Spectrometry Facility at the Mark Wainwright Analytical Centre, University of New South Wales.

2.11 Next-generation sequencing

2.11.1 Library preparation

Small RNA libraries were prepared using NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) (New England Biolabs, E7330S) according to the manufacturer's directions. Following PCR amplification, cDNA libraries were cleaned up using the QIAquick PCR Purification Kit (Qiagen, 28104) and samples subjected to quality control (QC) and size selection via gel purification. DNA fragments were separated on a vertical 6% polyacrylamide gel at 120V, 2h. Gel was stained with SYBR® Gold Nucleic Acid Stain (Life Technologies, S-11494) for 5 min and DNA visualised on a Fujifilm FLA-5100 Imager using a 473nm laser and LPB filter. Bands between 135-155bp (corresponding to miRNA and piRNA-sized small RNAs) were excised and crushed. Libraries were eluted in DNA Gel Elution Buffer for 3h to overnight at room temperature with rotation followed by the addition of 1µl linear acrylamide and 0.3M sodium acetate, pH 5.5. DNA was precipitated using 75% ethanol for at least 30 min at -80°C. DNA was pelleted by centrifugation at 14,000xg, 4°C for 30 min, washed in 80% ethanol, centrifuged at 14,000xg, 4°C for 30 min, air-dried and resuspended in 12µl H_2O for next-generation sequencing. Illumina HiSeq® sequencing was conducted by the Kinghorn Centre for Clinical Genomics, Garvan Institute, Darlinghurst, Sydney.

2.11.2 NGS data analysis

All bioinformatic processing was performed by Paul Young (Suter Laboratory, VCCRI). Raw sequence data files were obtained in FASTQ format. Using cutadapt (http://code.google.com/p/cutadapt/), sequencing adapters were trimmed by searching for the adapter sequence "GATCGGAAGAGCACACGTCTGAACTCCAGTCAC" and reads with a trim length below 18bp were removed at this step. The remaining trimmed reads were aligned to the most recent assembly (December 2011) of the mouse genome (mm10/GRCm38) using Bowtie (Langmead, 2010) with the following parameters: seed length 18; 1 mismatch in seed; maximum quality score of mismatched bases 70. Mapped reads stored in SAM/BAM format and were then

filtered for one mismatch to the genome across the entire length of the read. Annotation of each read was performed using genomic coordinates of genomic features obtained from the following sources: miRNA from mirBase v20, RefGene, Nested Repeats and tRNA data sets from UCSC Golden Path (http://hgdownload.soe.ucsc.edu/downloads.html). The genomic locations of 2750 reported murine piRNA clusters were obtained from piRNABank ((Sai Lakshmi and Agrawal, 2008), http://pirnabank.ibab.ac.in/) and these coordinates were converted for compatibility with the most recent version of UCSC Genome Browser (Dec. 2011, GRCm38/mm10, http://genome.ucsc.edu/cgi-bin/hgGateway) using NCBI Remap (https://www.ncbi.nlm.nih.gov/genome/tools/remap). Reads overlapping these clusters were annotated as falling into piRNA clusters. Sequences failing to overlap coordinates from these data sets were labelled as "unannotated".

2.12 List of buffers and reagents

HEPES-buffered saline (HBS) (2x)

- 0.28M NaCl
- 0.05M HEPES
- Na_2HPO_4

Made up to volume with MilliQ H_2O and pH adjusted to exactly 7.15. Solution was filtered using a 0.2µm Minisart filter (Sartorius, 16534K) and stored at -80°C. Once thawed, HBS was stored at 4°C up to three months without any reduction in transfection efficiency.

0.5M $CaCl_2$ (4x)

- 0.5M $CaCl_2$
- 1M HEPES, pH 7.3

Solution was made up to volume with MilliQ H_2O, filtered using a 0.2µm Minisart filter (Sartorius, 16534K) and stored at -80°C. Once thawed, HBS was stored at 4°C up to three months without any reduction in transfection efficiency.

Fractionation Buffer A

- 10mM HEPES

- 10mM KCl
- 1.5mM $MgCl_2$
- 0.1mM Ethylenediaminetetraacetic acid (EDTA)

Immediately prior to use, added 1mM DL-Dithiothreitol (DTT) and cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001).

Fractionation Buffer C

- 50mM Tris-HCl (pH 7.5)
- 0.5% Triton X-100
- 137.5mM NaCl
- 10% Glycerol
- 5mM EDTA
- 0.5% SDS

Immediately prior to use, added 1mM DL-Dithiothreitol (DTT) and cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001).

RIPA buffer

- 50mM Tris-HCl (pH 8.0)
- 150mM NaCl
- 1% NP-40
- 0.5% sodium deocycholate
- 0.1% SDS

Immediately prior to use, added 1mM DL-Dithiothreitol (DTT) and cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001).

Polysome Lysis Buffer

Obtained from (Keene et al., 2006)

- 100mM KCl
- 5mM $MgCl_2$

- 10mM HEPES (pH 7.0)
- 0.5% NP-40

Immediately prior to use, added 1mM DL-Dithiothreitol (DTT), cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001) and 40U/ml RNaseOUT ribonuclease inhibitor (Life Technologies, 10777-019).

NT2 buffer

- Obtained from (Keene et al., 2006)
- 50mM Tris-HCl (pH 7.5)
- 150mM NaCl
- 5mM $MgCl_2$
- 0.05% NP-40

Immediately prior to use, added 1mM DL-Dithiothreitol (DTT), cOmplete Mini Protease Inhibitor Cocktail (Roche, 11836170001) and 40U/ml RNaseOUT ribonuclease inhibitor (Life Technologies, 10777-019).

3. Miwi2 expression and function in murine embryonic stem cells

Introduction

Functional studies of the Piwi/piRNA pathway in mammals have largely been restricted to the germline. However, expression of Piwi genes and proteins in somatic cells has been long described – one of the first reports of the Piwi pathway was in the somatic cells of the *Drosophila* ovary (Cox et al., 1998). *Piwi* is expressed in the somatic follicle cells enveloping the developing oocyte; these cells exhibit primary, but not secondary piRNA biogenesis (Pelisson et al., 2007, Malone et al., 2009). Given that secondary piRNA biogenesis occurs as a by-product of primary piRNA-mediated "slicing" of transposon transcripts in *Drosophila* (Brennecke et al., 2007) and mice (Aravin et al., 2008), a repeat silencing-independent function of the Piwi/piRNA pathway in somatic cells is feasible, particularly in organisms such as the vertebrates where transposons are usually stably silenced by DNA methylation in somatic cells.

A number of studies have identified expression of members of the Piwi family in somatic cells of primitive organisms: these Piwi orthologues have been implicated in the regulation of stem-like features of regenerative cells in these organisms (Reddien et al., 2005, Alie et al., 2011, Funayama et al., 2010, Rinkevich et al., 1995). Expression of members of the Piwi family has also been reported in mammalian somatic multipotent stem cells. For example, *Hiwi* is expressed in human CD34+ haematopoietic stem cells and decreases upon their differentiation (Sharma et al., 2001). *Miwi2* expression is also enriched in stem-like cells in the mouse bone marrow – such as haematopoietic stem cells, multipotent progenitors and granulocyte-macrophage progenitors – relative to differentiated cell types including red blood cells, macrophages, neutrophils, B-cells, and T-cells, where *Miwi2* is not expressed (Nolde et al., 2013). A subset of piRNAs has also been shown to be expressed in mouse hippocampal tissue and in cultured hippocampal neurons, where they associate with Miwi protein (Lee et al., 2011).

Given the role of Piwi family members in maintaining stem-like cells in primitive organisms together with their enrichment in stem-like progenitors in mammalian somatic tissues relative to differentiated cells, I hypothesised that one or more of the three murine *Piwi* genes are expressed in murine embryonic stem cells or somatic tissues of the mouse (other than the hippocampus and bone marrow), along with piRNAs. A positive result would lead to the question of whether somatic Piwi proteins bind piRNAs in somatic cells.

Therefore, the specific aims of this chapter are to:

i. Determine whether members of the Piwi family are expressed in murine somatic tissues and embryonic stem cell lines
ii. Observe Piwi family member transcript expression changes during embryonic stem cell *in vitro* differentiation
iii. Investigate the phenotypic effects and gene and small RNA expression changes in murine embryonic stem cells following knockdown of somatic mouse Piwi expression.

I examined the transcript level of the three murine Piwi genes in various murine somatic cell types and ES cells using quantitative RT-PCR (qPCR). Based on these results, I focused further investigation on Miwi2 protein expression and localisation using western blot and immunofluorescence. I established an ES cell *in vitro* differentiation model to investigate *Miwi2* expression dynamics in relation to normal gene and small RNA expression changes during lineage specification. Finally, I employed lentiviral short hairpin RNA (shRNA) knockdown of *Miwi2* in ES cells to observe phenotypic changes, as well as changes to gene and small RNA expression profiles, across an *in vitro* differentiation trajectory.

Results

3.1 Murine *Piwi* transcript expression in somatic tissue and cell lines

I surveyed expression of the three murine Piwi transcripts, *Miwi*, *Mili*, and *Miwi2*, in a panel of adult somatic tissues and mouse embryonic stem cell lines. Unless otherwise stated, tissues were isolated from 21 day-old (D21) C57Bl/6J mice. I chose tissues with varying regenerative potential: spleen, liver, kidney and heart. The liver is highly regenerative (Overturf et al., 1997, Fausto, 2000) and the kidney moderately so (Gupta et al., 2002), whereas the heart consists predominately – in terms of mass – of terminally differentiated cardiomyocytes that essentially do not divide in the adult (Jacobson and Piper, 1986). The murine spleen is a site of haematopoiesis and thus, contains multipotent stem and progenitor cells (Wolber et al., 2002).

As this analysis compares murine Piwi transcript expression in tissues and cells with varying degrees of metabolic and proliferative activity and regenerative capacity, expression levels of commonly used reference genes such as *Rpl13a* and *β-actin* are not suitable. I generated a list of

top reference gene candidates for the tissues and cell lines of interest using Genevestigator V3 (Hruz et al., 2008); Calnexin (*Canx*) and Ubiquitin-conjugating enzyme E2Z (*Ube2z*) displayed similar stable patterns of expression across relevant tissues, hence these genes were selected as the reference for all qPCR experiments.

I used RNA from mouse testis as a positive control in all analyses: *Mili* is expressed from embryonic day 12.5 (E12.5) and *Miwi* is detectable from 14 days post-partum; both are expressed in the adult testis (Aravin et al., 2008). This group reported that Miwi2 is most highly expressed during a specific developmental window from embryonic day 15.5 to postnatal day 3. However, I have found that *Miwi2* transcript expression persists into adulthood (see Figure 8-1, Appendix). I therefore used D21 testis in all murine Piwi transcript analyses as a positive control; I also used neonatal (D1) testis in both *Miwi2* analyses and D19 testis in the qPCR for *Miwi2* in ES cells.

I used two embryonic stem cell lines in this study: W9.5 ES cells, derived from 129S1/Sv mice (Jackson Labs), and R1 ES cells, derived from a male blastocyst from the 129X1 x 129S1 strain (Nagy et al., 1993). Prior to qPCR for Piwi family member transcript expression, I used fluorescence-activated cell sorting (FACS) to deplete feeder MEFs and enrich the ES cell population for pluripotent cells by selecting for expression of the pluripotent cell surface marker, stage-specific embryonic antigen 1 (SSEA1) (Solter and Knowles, 1978, Knowles et al., 1978). Gating strategy for collection of $SSEA1^+$ ES cells is shown in Figure 8-2, Appendix.

Relative to D21 testis, *Mili* and *Miwi* expression levels were either undetectable or negligible in both mouse embryonic stem (mES) cells and somatic tissues (Figure 3-1). However in $SSEA1^+$ mouse embryonic stem cells, *Miwi2* transcript expression approaches that of the testis. Notably, expression of *Miwi2* is enriched in pluripotent W9.5 ES cells selected for SSEA1 expression relative to unsorted W9.5 or unsorted R1 ES cells. *Miwi2* transcripts were detectable in somatic tissues at around 10% of the levels observed in the neonatal testis.

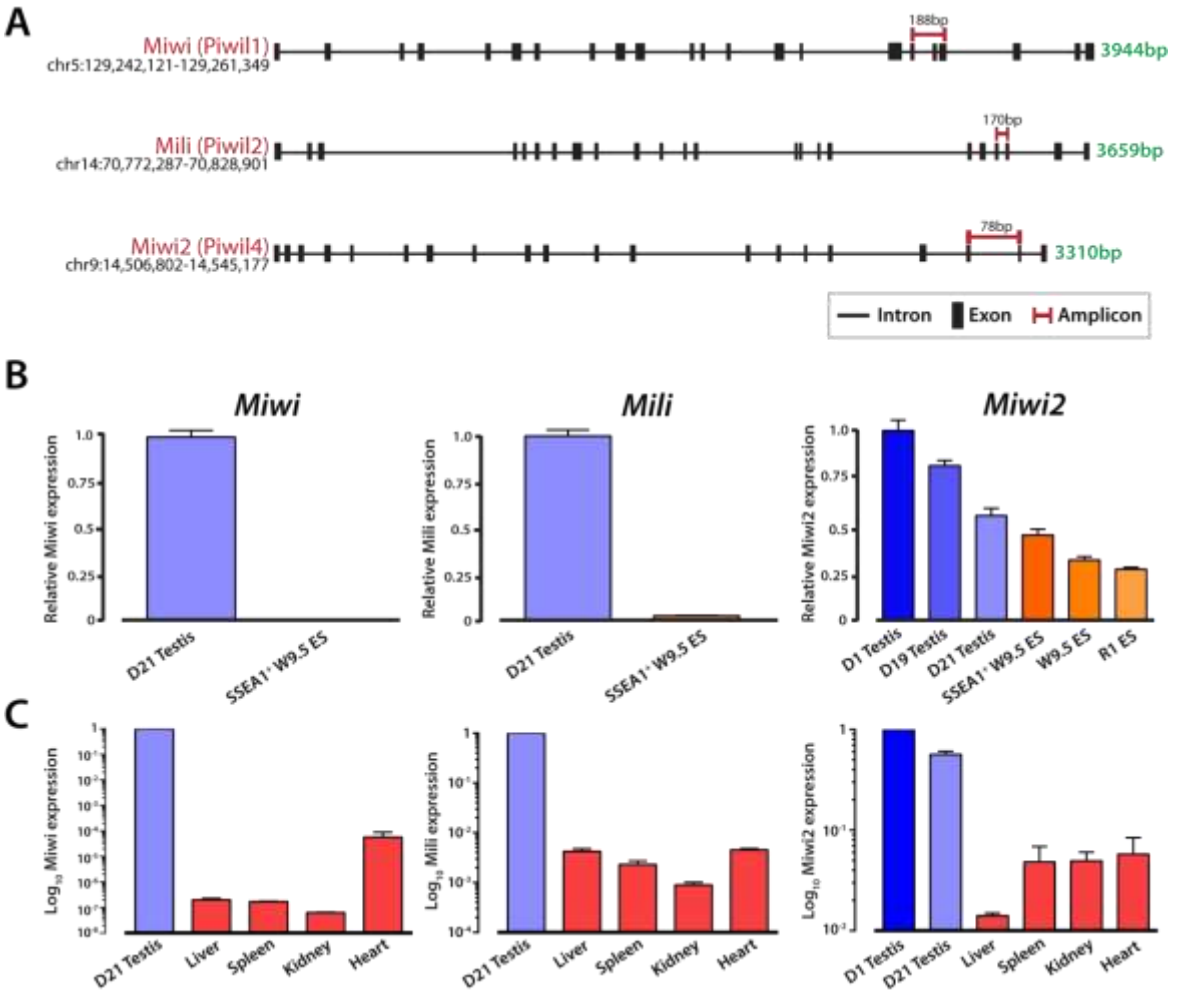


Figure 3-1: Murine Piwi transcripts can be detected in embryonic stem cells and somatic tissues

(A) Schematic representation of genomic loci for *Miwi, Mili,* and *Miwi2* with UCSC-annotated intron and exon boundaries. Representative quantitative PCR primer binding sites with amplicon sizes are shown; each primer set spans at least one exon-exon boundary. (B-C) Quantitative PCR data for *Miwi, Mili,* and *Miwi2* expression. All gene expression is shown relative to testis. All values are averaged from four replicate PCR reactions. Error bars represent standard error of the mean. (B) Expression in testis and murine embryonic stem cell lines. (C) Expression in testis and somatic tissues. Y-axis values are depicted on a log scale.

I also investigated *Miwi* (Piwil1)*, Mili* (Piwil2) and *Miwi2* (Piwil4) transcript expression in a number of murine tissues using data from the ENCODE (ENCODE Project Consortium, 2011), visualised with the UCSC Genome Browser graphical interface (July 2007, NCBI37/mm9, http://genome.ucsc.edu/cgi-bin/hgGateway). Results are depicted in Figure 3-2. It is important to note that in this analysis I did not compare levels of the three murine Piwi family members to each other; the scale of the tracks is adjusted for each transcript according to expression in 8 week old (adult) testis. Relative to testis, there is negligible expression of Miwi and Mili in a range of 8 week old mouse somatic tissues. But *Miwi2* transcription occurs across many tissues: spleen, adrenal gland, bladder, colon, kidney, large intestine, lung, mammary gland, ovary, placenta, and small intestine. There is low-level *Miwi2* expression in the cerebellum and heart, and negligible expression in the liver. While data from neonatal testis tissue was not publicly available at the time of analysis, this data highlights the almost-ubiquitous expression of *Miwi2* transcript in somatic tissues relative to testis, which is consistent with my qPCR findings.

Figure 3-2: ENCODE data for *Miwi, Mili* and *Miwi2* transcripts in testis and somatic tissues.

Image generated using UCSC genome browser with ENCODE wiggle (WIG) format tracks displayed. *Miwi* (Piwil1), *Mili* (Piwil2) and *Miwi2* (Piwil4). RefSeq genes are displayed above RNA-seq tracks

showing transcript from the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (ENCODE/CSHL). Peaks represent regions of *Miwi2* transcript enrichment. Within each transcript analysis, all tracks are viewed on the same scale, adjusted relative to expression in the 8 week Testis. Y-axis vertical viewing range for *Miwi* and *Miwi2* tracks min = 0, max = 15. Y-axis vertical viewing range for *Mili* track min = 0, max = 800.

Miwi2 protein expression in somatic tissue and ES cells

The presence of *Miwi2* mRNA in somatic tissue and embryonic stem cells implies the presence of Miwi2 protein. Here I sought to confirm this and determine the cellular localisation of Miwi2 before proceeding with functional analysis. Miwi2 has been reported to localise predominately to the nucleus of male germ cells, although the protein is restricted to the cytoplasm in the absence of Mili (Aravin and Bourc'his, 2008). Furthermore, in *Drosophila,* it has been demonstrated that Piwi function can differ depending on its cellular localisation (Klenov et al., 2011).

Miwi2 expression in somatic tissue and localisation in mouse embryonic stem cells

There are few commercially-available antibodies for immunodetection of Miwi2. I selected specific lots of a rabbit polyclonal antibody against Miwi2 (Abcam ab21869, lots #841630 and #GR130323-1) for western blot detection, as I had found previously that they were the most reliable. I have performed experiments to confirm specificity of this antibody (see Figure 8-3, Appendix). For positive controls for Miwi2 detection, I used D19/D21 and/or neonatal (D1) testis total lysate and/or HEK293FT cells transiently overexpressing a FLAG-tagged version of Miwi2 (293FT-Miwi2-FLAG). Miwi2 protein (98kDa) was detected in testis and 293FT-Miwi2-FLAG cells, as well as spleen and low levels in the heart of D21 C57BL/6 mice (Figure 3-3A). *Miwi2* is also translated into protein in W9.5 mouse ES cells; its expression appears less in embryoid bodies (Figure 3-3B).

I also performed nuclear/cytoplasmic fractionation of W9.5 ES cells and found that the protein is predominately cytoplasmic, although there appears to be a small amount present in the nucleus (Figure 3-3B). However, a small amount of β-tubulin is detected in the nuclear lysate, indicating slight cytoplasmic contamination of the nuclear fraction. It is therefore possible that small amounts of cytoplasmic Miwi2 could have falsely appeared in the nuclear fraction. There is no band present for H2B in the D19 testis sample as the antibody detects somatic H2B only.

I also performed immunofluorescence on ES cell colonies using the same anti-Miwi2 antibody to visualise Miwi2 expression *in situ*. Multiple negative controls were used per experiment: ES cells

were either stained with a peptide-blocked α-Miwi2 antibody or the secondary antibody only; Figure 3-3 shows the peptide-blocked control only. In agreement with the western data, the immunofluorescence shows that Miwi2 resides predominately in the cytoplasm in W9.5 ES cells (Figure 3-3C). Miwi2 expression is punctate and appears to be perinuclear. Some nuclear staining is apparent; this could be attributed to fluorescence bleed-through from perinuclear Miwi2 present in layers of the ES cell section immediately above or below the focal point of interest, or it may represent a small amount of protein that is present in the nucleus.

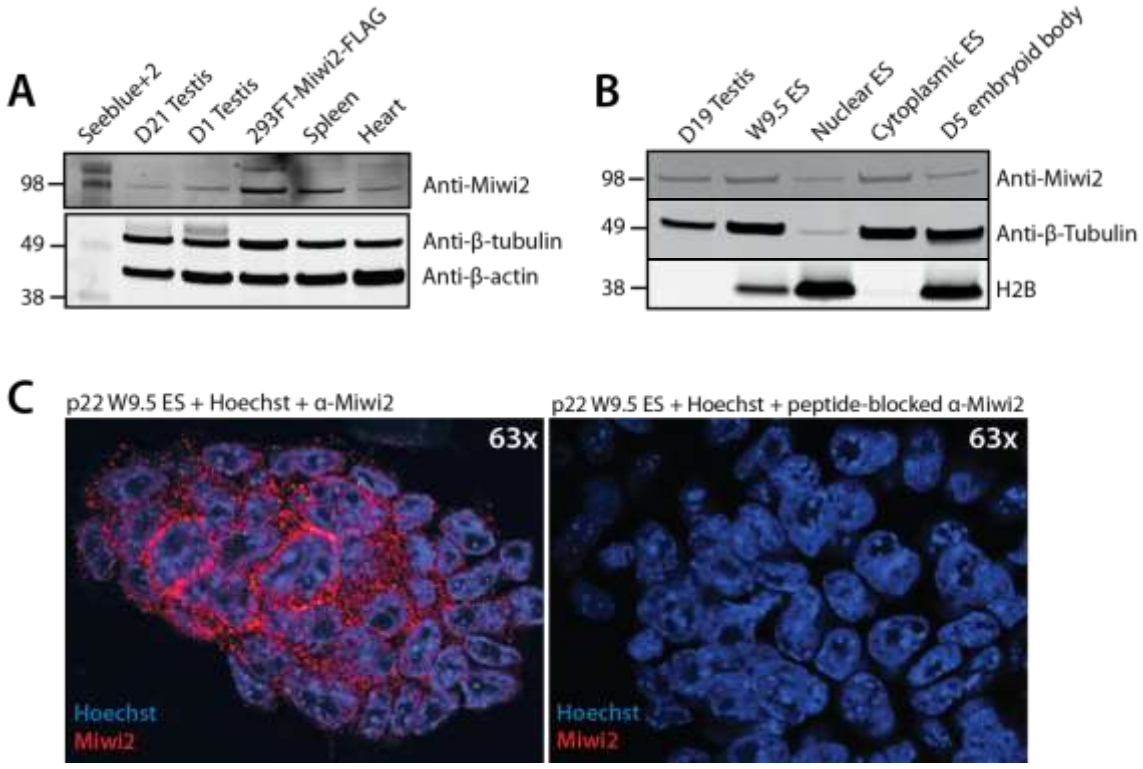


Figure 3-3: Miwi2 protein is present in somatic tissue and cytoplasmic in ES cells

(A) Western blot showing Miwi2 expression (top panel) in whole testis and somatic cells and tissue. B-tubulin and β-actin (bottom panel) were used as loading controls. (B) Western blot showing Miwi2 expression in testis, whole and fractionated mouse ES cells and embryoid bodies. β-tubulin and somatic H2B were used as loading controls and also to assess separation of cytoplasmic and nuclear fractions. (C) Immunofluorescence showing cytoplasmic localisation of *Miwi2* in W9.5 ES cell colonies. Antibody specificity for Miwi2 is demonstrated using negative control ES cells stained with a peptide-blocked α-Miwi2 antibody.

3.2 Miwi2 expression during *in vitro* differentiation

If – as my hypothesis predicts – Piwi gene and protein expression promotes a stem-like phenotype, it would follow that Miwi2 expression decreases upon cellular differentiation. Such a decrease has been observed in other studies of stem and stem-like cells. For example, Piwi expression is enriched in stem-like cells compared to their differentiated counterparts in primitive organisms such as *S. mediterranea, P. pileus, E. fluviatilis* and *B. leachi* (Reddien et al., 2005, Alie et al., 2011, Funayama et al., 2010, Rinkevich et al., 1995). In humans, *Hiwi* is expressed in human $CD34^+$ multipotent haematopoietic stem cells, which have the potential to form a number of downstream lymphoid and myeloid blood cell types (Sharma et al., 2001). Upon differentiation, *Hiwi* transcript is no longer detectable. Hence, I investigated whether *Miwi2* – which is expressed in pluripotent mouse ES cells – also decreases during murine lineage specification.

Miwi2 *expression decreases during* in vitro *differentiation*

In preliminary work, I found that Miwi2 protein expression was decreased in 5 day embryoid bodies relative to undifferentiated ES cells (Figure 3-3B). Thus, I aimed to pinpoint the time point at which *Miwi2* transcript expression changes in the context of the well-characterised changes in pluripotency gene expression during ES cell differentiation. To do this, I designed qPCR primers to amplify transcript regions of the pluripotency genes *Oct4, Nanog* and *Sox2* in addition to a number of germ layer markers: the ectodermal markers *Nestin* and *Otx2*, the mesodermal genes *Brachyury* (*Bry/T*) and *PECAM/CD31*, and the endodermal markers *HNF4α* and *Sox17*. I performed *in vitro* differentiation of ES cells over 5 days using the hanging drop method (see Section 2.4.4, Materials and Methods).

Miwi2 qPCR results in W9.5 ES cells during differentiation are depicted in Figure 3-4. *Miwi2* expression begins to sharply decrease after Day 1 of *in vitro* differentiation and is negligible from Day 3 onwards. This drop in *Miwi2* expression precedes a change in *Oct4*, occurs after an initial decrease in *Nanog* expression and is concomitant with diminishing levels of *Sox2*. *Oct4* and *Sox2* expression decreases as expected (Kim et al., 2008); *Nanog* levels decrease immediately, sharply revert to pluripotent expression levels on Day 3, and then exhibit a steady decrease after Day 3.

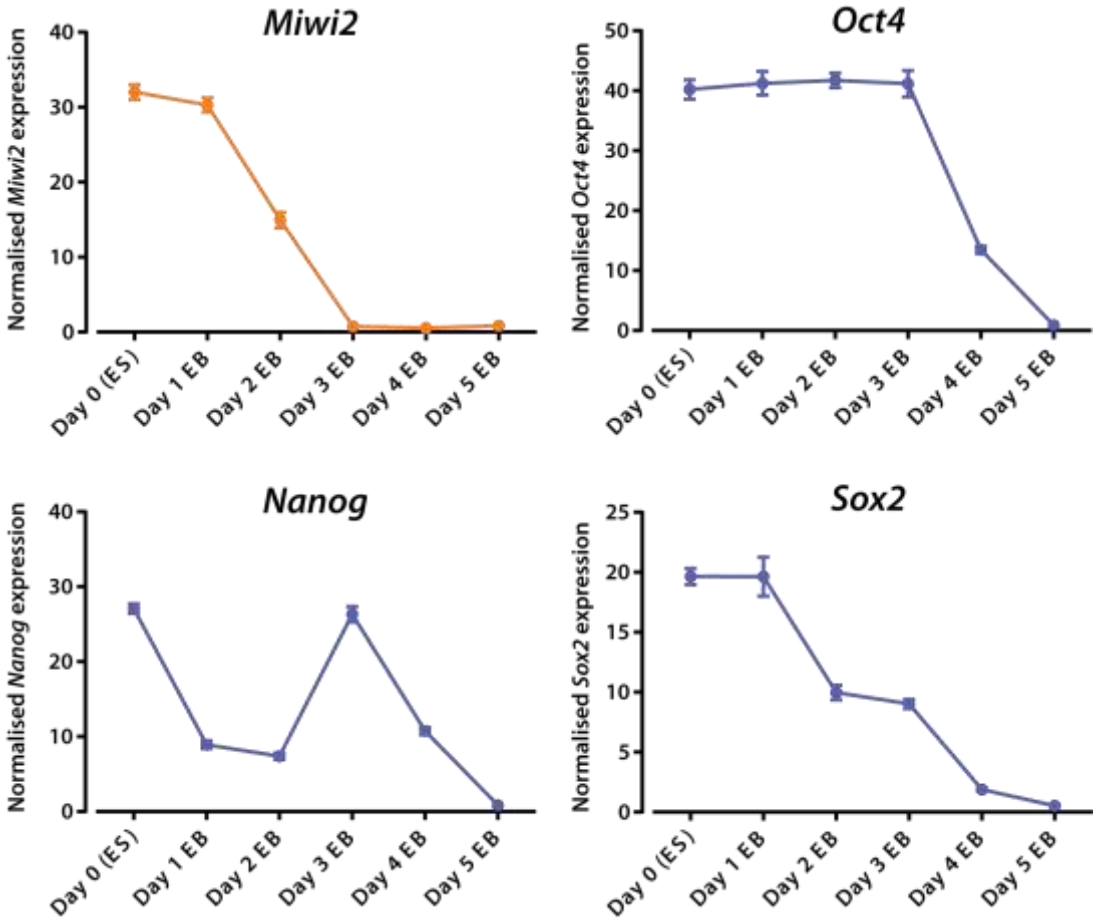


Figure 3-4: *Miwi2* and pluripotency gene expression decreases during differentiation

Quantitative RT-PCR data representing transcript expression of *Miwi2, Oct4, Nanog* and *Sox2* during normal *in vitro* differentiation of murine embryonic stem cells. Values are the average of technical triplicate PCR reactions; error bars represent standard error of the mean. Y-axis values are normalised expression values derived from Comparative C_T analysis. Raw C_T values were normalised to *Canx* and *Ube2z* reference genes. Values are different on each axis as transcript abundance varies for each gene.

I then examined expression of differentiation genes to i) confirm that differentiation is proceeding as expected in this system, and ii) place diminishing *Miwi2* expression in the context of primitive germ layer formation. I selected two primitive markers for each embryonic germ layer for expression analysis; quantitative RT-PCR data is presented in Figure 3-5.

The ectoderm is the first germ layer to emerge during *in vivo* gastrulation (Lawson et al., 1991), and this is reflected in my *in vitro* system as I observed peaks in *Nestin* and *Otx2* expression by Day 3 (Figure 3-5A). The primitive mesoderm marker Brachyury (*Bry/T*) peaks at Day 4 but the mature mesodermal gene Cluster of Differentiation 1 (*CD31, PECAM*) exhibits a 20-fold decrease in expression from Days 0-4 (Figure 3-5B). The endodermal lineage is the final germ layer to emerge in the *in vitro* differentiation model (Figure 3-5C). Expression of primitive endodermal markers Hepatocyte Nuclear Factor 4 alpha (*HNF4α*) and SRY (sex determining region Y)-box 17 (*Sox17*) rises dramatically after Day 3 of differentiation. These gene expression changes are depicted collectively and schematically in relation to *Miwi2* in Figure 3-6.

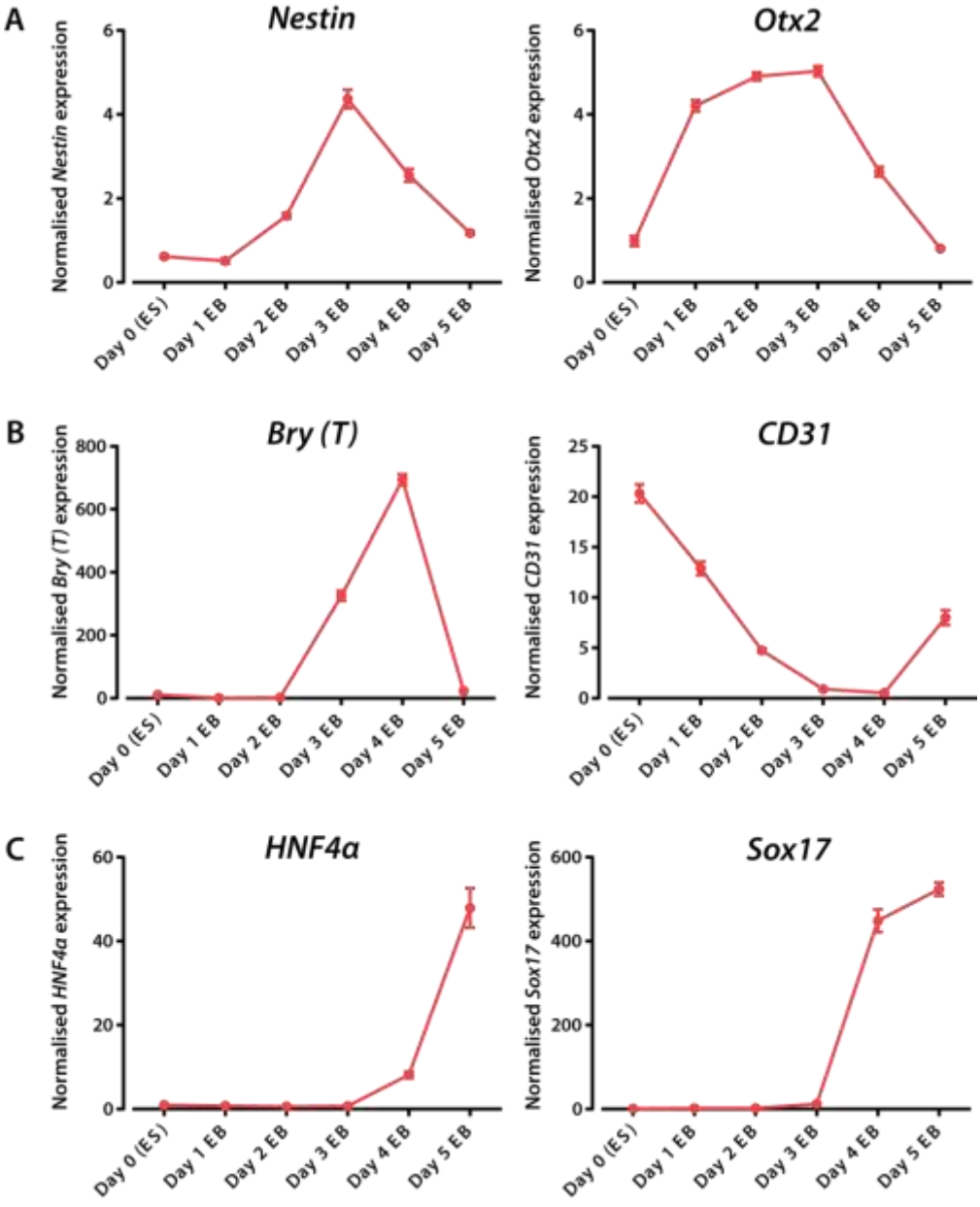


Figure 3-5: Germ layer marker gene expression during *in vitro* differentiation

Quantitative RT-PCR data of primitive germ layer markers during normal *in vitro* differentiation of murine embryonic stem cells. (A) Ectodermal markers *Nestin* and *Otx2*. (B) Mesodermal markers *CD31* (PECAM) and *Brachyury* (Bry/T). (C) Endodermal markers *HNF4α* and *Sox17*. Values are the average of technical triplicate PCR reactions; error bars represent standard error of the mean. Y-axis values are normalised expression values derived from Comparative C_T analysis. Raw C_T values were normalised to *Canx* and *Ube2z* reference genes. Values are different on each axis as transcript abundance varies for each gene.

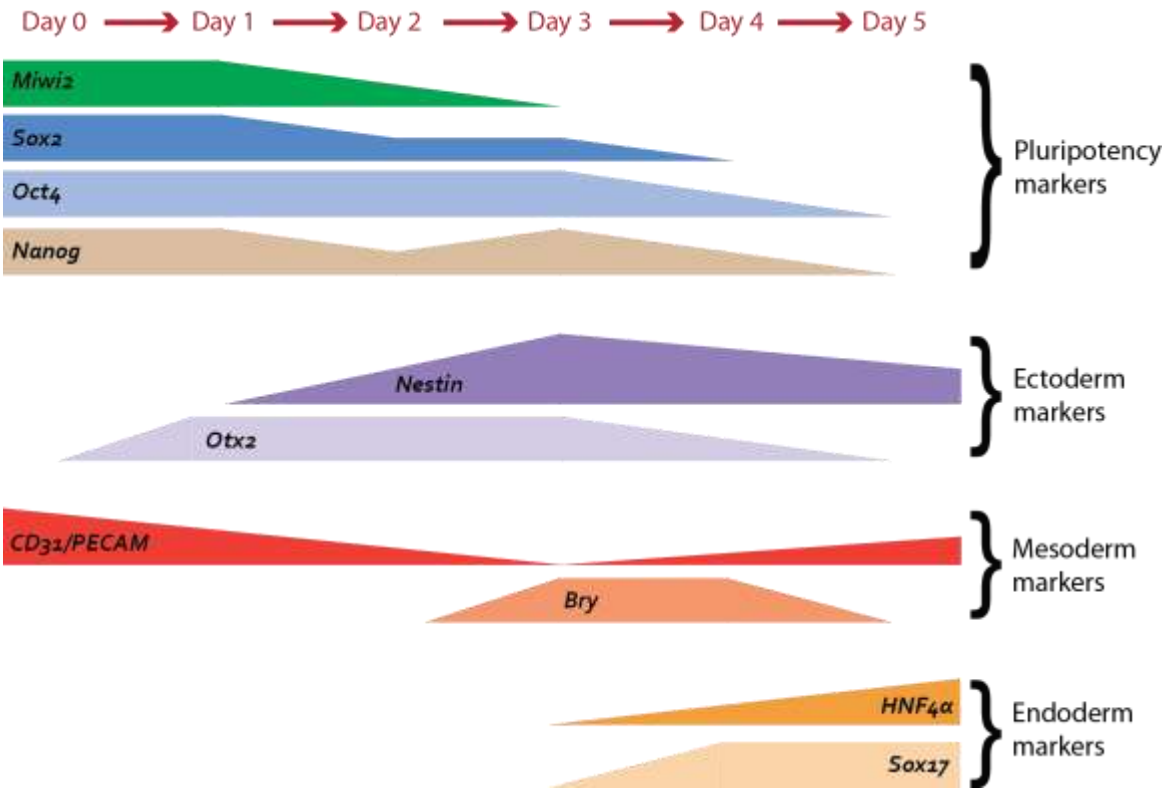


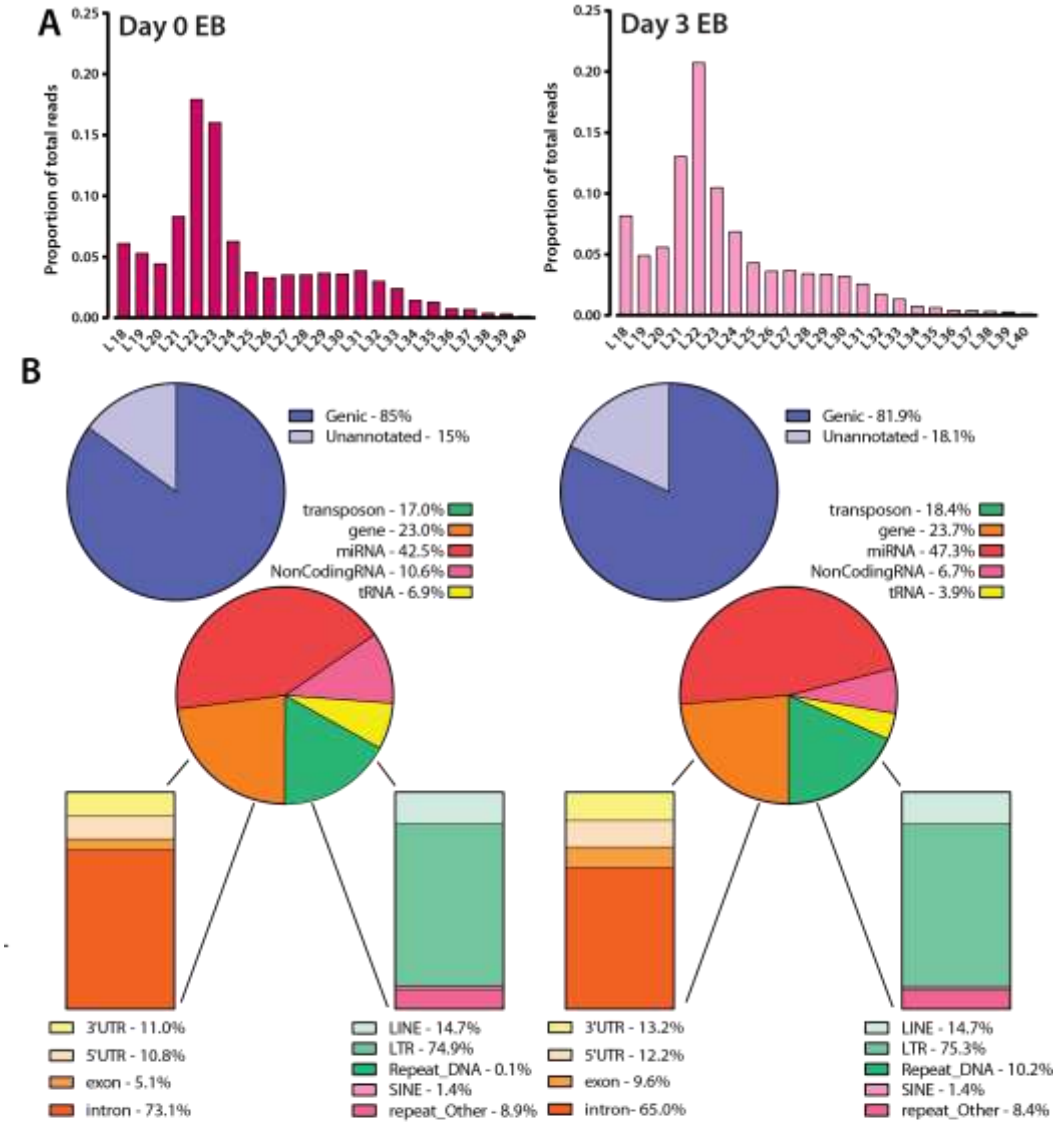
Figure 3-6: Schematic of *Miwi2* gene expression changes in the context of normal differentiation

Image drawn from qPCR data for individual gene expression during 5 days *in vitro* differentiation of R1 mouse embryonic stem cells.

3.3 Small RNA expression during in vitro differentiation

I performed small RNA sequencing of ES cells (Day 0) and Day 3 embryoid bodies to complement the targeted gene expression analysis (Figure 3-4 and Figure 3-5) and provide some extra molecular context to decreased *Miwi2* expression I observed after three days of *in vitro* differentiation. Furthermore, I wished to investigate whether there was a concurrent decrease in pi-like RNAs alongside the disappearance of *Miwi2* expression during ES cell differentiation. I chose Day 0 and Day 3 time points for analysis as they correspond to high and negligible expression of *Miwi2*, respectively (Figure 3-4). For clarity, the ES cell sample will be referred to as "Day 0 embryoid bodies" in this chapter. I performed Illumina HiSeq® next-generation sequencing of total small RNA for each timepoint (see Section 2.11.1, Materials and Methods). Bioinformatic analysis was performed with Paul Young, Suter Laboratory. Following bioinformatic processing (see Section 2.11.2, Materials and Methods), I had over 10 million reads for my Day 0 EB sample and almost 15 million reads from the Day 3 EB sequencing (Table 8-1, Appendix). Of those reads between 18 and 50nt in length, the majority mapped successfully to the mouse genome after filtering for 1 error (65% for Day 0; 64% for Day 3). I obtained 73,356 unique sequences from the Day 0 sequencing and 83,037 from Day 3. My sequencing dataset included a small number of single copy reads (10% for Day 0 and 4% for Day 3), suggesting that the small RNA sequencing was not exhaustive. With the assistance of Paul Young (Suter Laboratory, VCCRI), I performed piRNA cluster analysis as described in Section 2.11.2, Materials and Methods. Of all sequences with greater than 10 reads in the Day 0 sample and Day 3 samples, 0.0637% and 0.0538% aligned with known piRNA clusters, respectively.

There are no overt differences in length distribution or genomic origin of small RNA reads between Day 0 and Day 3 samples (Figure 3-7). In both samples, the majority of small RNA reads are miRNA-sized (21-23nt, Figure 3-7A) and most map to annotated regions of the genome (Figure 3-7B). Most genic reads map to introns, and most repeat-derived reads originate from long terminal repeats (LTRs), particularly LTR-ERVL-MaLR repeats (Figure 3-7B and C). However, the proportion of tRNA-derived small RNA reads after 3 days differentiation (3.9%) is close to half that of Day 0 embryoid bodies (6.9%) (Figure 3-7B).



C

Repeat element	% all repeats
LTR_ERVL-MaLR	60.32%
LINE_L1	14.21%
rRNA_rRNA	8.16%
LTR_ERVK	9.29%
LTR_ERVL	0.40%
LTR_ERV1	3.22%
SINE_B2	0.62%

Repeat element	% all repeats
LTR_ERVL-MaLR	65.24%
LINE_L1	14.34%
rRNA_rRNA	7.70%
LTR_ERVK	4.44%
LTR_ERVL	2.30%
LTR_ERV1	2.15%
SINE_B2	0.71%

Figure 3-7: Characterisation of total small RNA population during *in vitro* differentiation

Small RNA sequence characteristics of Day 0 and Day 3 embryoid bodies. (A) Length distribution of small RNA reads between 18-40nt. Values represented as a proportion of total reads. (B) Categorical annotations of small RNA reads. (C) List of most abundant repeat elements deriving small RNA reads.

To further characterise any relevant changes in small RNA composition I analysed those with large (greater than 10-fold) expression changes (Figure 3-8). I did this to i) reduce noise in my analysis as there were many large changes (5829 >2-fold changes) and ii) to account for the fact that I did not perform sequencing replicates for each sample and was therefore unable to perform T-tests to assess statistical significance of any changes. Read counts were converted to a proportion of total reads/sample and linear fold change calculated in Day 3 relative to Day 0 embryoid bodies. I filtered results for i) >100 reads in Day 0 for analysis of down-regulated genes, ii) >100 reads in Day 3 for analysis of up-regulated genes and iii) > ±10-fold change. In doing this, I found that 1865 small RNAs are up-regulated and 81 are down-regulated >10-fold in Day 3 relative to Day 0.

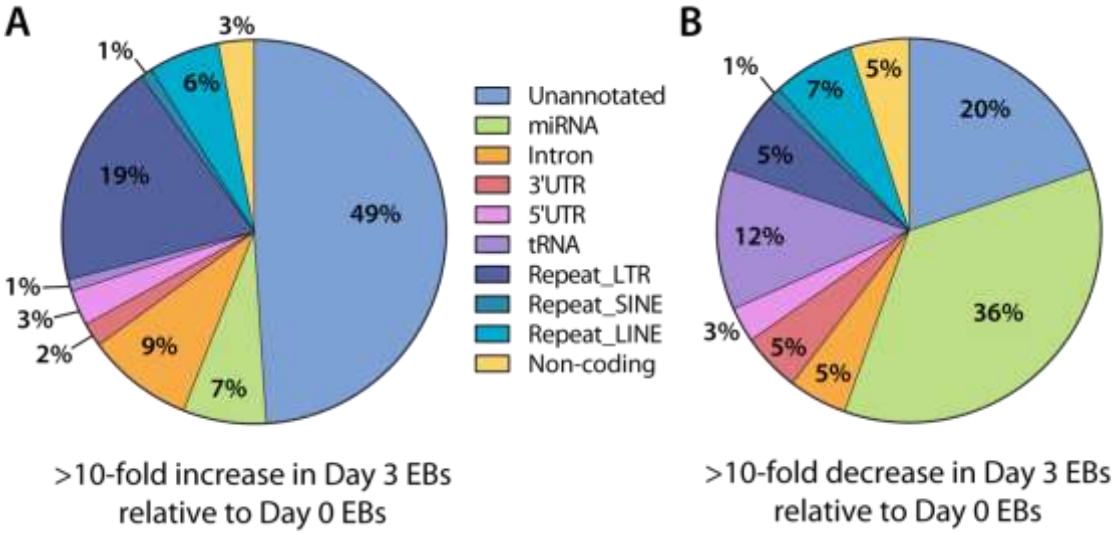


Figure 3-8: Small RNA expression changes during *in vitro* differentiation

Categorical annotations of unique sequences with expression changes in Day 3 relative to Day 0 embryoid bodies. (A) 10-fold increase. (B) 10-fold decrease.

Of those unique sequences with changes in expression, a large percentage of up-regulated (49%) and a moderate proportion of down-regulated (20%) sequences map to unannotated regions of the genome (Figure 3-8); this corresponds to 938 and 18 unique sequences, respectively. Of the 938 unannotated sequences that are up-regulated, there is a bias for smaller (<28nt) lengths. There are no overt pi-like RNA sequence features such as a bias for a uracil at position 1 (1U) or an adenine at position 10 (10A), as the up-regulated sequences exhibit an overrepresentation of A (57%) at position 1 and no overt nucleotide bias at position 10 (Figure 8-4, Appendix).

Annotated miRNAs comprised 7% of sequences with 10-fold increased expression and 36% of down-regulated sequences following 3 days differentiation (Figure 3-8). Multiple unique miRNA sequences map to individual miRNA genes: these include 24 down-regulated and 30 up-regulated miRNAs and are listed in Table 3-1 and Table 3-2, respectively.

Of down-regulated small RNA reads, 12% are derived from specific tRNAs: HisGTG, GluTCC, AspGTC and GlyGCC. These reads also exhibit a size bias and are either 19, 21-22 or 31-33nt in length (Figure 3-9). Small tRNA-derived RNAs stack at the 5' end of their parent tRNA gene: Figure 3-9C depicts reads aligning with tRNA1431-HisGTG. Furthermore, a moderate proportion (19%) of small RNAs that are up-regulated during differentiation map to LTR repeat elements. These are from 19-34nt in length with no over-representation of a particular length; they predominately map to LTR_ERVL-MaLR (MaLR) repeats, which are a type of transposable element enriched in 2-cell embryos (Macfarlan et al., 2012) (Figure 3-10). Figure 3-8 reveals that small RNAs with gene expression changes map to genic regions: 3'UTR, 5'UTR and introns. Most of these genic reads map to introns; Figure 3-11 depicts reads aligned with the intronic region of *Cam1kd*. Intriguingly, many small RNAs align with both introns and repeat elements; these 24 genes are listed in Table 3-3. Using Ingenuity Pathway Analysis, I performed pathway analysis of these genes: they are involved predominately in cellular assembly and organisation; cellular function and maintenance; carbohydrate metabolism; cell morphology; and cell-to-cell signalling and interaction.

Table 3-1: miRNAs with >10-fold decreased expression in Day 3 relative to Day 0 Embryoid bodies

miRNA ID	Reads at D0	Reads at D3	Fold Change
miR-1298-5p	960	8	-191.9
mir-367	1588	14	-181.4
mir-302a	142	2	-113.5
miR-302c-3p	19642	303	-103.7
miR-135a-1-3p	122	2	-97.5
mir-653	4962	144	-55.1
miR-302b-5p	702	22	-51.0
miR-302c-5p	1388	50	-44.4
miR-302b-3p	264090	9536	-44.3
miR-302d-3p	98573	3801	-41.5
miR-302a-3p	38927	1709	-36.4
miR-367-3p	1739	78	-35.7
miR-302d-5p	1093	59	-29.6
miR-302a-5p	64223	3564	-28.8
mir-302c	234	13	-28.8
miR-653-5p	1074	61	-28.2
miR-488-3p	304	19	-25.6
mir-141	122	9	-21.7
miR-135a-5p	205	16	-20.5
miR-106a-5p	73185	7295	-16.0
miR-141-5p	140	14	-16.0
miR-218-5p	16115	1749	-14.7
miR-20b-5p	141805	17155	-13.2
miR-452-5p	288	43	-10.7

Table 3-2: miRNAs with >10-fold increased expression in Day 3 relative to Day 0 Embryoid bodies

miRNA ID	Reads at D0	Reads at D3	Fold Change
miR-6238	1	425	265.8
miR-211-5p	3	665	138.6
miR-143-5p	1	144	90.0
miR-199a-5p	74	8386	70.9
miR-150-3p	1	112	70.0
miR-181d-5p	239	15128	39.6
miR-145a-5p	17	1064	39.1
miR-150-5p	82	4581	34.9
miR-214-5p	4	203	31.7
miR-214-3p	11	539	30.6
mir-297a-2	12	445	23.2
miR-129-1-3p	3	103	21.5
miR-143-3p	1252	42376	21.2
miR-199b-3p	105	3335	19.9
miR-125b-5p	138	4100	18.6
miR-199a-3p	210	5366	16.0
miR-1964-3p	7	174	15.5
miR-145a-3p	10	238	14.9
miR-615-3p	8	183	14.3
miR-871-3p	7	136	12.1
let-7d-3p	9	155	10.8
miR-365-3p	13	222	10.7
miR-181c-5p	124	2111	10.6
miR-100-5p	111	1862	10.5
let-7i-5p	377	6255	10.4
miR-3091-3p	7	115	10.3
miR-669e-5p	76	1240	10.2
let-7d-5p	27	436	10.1
miR-669f-3p	41	661	10.1
let-7a-5p	48	769	10.0

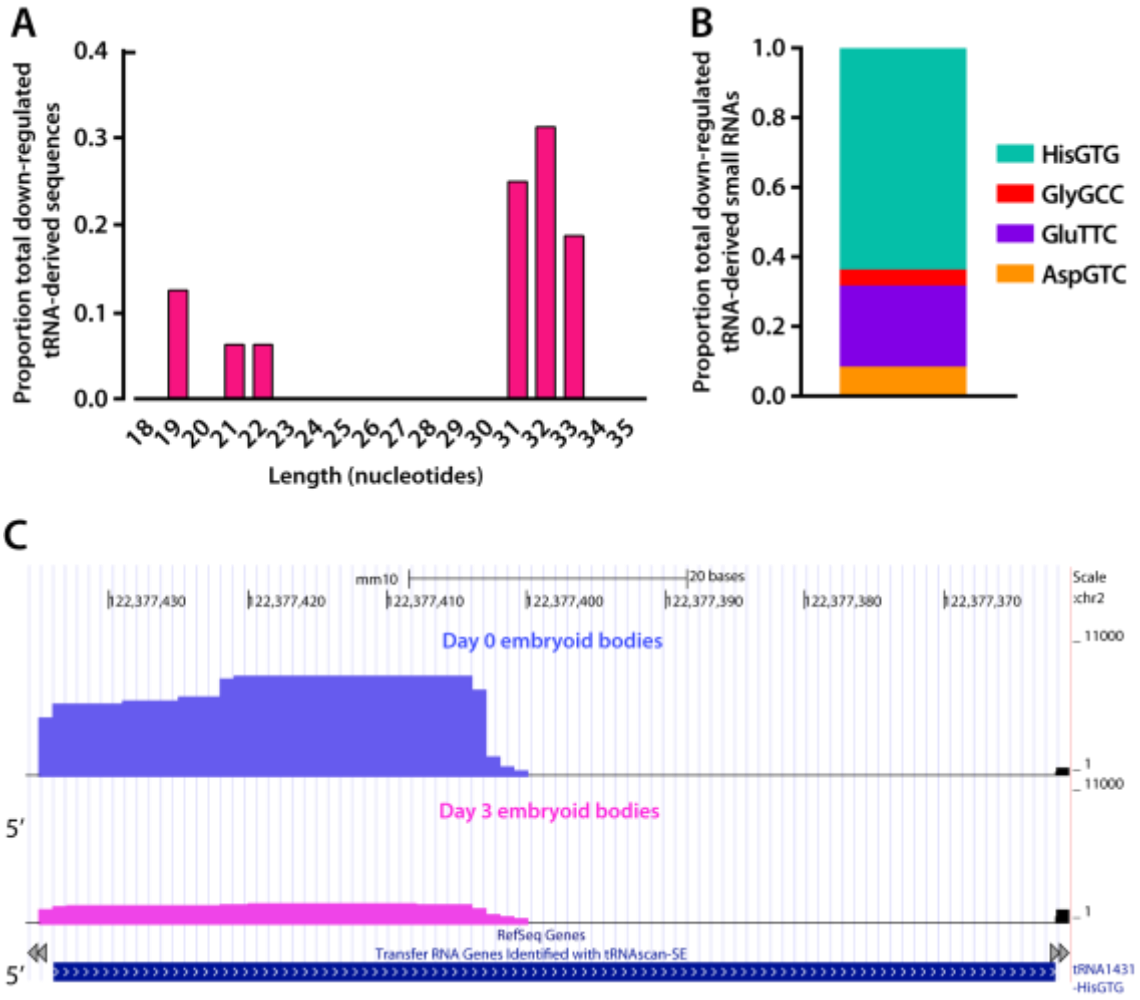


Figure 3-9: Sequence characteristics and genomic mapping of down-regulated tRNA-derived small RNAs

(A) Length distribution and (B) specific tRNA annotations presented as a proportion of total down-regulated tRNA-derived reads. (C) Graphical representation of tRNA1431-HisGTG-derived small RNAs from Day 0 and Day 3 embryoid bodies in WIG format aligned with the mouse genome (UCSC Genome Browser, GRCm38/mm10).

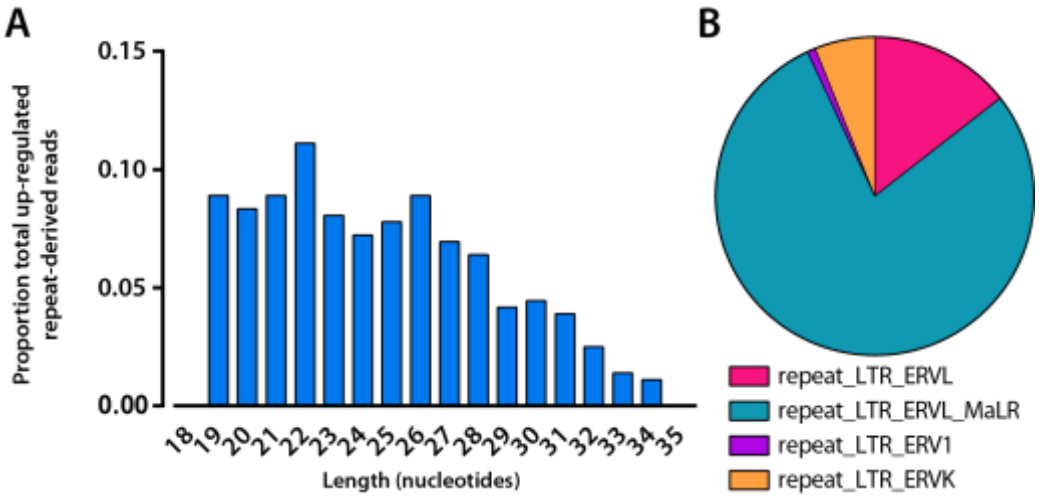


Figure 3-10: Sequence characteristics of up-regulated small RNAs derived from LTR repeats

(A) Length distribution and (B) specific LTR repeat annotations presented as a proportion of total up-regulated LTR repeat-derived reads.

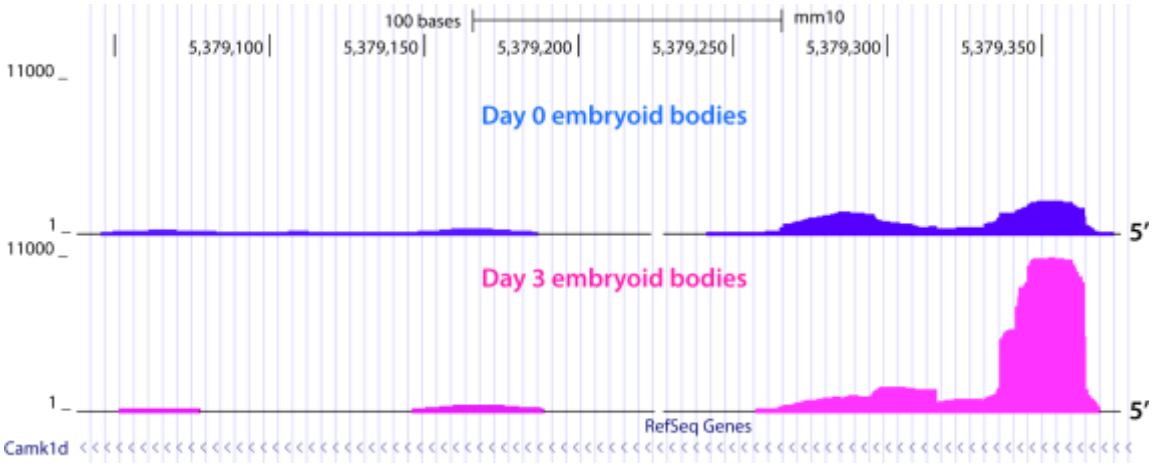


Figure 3-11: Small RNAs mapping to intronic regions of *Cam1kd*

Graphical representation of *Cam1kd* intron-derived small RNAs from Day 0 and Day 3 embryoid bodies in WIG format aligned with the mouse genome (UCSC Genome Browser, GRCm38/mm10).

Table 3-3: List of gene with introns giving rise to small RNAs that also map to repeat elements

Gene name	Repeat name
Arhgap21	repeat_LINE_L1
Cacna1i	repeat_LINE_L1
Camk1d	repeat_LINE_L1
Car5a	repeat_LINE_L1
Ccdc53	repeat_LTR_ERVL-MaLR
Cdk8	repeat_LTR_ERVK
Clasp2	repeat_LTR_ERVL-MaLR
Dclk2	repeat_LTR_ERVK
Enpp3	repeat_Satellite_Satellite
Gphn	repeat_rRNA_rRNA|piRNA_Cluster
Grip1	repeat_Simple_repeat_Simple_repeat
Itga9	repeat_rRNA_rRNA|piRNA_Cluster_AS
Jarid2	repeat_LTR_ERVL-MaLR
Kif3c	repeat_LINE_L1
Lpl	repeat_LTR_ERVL
Lx9	repeat_LINE_L1
Nrxn1	repeat_LTR_ERVL
Parp4	repeat_LINE_L1
Prkg1	repeat_LINE_L1|piRNA_Cluster
Psmg3	repeat_LTR_ERVK
Rapgef6	repeat_SINE_B4
Sh2b2	repeat_SINE_Alu
Slc9a4	repeat_Simple_repeat_Simple_repeat
Spata5	repeat_LINE_L1

In summary, the main findings of the small RNA profiling experiment include no overt change in expression of classic piRNAs, and very large changes in miRNA composition during *in vitro* differentiation. Furthermore, there is a halving in the total proportion of small RNAs derived from the 5' end of tRNAs, and most of these map to HisGTG. Finally, I observed an increase in expression of small RNA sequences mapping to i) MaLR repeats and ii) both introns and repeat elements in Day 3 relative to Day 0 embryoid bodies.

3.4 *Miwi2* knockdown in mouse embryonic stem cells

I have found that *Miwi2* expression decreases during *in vitro* cellular differentiation of ES cells when large-scale changes in pluripotency genes, lineage markers, and small noncoding RNAs are occurring. Several studies have demonstrated that RNAi-mediated knockdown of particular Piwi isoforms ablates stem-like properties of regenerative cells in primitive organisms such as *S. mediterranea* (Reddien et al., 2005, Palakodeti et al., 2008) and *B. leachi* (Rinkevich et al., 1995). Furthermore, defects in germline development observed in Piwi-null *Drosophila* have been attributed to aberrant differentiation of germline stem cells without self-renewal (Cox et al., 1998). Here, I asked whether knockdown of *Miwi2* in murine embryonic stem cells has a similar effect. It is unlikely that one would observe a complete exit from the pluripotent state, as Miwi2 mutant mice – despite exhibiting male sterility – are viable and appear to develop normally (Carmell et al., 2007). However, reduction of *Miwi2* may impart subtle changes in the maintenance of self-renewal and pluripotency *in vitro* that may not have been reported or even observed previously. Miwi2-knockout mice were engineered from heterozygous ES cells lacking only one copy of the Miwi2 allele (Carmell et al., 2007, Kuramochi-Miyagawa et al., 2008) and although ES cells would exist in Miwi2-null blastocysts, there are no published reports of *in vitro* ablation of Miwi2 function in mouse embryonic stem cells.

For knockdown experiments, I used lentivirus to deliver a plasmid encoding an inducible *Miwi2*-targeting shRNA sequence. Long terminal repeats (LTRs) flanking the gene of interest on the lentiviral vector assist in the genomic integration of foreign DNA and the introduced sequence is heritable through cell division. I was therefore able to engineer cell lines with stable integration of the α-*Miwi2* hairpin prior to global doxycycline-mediated induction of its expression and *Miwi2* knockdown. I found that W9.5 ES cells resisted *Miwi2* knockdown, hence all knockdowns and downstream analyses were performed in R1 mouse ES cells, which did not exhibit the same resistance.

All vector maps for plasmids used in lentiviral optimisation and *Miwi2* knockdown experiments are shown in Section 2.2, Materials and Methods. In this study I used three lentiviral vector backbones: pTRIPZ and pPGK-GFP were used for optimisation of lentiviral packaging and the pRSIT system was used for *Miwi2* knockdown. For packaging of the shRNA-containing vector into infectious lentiviral particles, I used a second-generation system, which refers to the use of two

additional helper plasmids: one encoding the viral packaging components Gag, Pol, Rev, and Tat (psPAX2) and the other expressing the viral envelope protein VSV-G (pMD.2G). For transfection of each lentiviral vector of interest plus helper plasmids into HEK293FT cells, I used the calcium chloride method: a HEPES-buffered solution is used to precipitate calcium phosphate and DNA and the insoluble precipitate attaches to the cell surface and is brought into the cells by endocytosis (Kingston et al., 2003). HEK293FT cells also express the SV40 Large T antigen, which binds and inactivates tumour suppressor proteins including Retinoblastoma (Rb) protein and p53, causing cells to leave G1 phase and enter S phase, promoting DNA replication (Ali and DeCaprio, 2001). This facilitates viral genome replication along with uncontrolled host cell DNA replication.

Optimisation of lentiviral packaging and mouse embryonic stem cell stable selection

I optimised multiple experimental parameters to maximise the amount of virus produced (viral titre). For optimisation, I used HEK293FT cells for both viral packaging and transduction. Given that fluorescent markers are encoded by all lentiviral vectors used in this study, I used RFP (pTRIPZ) and GFP (pPGK-GFP) expression as a surrogate for viral titre; these findings are depicted in Figure 3-12A. One of these parameters was plasmid size: packaging of a small lentiviral vector (pPGK-GFP, 6.9kb) results in considerably higher viral titre than a large plasmid (pTRIPZ, 13.7kb). The bulk of viral production occurs during the 24h period following calcium chloride transfection and can be further increased by manual concentration of virus-containing supernatant relative to neat supernatant. Figure 3-12B depicts an experiment performed by Nicole Schonrock (Harvey Laboratory, Victor Chang Cardiac Research Institute) which tests the effect of polybrene on virus-mediated uptake of a GFP-expressing vector into murine HL-1 cells. These findings demonstrate that the addition of 8µg/ml polybrene at the time of transduction results in greater transduction efficiency and can be attributed to the fact that polybrene minimises electrostatic repulsions between the lentiviral envelope and the host cell membrane (Davis et al., 2002). The effect of polybrene is also mimicked by centrifugation during transduction: centrifugal forces overcome these membrane electrostatic repulsions (Paya et al., 2006). Following centrifugation, I incubated cells and viral supernatant at 32°C to aid transduction efficiency, as it has been reported that the half-life of retroviruses is higher at 32°C than 37°C (Kotani et al., 1994, Lee et al., 1996, McTaggart and Al-Rubeai, 2000). I also established the optimal concentration of antibiotic for selection of stably-transduced ES cells using a dose-response curve (Figure 3-12C). R1 ES cells were concurrently exposed to multiple concentrations of puromycin and monitored over 8 days. The

optimal antibiotic concentration is the minimum dose that results in 100% cell death after several days. Thus, I selected for ES cells with stable expression using 1.0µg/ml puromycin for at least 96h.

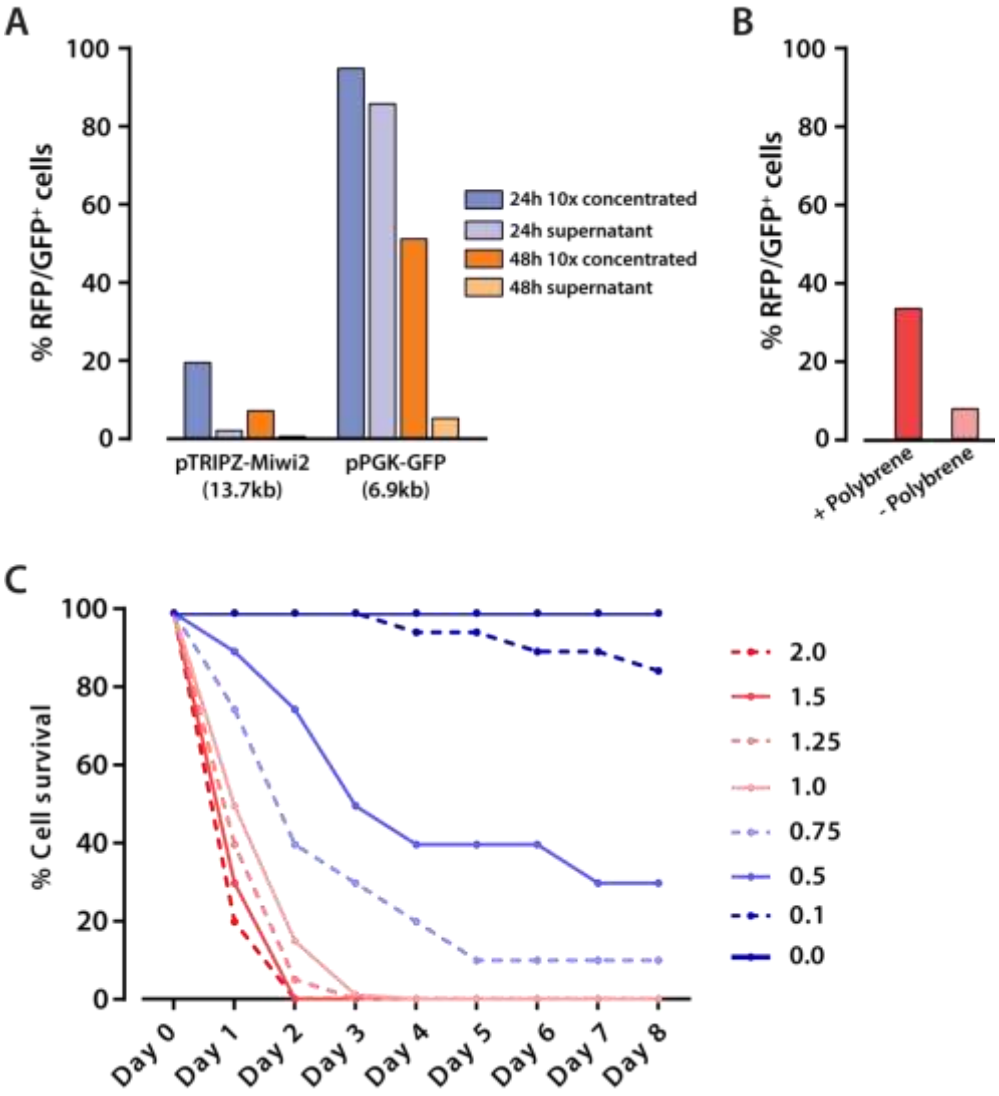


Figure 3-12: Optimal lentiviral packaging and transduction conditions and puromycin dose response in R1 mouse embryonic stem cells

(A) Effects of vector size, timing of viral supernatant harvest and concentration on viral titre using pTRIPZ-Miwi2 and pPGK-GFP. (B) Effect of polybrene on viral uptake. (C) Puromycin dose response curve for R1 mouse ES cells. Concentrations of puromycin inducing 100% cell death are depicted in shades of red; concentrations resulting in incomplete cell death are shown in shades of blue.

Miwi2 knockdown in mouse embryonic stem cells

Both pTRIPZ and pRSIT are lentiviral vectors that can be customised to encode α-*Miwi2* shRNA. Initial knockdown experiments were performed with pTRIPZ: this is a commercially-available vector set with doxycycline-inducible shRNA and RFP, plus resistance to puromycin. It is preferable to direct at least three hairpins against a gene of interest to control for off-target effects of random viral integration (Jackson and Linsley, 2010), however, the pTRIPZ system has only two different α-*Miwi2* shRNAs available. Furthermore, the pTRIPZ vector backbone is large (13.7kb) and I experienced limited success in preparing high-titre viral supernatant. For these reasons, in subsequent experiments I utilised the pRSIT lentiviral system with five vectors containing unique hairpins directed against *Miwi2*. The pRSIT system was not available at the time of lentiviral optimisation in HEK293FT cells (Figure 3-12A). The pRSIT vectors are fully customisable in terms of choice of promoter, doxycycline inducibility, fluorescent tag, and antibiotic resistance. Moreover, the vector backbone is much smaller (8.7kb), enabling high viral titre. I customised a vector backbone with constitutive GFP expression, puromycin resistance, and shRNA under the control of a doxycycline-inducible U6 promoter (see Section 2.2, Materials and Methods). Viral particles were produced containing each of the five α-*Miwi2* shRNA-containing vectors, all α-*Miwi2* vectors combined ("all"), plus a scrambled, non-gene-targeting ("Scr") shRNA. *Miwi2*-knockdown (KD) cells were transduced in triplicate and Scr control cells in duplicate; individual clones with stable genomic integration of the hairpins were isolated via antibiotic selection (see Section 2.4.5, Materials and Methods).

Despite uniform puromycin resistance, ES cells within an individual colony displayed mosaic-like GFP expression in all 18 transduced lines (Figure 3-13A). It could thus be inferred that expression of the shRNA is also mosaic within a puromycin-resistant colony. To collect cells that I was confident were expressing the shRNA, I induced hairpin expression with doxycycline for 72h and performed flow cytometry on each of the 18 samples to i) remove contaminating MEFs and ii) collect only the living (7AAD-negative, Figure 3-13B, y-axis) and GFP-positive (Figure 3-13B, x-axis) cells for *Miwi2* expression analysis. Of the 16 *Miwi2*-KD samples, 14 yielded high quality RNA, hence were analysed for *Miwi2* expression. Of these 14 samples, only three cell lines (R1-pRSIT_4c, R1-pRSIT_5c and R1-pRSIT_all) exhibited a *Miwi2* knockdown of greater than 80% relative to the two scrambled control lines (Figure 3-13C and Figure 8-5, Appendix). Five ES cell lines (two control and three *Miwi2*-knockdown) were used in downstream expression analyses.

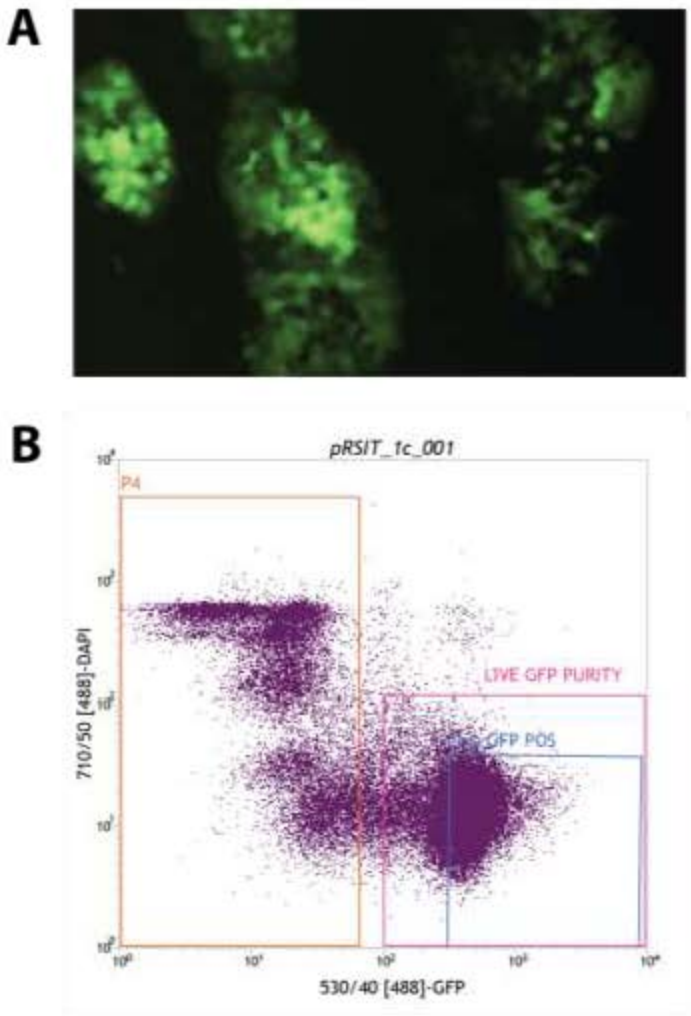

A
B
pRSIT_1c_001
P4
LIVE GFP PURITY
GFP POS
710/50 [488]-DAPI
530/40 [488]-GFP
C
***
Miwi2 expression relative to Scr_A
1.0
0.75
0.5
0.25
0.0
R1-pRSIT-Scr_A
R1-pRSIT-Scr_B
R1-pRSIT-Miwi2_4c
R1-pRSIT-Miwi2_5c
R1-pRSIT-Miwi2_all

Figure 3-13: Quantitative PCR detection of *Miwi2* transcript in GFP$^+$ R1 mouse embryonic stem cells after induced shRNA expression

(A) Fluorescent microscope image of transduced ES cells, 200x zoom. (B) Scatter plot defining FACS gating strategy for collection of GFP-positive R1 ES cells. (C) Quantitative RT-PCR showing *Miwi2* expression in control (R1-pRSIT-Scr) and *Miwi2*-knockdown cell lines (R1-pRSIT-Miwi2). Values are the average of technical triplicate PCR reactions; error bars represent standard error of the mean. P-value was determined via student's t-test (two-tailed, unpaired) analysis of two R1-pRSIT-Scr cell lines versus three R1-pRSIT-Miwi2 lines (*** $p < 0.001$).

Effects of Miwi2 knockdown in mouse embryonic stem cells: crude phenotype

I performed phenotypic observation of the two cell lines with greatest Miwi2 knockdown (R1-pRSIT_4c and R1-pRSIT_5c) and two Scr control lines (R1-pRSIT-Scr_A and R1-pRSIT_Scr_B) over 10 days following doxycycline induction of shRNA expression. Firstly, I performed daily crude visual observation of cell appearance. Pluripotent mouse ES cells form tight colonies *in vitro* with bright, round edges and have a syncytial appearance; differentiated cells display jagged, dull edges and individual cells within the colony are easily discerned. Secondly, cell density was measured every 48h after equal replating as a measure of doubling time. Finally, mean size was calculated of cells between 10-20µm every 48h. Placement of such cutoffs ensured that the majority of MEFs - which tend to be greater than 20µm in size - did not contribute to the size analysis. Figure 3-14 depicts these conflated data from the two *Miwi2*-knockdown cell lines and scrambled control duplicate lines.

Should *Miwi2* knockdown affect ES cell self-renewal or induce early differentiation undetectable by crude observation, it would be expected that the cells divide at a slower rate, or that the cell size would increase. It is not expected that most differentiated cells would exceed 20µm in size. However, there were no appreciable or significant differences in appearance, growth rate or cell size of *Miwi2*-KD cells relative to the Scr controls (Figure 3-14).

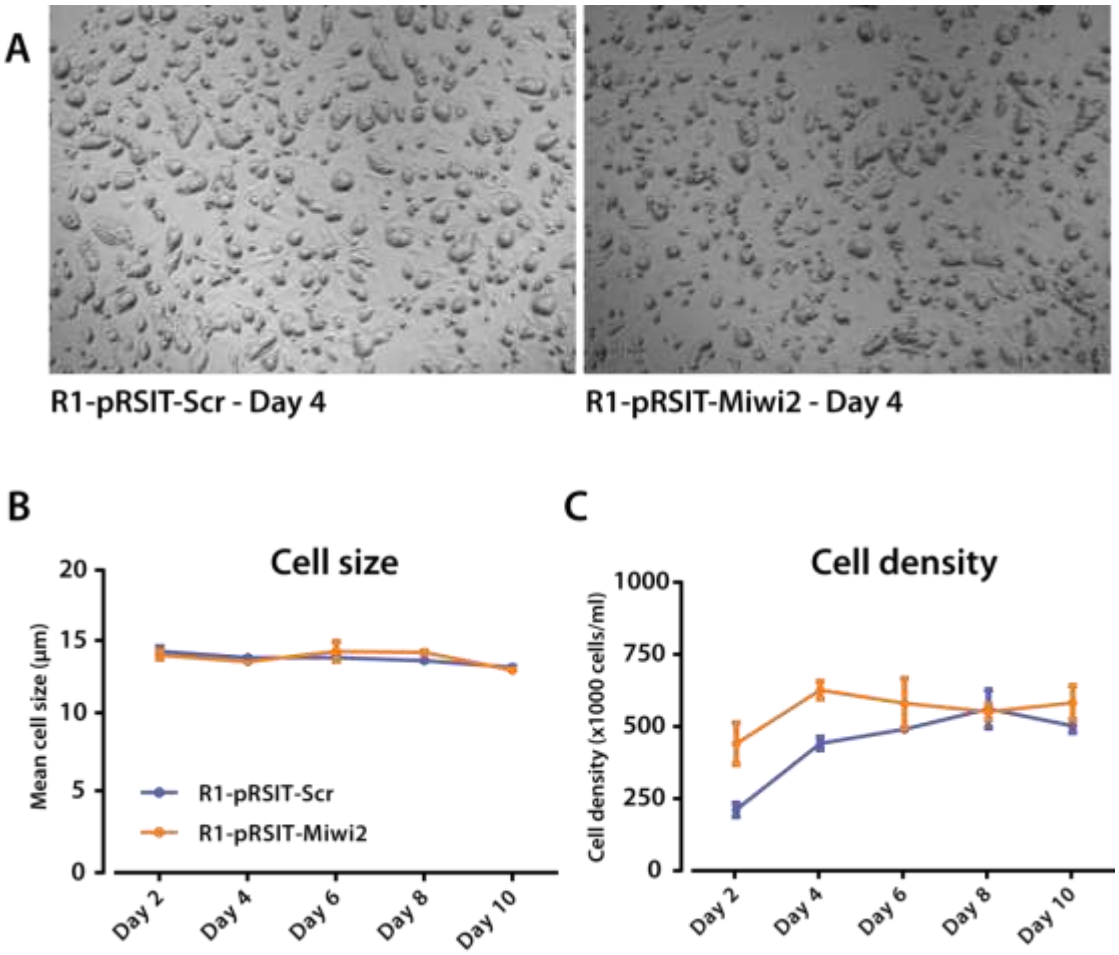


Figure 3-14: Phenotypic observation of *Miwi2* knockdown embryonic stem cells

(A) Images were captured following 96h doxycycline treatment of control cells (R1-pRSIT-Scr, n=2) and *Miwi2*-KD cells (R1-pRSIT-Miwi2, n=2) at 40x zoom. (B) Mean cell size. (C) Cell density. Data is averaged from duplicate cell lines per condition.

Effects of Miwi2 knockdown in mouse embryonic stem cells: pluripotency and early differentiation gene expression

Miwi2 knockdown in mES cells resulted in neither an overt effect on growth rate or cell size, nor obvious induction of differentiation. However, it is possible that it induced a more subtle phenotype, so I investigated whether any changes in gene expression of pluripotency genes had occurred. I analysed *Oct4, Sox2* and *Nanog* expression every 48h over 10 days doxycycline induction of hairpin expression (Figure 3-15). Although I confirmed that *Miwi2* expression remains knocked down in R1-pRSIT-Miwi2 samples relative to control R1-pRSIT-Scr samples, there are no consistent significant perturbations evident in expression of the pluripotency factors. Of note, there is an apparent increase in *Miwi2* in R1-pRSIT-Scr from Days 2-4. This was unexpected, as culture conditions are kept consistent for the duration of the 10 days.

I then asked if any aberrant peaks in lineage-specific gene expression were apparent following *Miwi2* knockdown that would indicate premature differentiation (Figure 3-16). I used the same lineage markers as in Figure 3-5. While expression of these genes is low in ES cells, four out of the six are quantifiable using qPCR (*HNF4α* and *Sox17* expression was below the limit of detection and is not shown). Where possible, the points are plotted using the same y-axis values as Figure 3-5 to indicate whether gene expression changes are indicative of premature differentiation. *Miwi2*-knockdown cells exhibit decreased expression of *Nestin, Otx2* and *Brachyury* relative to scrambled controls; this is evident at multiple time points over the 10 days. However, it should be noted that these genes – with the exception of CD31 – are expressed at a relatively low level in embryonic stem cells (see “Day 0 (ES)” gene expression, Figure 3-5).

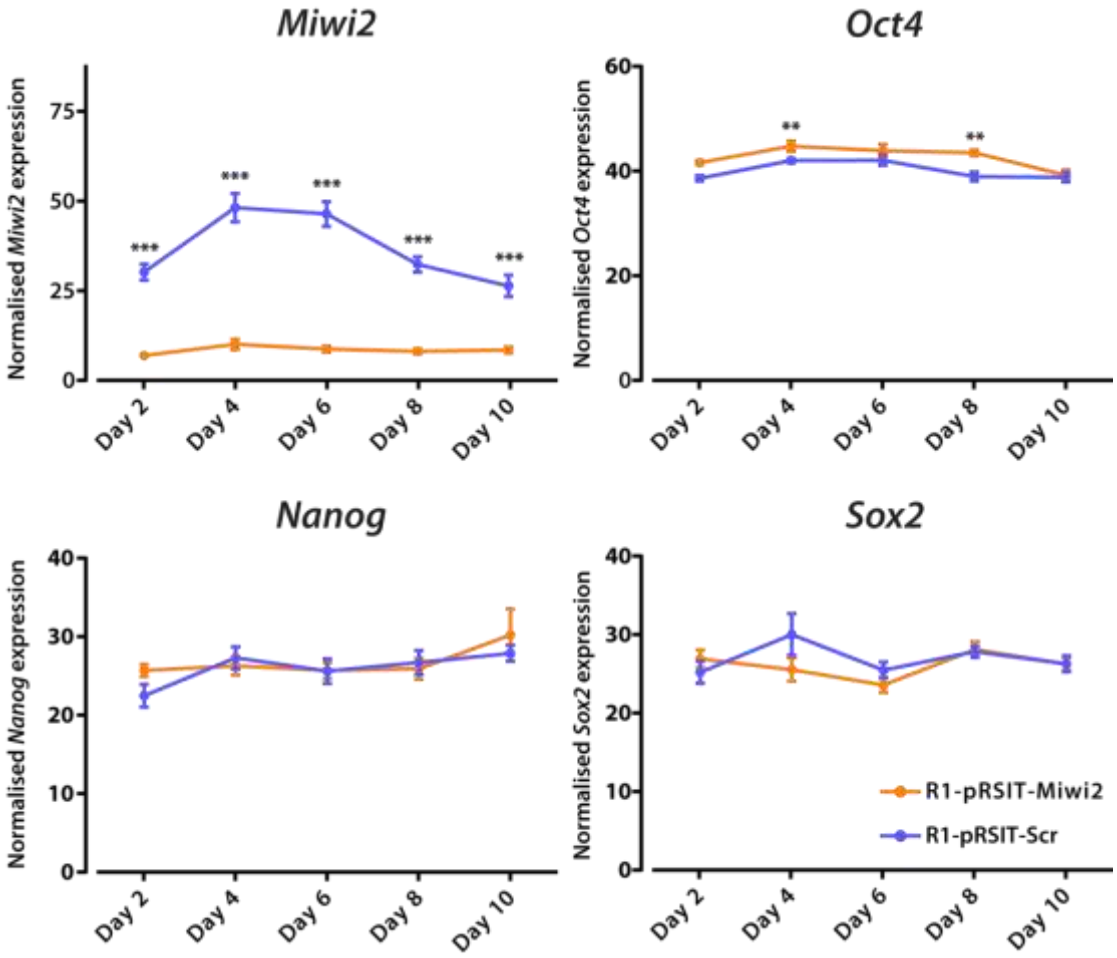


Figure 3-15: Quantitative PCR data showing *Miwi2* and pluripotency gene expression during doxycycline induction of *Miwi2* knockdown in mouse embryonic stem cells

Quantitative RT-PCR data representing transcript expression of *Miwi2, Oct4, Nanog* and *Sox2* during 10 days doxycycline-induced shRNA expression in scrambled control (blue line, n=2) versus *Miwi2* knockdown cells (orange line, n=2). Y-axis values are averaged between replicate cell lines and are normalised expression values derived from Comparative C_T analysis. Raw C_T values were normalised to *Canx* and *Ube2z* reference genes. Values are different on each axis as transcript abundance varies for each gene. Error bars represent standard error of the mean. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant differences in gene expression between scrambled control and *Miwi2* knockdown cell lines (** $p < 0.01$; *** $p < 0.001$).

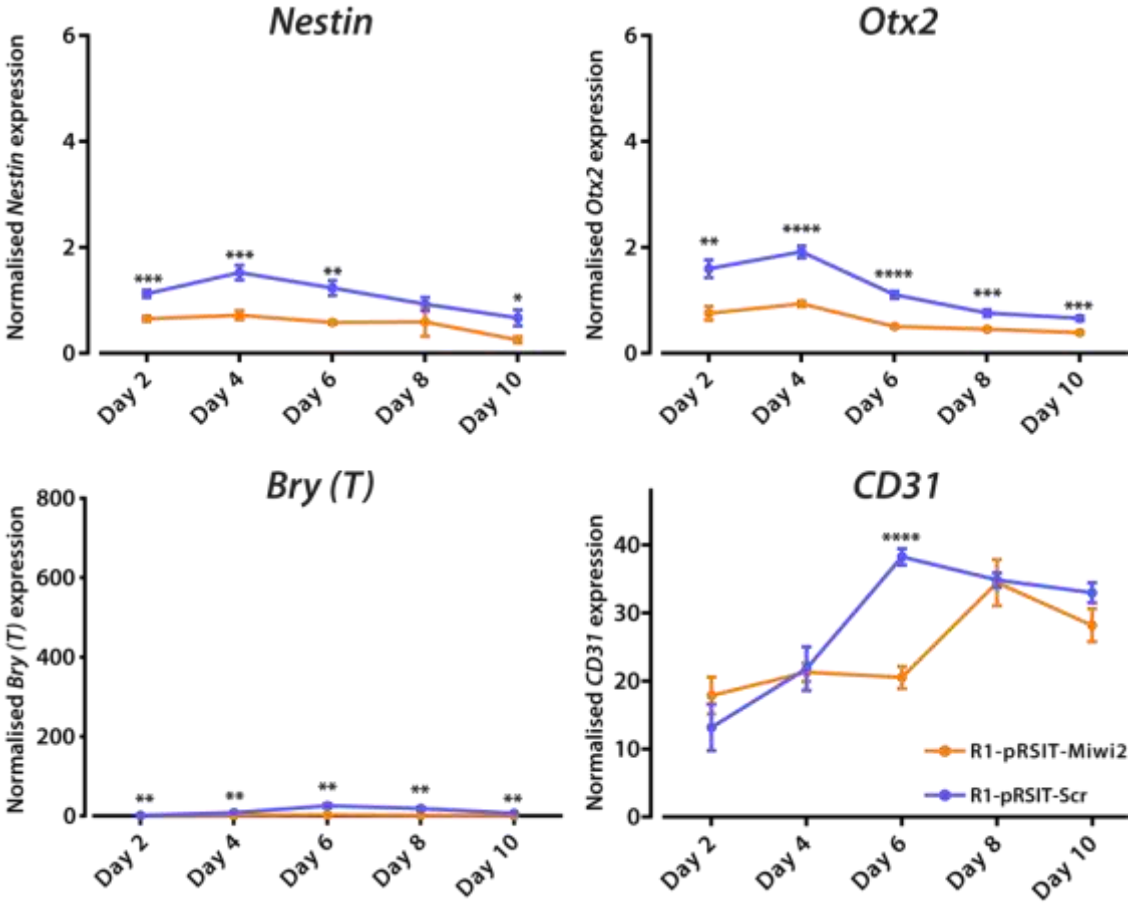


Figure 3-16: Quantitative PCR data showing primitive germ layer marker expression during doxycycline induction of *Miwi2* knockdown in mouse embryonic stem cells

Quantitative RT-PCR data representing transcript expression of ectodermal markers *Nestin* and *Otx2* and mesodermal markers *Brachyury* and *CD31* during 10 days doxycycline-induced shRNA expression in scrambled control (blue line, n=2) versus *Miwi2* knockdown cells (orange line, n=2). Y-axis values are averaged between replicate cell lines and are normalised expression values derived from Comparative C_T analysis. Raw C_T values were normalised to *Canx* and *Ube2z* reference genes. Values are different on each axis as transcript abundance varies for each gene. Error bars represent standard error of the mean. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant differences in gene expression between scrambled control and *Miwi2* knockdown cell lines (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).

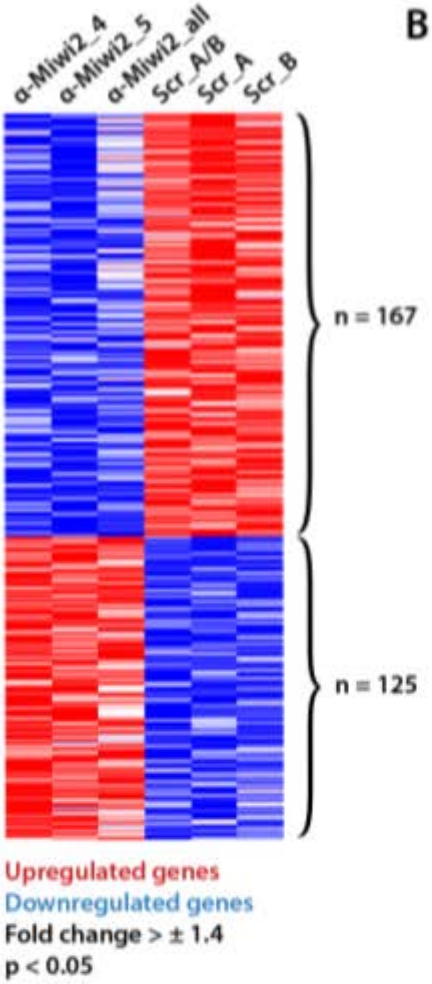
Effects of Miwi2 knockdown in mouse embryonic stem cells: global gene expression

While I saw no significant difference in the pluripotency markers with *Miwi2* knockdown over 10 days, there may have been some expression changes in other genes; I therefore decided to take a global view of any gene expression changes induced by knockdown of *Miwi2* in ES cells using Affymetrix® GeneChip® ST Arrays. This was a 3x3 experiment with single colour arrays: for scrambled control cells I had two biological replicates (R1-pRSIT-Scr_A and Scr_B) and one technical replicate (pooled Scr_A and Scr_B); I included three biological replicates for *Miwi2*-knockdown samples (R1-pRSIT-4c, -5c and -all).

Analysis was performed as described in Section 2.6.4, Materials and Methods. With a false discovery rate of 0.25 – which requires rigorous qPCR validation of any gene hits from the microarray analysis – a total of 167 genes were significantly down-regulated >1.4-fold and the expression of 125 genes was significantly increased >1.4-fold following *Miwi2* knockdown ($p < 0.05$, Figure 3-17A). Of note, *Otx2* is one of the more significantly down-regulated genes following *Miwi2* knockdown as revealed by the Affymetrix® array; this supports my previous findings (refer to Figure 3-16). However, this gene list included neither the pluripotency markers nor the remaining lineage markers examined above. Notably, there were no statistically significant changes in either *Mili* or *Miwi* following *Miwi2* knockdown. Genes (including noncoding RNAs) with the largest changes in expression from the microarray analysis are listed in Figure 3-17B. Of these top hits, those genes that are down-regulated had the greatest fold change in expression and were focused upon for further analysis. Putative functions of these genes are outlined in Table 8-2, Appendix.

To validate the array results, I performed SYBR Green qPCR analysis of expression of the 9 genes with greatest negative fold change (listed in Figure 3-17B) using the same RNA that was analysed on the microarray. In all cases, I confirmed highly significant ($p < 0.0001$ in 8/9 samples, $p < 0.001$ in 1/9 samples) down-regulation of these genes (Figure 3-17C). This provided confidence to perform pathway analysis using the dataset generated from the array analysis (Figure 3-17A). I used Ingenuity Pathway Analysis to generate a list of putative gene regulatory networks affected by *Miwi2* knockdown (Table 3-4). The top two predicted networks were i) cellular development, cellular growth and proliferation, haematological system development and function and ii) auditory and vestibular system development and function, embryonic development and organ development (Figure 3-18).

A



B

Top up-regulated	Fold change
miR-365	↑2.11
Nkx6-3	↑1.89
1600029D21Rik	↑1.83
Psg18	↑1.62
SLC17A6	↑1.51
DNAJC15	↑1.48
Mbnl2	↑1.46
Ank1	↑1.44
miR-491	↑1.43
PLA2G5	↑1.42

Top down-regulated	Fold change
Uty	↓4.46
Ddx3y	↓3.76
Calb2	↓2.60
Mmp13	↓2.20
miR-302c	↓2.19
Mmp3	↓2.13
Otx2	↓1.98
Enpp1	↓1.90
Mcpt8	↓1.87
Rasgrp1	↓1.75

C

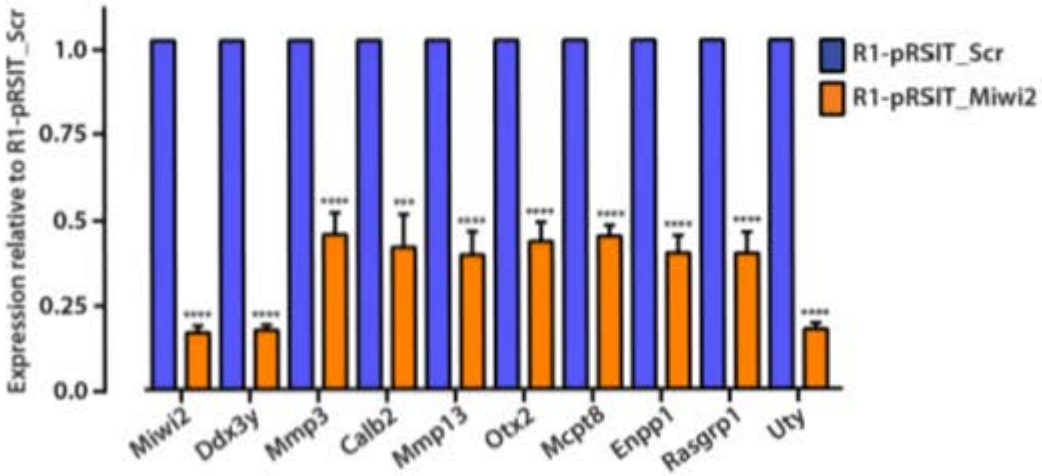


Figure 3-17: Affymetrix GeneChip® Mouse Gene 2.0 ST Array downstream analysis following *Miwi2* knockdown in mouse embryonic stem cells

(A) Heat map generated by GeneE (Broad Institute) for genes with largest significant expression changes (fold change > ±1.4, $p < 0.05$) as revealed by microarray analysis following *Miwi2* knockdown. (B) LimmaGP-generated list of genes with largest fold change ($p < 0.05$). (C) Quantitative PCR validation of genes in (B). Values are the average of gene expression in ES cell lines from each group: scrambled control values are derived from R1-pRSIT-Scr_A and _B (n=2); *Miwi2*-knockdown values from R1-pRSIT-Miwi2-4c, -5c and –all (n=3) and are presented relative to scrambled control. Error bars represent standard error of the mean. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant changes in gene expression in R1-pRSIT-Scr versus R1-pRSIT-Miwi2 samples (*** $p < 0.001$; **** $p < 0.0001$).

Table 3-4: List of top predicted gene regulatory networks affected by *Miwi2* knockdown

List of networks identified involving > 10 gene/molecules identified as differentially expressed after *Miwi2* knockdown by LimmaGP (fold change > ±1.4, $p < 0.05$).

Network #	# Genes/molecules	Top functions and associated diseases
1	24	Cellular Development, Cellular Growth and Proliferation, Haematological System Development and Function
2	16	Auditory and Vestibular System Development and Function, Embryonic Development, Organ Development
3	14	Cellular Development, Cellular Growth and Proliferation, Embryonic Development
4	14	Cellular Movement, Cancer, Inflammatory Response
5	13	Cellular Movement, Immune Cell Trafficking, Haematological System Development and Function
6	13	Cellular Movement, Haematological System Development and Function, Haematological Disease
7	12	Cellular Function and Maintenance, Cell Morphology, Cell-To-Cell Signalling and Interaction

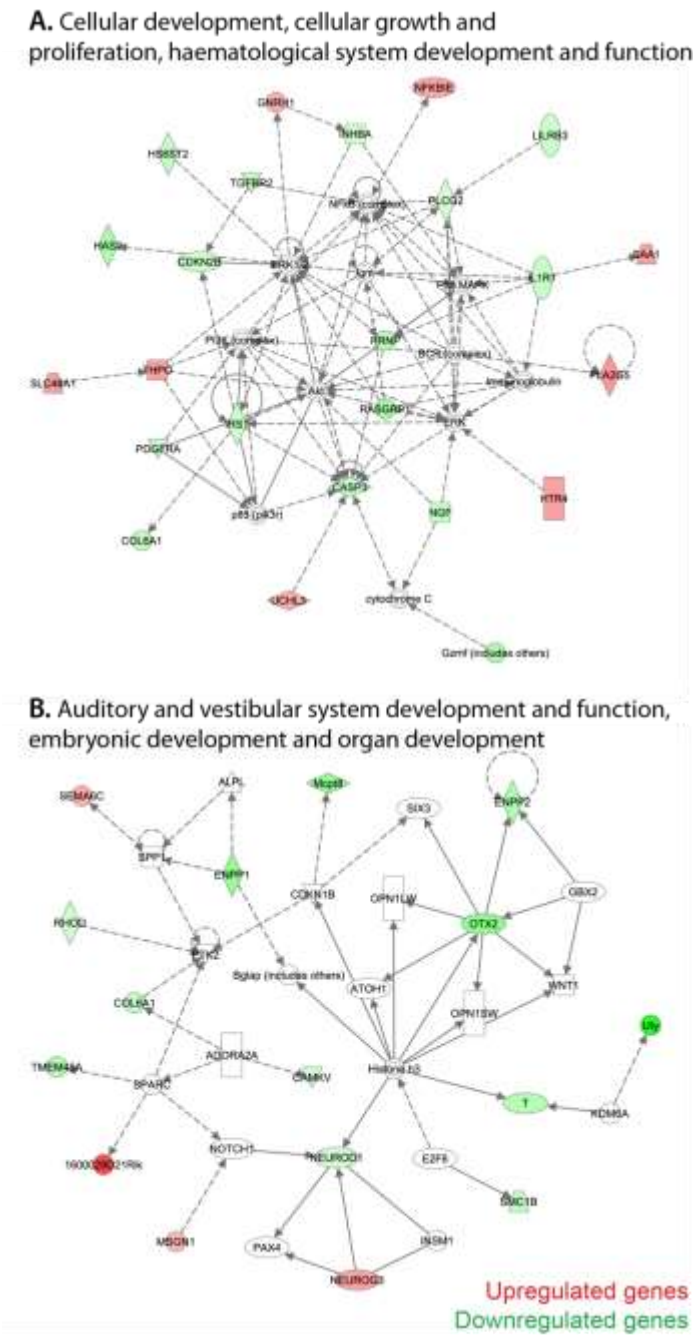


Figure 3-18: Ingenuity Pathway Analysis of genes with significant changes in gene expression following *Miwi2* knockdown

Top two predicted network interactions of differentially expressed genes from microarray data analysis (fold change > ±1.4, $p < 0.05$). Schematic generated using Ingenuity Pathway Analysis (Ingenuity Systems).

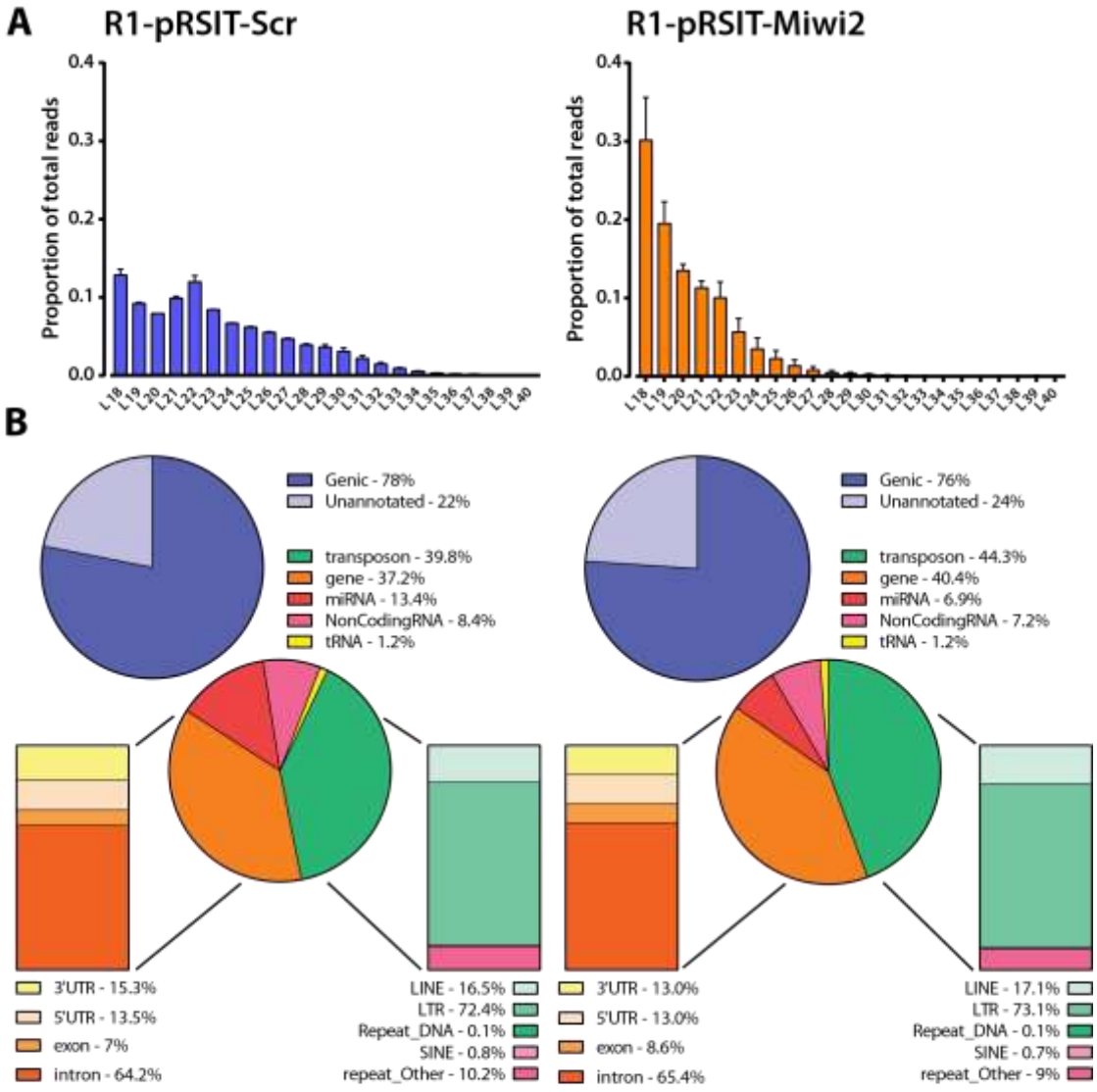
3.5 Small RNAome following *Miwi2* knockdown in mouse embryonic stem cells

Miwi2 knockdown in mouse ES cells generated perturbations in gene expression, and I hypothesised that some of these changes may be regulated by small noncoding RNAs. Hence, I performed next-generation sequencing of scrambled control and *Miwi2* knockdown ES cells to investigate changes in the total small RNA landscape. In particular, I wanted to address whether the expression of miRNAs and small RNAs with piRNA-like characteristics was affected by decreased *Miwi2* expression. I performed Illumina HiSeq® next-generation sequencing of total small RNA from the samples used for the microarray: duplicate scrambled controls (R1-pRSIT-Scr_A and –Scr_B) and triplicate *Miwi2*-knockdown ES cell lines (R1-pRSIT-Miwi2-4c, -Miwi2_5c and –Miwi2_all). Bioinformatic analysis was performed with Paul Young, Suter Laboratory. Following bioinformatic processing (see Section 2.11.2, Materials and Methods), I obtained an average of 5.8 million reads for scrambled controls and an average of 4.8 million reads for *Miwi2*-knockdown samples. Of those reads between 18 and 50nt in length, the majority mapped successfully to the mouse genome after filtering for 1 error (73% for R1-pRSIT-Scr and 77% for R1-pRSIT-Miwi2). I obtained 51,482 unique sequences for scrambled control samples and 24,616 from *Miwi2*-knockdown samples. My sequencing dataset included a number of single copy reads (24% for R1-pRSIT-Scr and 39% for R1-pRSIT-Miwi2), indicating again that the small RNA sequencing was not exhaustive. With the assistance of Paul Young (Suter Laboratory, VCCRI), I performed piRNA cluster analysis as described in Section 2.11.2, Materials and Methods.

Figure 3-19 depicts basic sequencing read analysis of scrambled control versus *Miwi2*-knockdown samples; all data is averaged from sequencing replicates. The small RNA profiles of R1-pRSIT-Scr and R1-pRSIT-Miwi2 samples were surprisingly dissimilar to that of Day 0 embryoid bodies at the ES cell stage (Figure 3-7). For example, scrambled control ES cells do not exhibit the expected miRNA-sized (21-24nt) peak (Figure 3-19A). To ensure that this size distribution was not indicative of degraded RNA, I estimated the quality of the RNA using the Agilent Bioanalyser; the RNA integrity number (RIN) was 9.4-10.0 for each sample (Figure 8-6, Appendix).

Following *Miwi2* knockdown, the proportion of piRNA-sized (24-32nt) reads decreases relative to control (Figure 3-19A); Miwi2 typically binds 27-29nt small RNAs (Aravin et al., 2008). Of those 27-29nt sequences in R1-pRSIT-Scr, 16% map to known piRNA cluster regions. Also, most 27-29nt sequences are derived from introns (19%), LTR repeats (23%) and unannotated regions (18%).

Almost one quarter of all small RNA reads map to unannotated regions and the majority of annotated reads are derived from genes and repeat regions for both groups. Most genic reads are derived from introns and most repeat-derived small RNAs map to LTR repeats (Figure 3-19B), predominately MaLR repeats (Figure 3-19C).



Repeat element	% all repeats
LTR_ERVL-MaLR	60.37%
LINE_L1	16.06%
rRNA_rRNA	9.56%
LTR_ERVK	6.80%
LTR_ERV1	3.47%

Repeat element	% all repeats
LTR_ERVL-MaLR	59.48%
LINE_L1	16.70%
rRNA_rRNA	8.23%
LTR_ERVK	7.46%
LTR_ERV1	4.52%

Figure 3-19: Characterisation of small RNA population in control and *Miwi2*-knockdown ES cells

Small RNA sequence characteristics of averaged R1-pRSIT-Scr duplicates versus averaged R1-pRSIT-Miwi2 triplicate samples. (A) Length distribution of small RNA reads between 18-40nt. Values represented as a proportion of total reads. (B) Categorical annotations of small RNA reads. (C) List of most abundant repeat elements giving rise to small RNA reads.

Small RNA expression changes after Miwi2 knockdown in ES cells

To investigate changes in small RNA expression, read counts were converted to a proportion of total reads/sample and linear fold change calculated in *Miwi2*-knockdown samples relative to scrambled controls. I filtered results for i) >100 reads in R1-pRSIT-Scr_A, ii) > ±2-fold change and iii) p-value <0.05 as generated by a two-tailed, unpaired Student's T-test. In R1-pRSIT-Miwi2 samples, 99 sequences were up-regulated and 386 sequences were down-regulated >2-fold ($p < 0.05$) relative to R1-pRSIT-Scr. Figure 3-20 outlines the length distribution and annotations of these up- and down-regulated sequences following *Miwi2* knockdown. There is an overt bias towards short sequence length for sequences with increased expression with majority of all reads 19-20nt in length. Those down-regulated sequences are mostly piRNA-sized (25-30nt) plus a small proportion of 22-24 and 31-34nt reads. 22% of up-regulated sequences and 16% of down-regulated small RNAs are generated from known piRNA clusters. Of those sequences derived from piRNA clusters that are down-regulated after *Miwi2* knockdown, the most common length is 27 nucleotides (20%), there is an underrepresentation of C at position 1 and an overrepresentation of C at position 10 (Figure 8-7, Appendix).

I adjusted the parameters for miRNA analysis to include sequences with at least 1.5-fold expression changes. I observed the down-regulation of six and the up-regulation of three miRNAs following *Miwi2* knockdown in ES cells (Table 3-5 and Table 3-6). Notably, this list does not include the ES cell-specific miR-302c which – in addition to exhibiting decreased expression during normal *in vitro* differentiation (Table 3-1) – was reported to be down-regulated in R1-pRSIT-Miwi2 cells by microarray (Figure 3-17). Members of the miR-199 and miR-669 family – which undergo large expression changes during *in vitro* differentiation – exhibit smaller decreases following *Miwi2* knockdown.

Most of the small RNAs with greater than 2-fold changes in expression in R1-pRSIT-Miwi2 cells relative to R1-pRSIT-Scr originate from repeat sequences (LTR and LINE repeats, Figure 3-20). Sequencing analysis revealed that the majority of LTR repeats are MaLR: 77% of up-regulated and 91% of down-regulated LTR reads. Many up-regulated (38%) and down-regulated (35%) small RNAs also map to genes: 3'UTR, 5'UTR and introns. Several genes produce small RNAs with large expression changes, representative snapshots from UCSC Genome Browser are shown in Figure 8-8, Appendix.

Some small RNAs map to both introns and repeat elements; these are listed in Table 3-8. In addition, a small proportion of down-regulated sequences align with tRNA genes (Table 3-7); no tRNA-derived small RNAs were up-regulated following *Miwi2* knockdown (Figure 3-20).

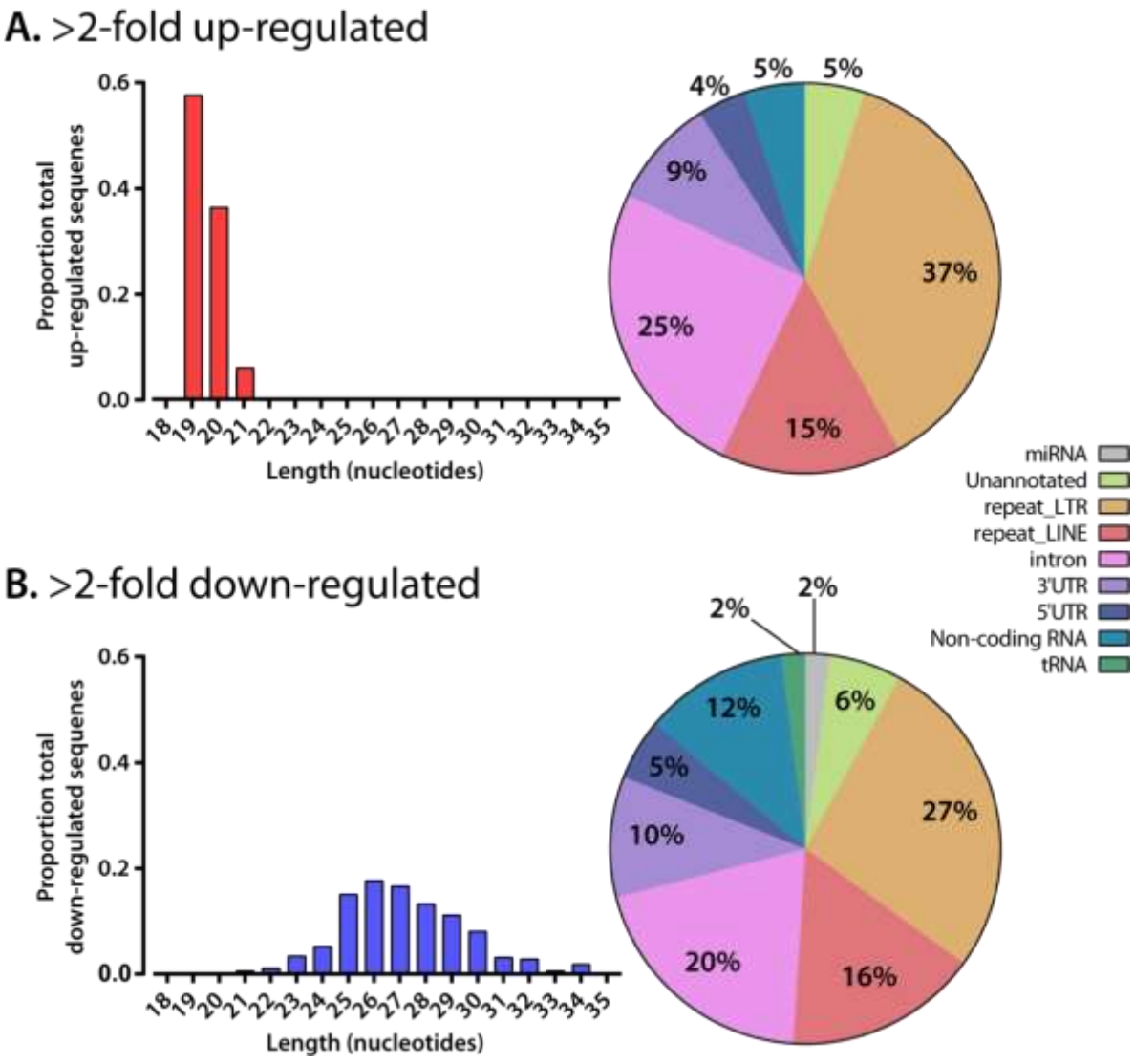


Figure 3-20: Small RNA expression changes following *Miwi2* knockdown

Length distribution (left hand side) and categorical annotations (right-hand side) of (A) >2-fold up-regulated and (B) >2-fold down-regulated small RNA reads presented as a proportion of total up- or down-regulated reads.

Table 3-5: List of >1.5-fold down-regulated miRNAs in *Miwi2* knockdown ES cells

miRNA ID	Read count Scr	Read count Miwi2	Fold Change
miR-199a-3p	996	91	1.94
miR-199a-5p	1646	153	1.85
miR-295-3p	3483	309	2.07
miR-409-5p	171	15	2.39
miR-541-5p	2661	259	1.62
miR-669c-5p	280	27	1.99

Table 3-6: List of >1.5-fold up-regulated miRNAs in *Miwi2* knockdown ES cells

miRNA ID	Read count Scr	Read count Miwi2	Fold Change
miR-376b-3p	131	42	2.14
miR-381-3p	868	265	1.88
miR-411-5p	386	113	1.92

Table 3-7: List of tRNA-derived small RNAs with >2-fold decrease in *Miwi2* knockdown ES cells

Length	Annotation(s)	tRNA	Fold Change
24	tRNA	HisGTG	-2.5244493
26	tRNA|piRNA	GluCTC	-2.4552296
27	tRNA	GluCTC	-4.1177555
30	tRNA	GlyGCC	-9.0782782
30	tRNA	GluCTC	-8.2913869
31	tRNA	GlyGCC	-4.7637759
31	tRNA	GluCTC	-9.903263
32	tRNA	LysCTT	-7.7276989

Table 3-8: List of gene with introns giving rise to small RNAs that also map to repeat elements with >2-fold decreases after *Miwi2* knockdown in ES cells

Gene name	Type of repeat element
Arhgap21	repeat_LINE_L1
Camk1d	repeat_LTR_ERV1
Cdk8	repeat_LTR_ERVK
Gphn	repeat_LTR_ERVL-MaLR
Gphn	repeat_LINE_L1
Impg1	repeat_LINE_L1
Iqcg	repeat_LTR_ERVL-MaLR
Itga9	repeat_LINE_L1|piRNA_Cluster_AS
Spata5	repeat_LINE_L1
Tmtc2	repeat_LINE_L1
Zfp185	repeat_LTR_ERVL-MaLR

To summarise the findings of the small RNA profiles in mouse ES cells following *Miwi2* knockdown, I observed a decrease in abundance of pi-sized (mostly 24-30nt) RNAs – many of which map to known piRNA clusters – and an increase in smaller (mostly 19-20nt) RNAs, although the identity and function of many of these sequences is yet to be revealed. Similar to the small RNA profiles during *in vitro* differentiation, I observed positive and negative expression changes in a number of small RNAs mapping to MaLR repeats and introns. Finally, there were small, but subtle, changes in expression of several miRNAs in *Miwi2*-knockdown ES cells.

3.6 Effects of Miwi2 knockdown in mouse embryonic stem cells: gene expression changes during *in vitro* differentiation

I observed subtle but significant differences in low-level lineage marker expression in *Miwi2*-knockdown embryonic stem cells maintained under pluripotent conditions for 10 days (Figure 3-16). I then examined whether aberrant changes in expression of these pluripotency and differentiation markers were evident during *in vitro* differentiation of *Miwi2*-knockdown ES cells. For the differentiation experiment, I tested seven cell lines: the same duplicate scrambled control lines (R1-pRSIT-Scr_A and _B) and same two *Miwi2*-knockdown cell lines expressing hairpins #4 and #5 (R1-pRSIT-Miwi2-4c and -5c) as in Section 3.4; an additional two lines expressing α-*Miwi2* hairpins #1 and #3 (R1-pRSIT-Miwi2-1a and -3a; plus a technical repeat of an ES cell line expressing hairpin #5 (R1-pRSIT-Miwi2-5c_rpt). I included the additional two *Miwi2*-knockdown cell lines (hairpins #1 and #3) as these shRNA sequences yielded a milder (yet still greater than 50%) knockdown of the transcript than using hairpins #4 and #5, hence I would be able to assess any dose-dependent effects on *Miwi2* knockdown on gene expression during embryoid body formation. Therefore, I investigated the differentiation potential of five *Miwi2*-knockdown cell lines expressing four unique α-*Miwi2* shRNAs relative to two scrambled control ES cell lines. I took this approach as it is important to direct multiple shRNAs towards the target gene in knockdown experiments to account for any off-target effects of the hairpin (Jackson and Linsley, 2010). All cell lines were treated with doxycycline to induce expression of the shRNA for 48h prior to *in vitro* differentiation (see Section 2.4.4, Materials and Methods). Figure 3-21 depicts average gene expression changes between both scrambled controls versus all *Miwi2*-knockdown cell lines during 6 days differentiation: *Miwi2, Oct4, Nanog* and *Sox2*. Data from R1-pRSIT-Miwi2-5c_rpt has been omitted from Figure 3-21 as it demonstrated signs of off-target effects. Graphs showing pluripotency gene expression for the seven individual cell lines are depicted in Figure 8-9, Appendix. I confirmed that *Miwi2* expression was significantly knocked down 48h after doxycycline induction of shRNA (Day 0, Figure 3-21A), although hairpins #4 and #5 (Miwi2_4c and Miwi2_5c) induced the most efficient knockdown, as previously observed in the original knockdown experiments (described above, Section 0). *Miwi2* expression remains significantly repressed in all the R1-pRSIT-Miwi2 cells relative to R1-pRSIT-Scr samples until Day 3 when *Miwi2* expression converges at a low level in both control and *Miwi2*-knockdown cells (Figure 3-21B); this reflects the pattern of *Miwi2* expression during *in vitro* differentiation of wild-type ES cells (see Figure 3-4).

There are no significant differences in *Oct4* expression until Day 4, where R1-pRSIT-Miwi2 samples display reduced expression; *Oct4* levels converge at Day 5. I observed a premature peak in *Nanog* expression at Day 3 in *Miwi2*-knockdown cell lines, compared to a peak in expression seen at Day 4 in scrambled control cells; however a peak in *Nanog* expression at Day 3 is observed during WT differentiation (refer to Figure 3-4). There were no significant differences in levels of *Sox2* during this timecourse: R1-pRSIT-Miwi2 cell lines displayed almost identical expression patterns to R1-pRSIT-Scr cell lines.

I previously observed changes in low-level expression of several lineage markers in *Miwi2*-knockdown ES cells (Figure 3-16), therefore I also surveyed the regulation these markers during *in vitro* germ layer formation in the two scrambled control lines and the five *Miwi2*-knockdown ES cell lines described above (Figure 3-22). Data for R1-pRSIT-Miwi2-5c_rpt are again omitted as this cell line demonstrated signs of off-target effects. Graphs showing differentiation gene expression for individual cell lines are depicted in Figure 8-10, Appendix. All *Miwi2*-knockdown cell lines exhibit relatively similar patterns of gene expression of *Nestin, Brachyury, CD31, HNF4α* and *Sox17* to Scr controls. The only significant difference in expression is in *Sox17* at Day 6 (Figure 3-22C), although the expression pattern of R1-pRSIT-Miwi2 cells lines is similar to R1-pRSIT-Scr. The expression pattern of *Otx2* (Figure 3-22A) is variable between cell lines as indicated by the large error bars; these changes may be attributed to the ~2-fold decrease in low-level *Otx2* expression in R1-pRSIT-Miwi2_4 and Miwi2_5 (but not Miwi2_1 and Miwi2_3) relative to R1-pRSIT-Scr; this was revealed by the microarray analysis (refer to Figure 3-17) and confirmed via qPCR (Figure 3-22D).

In summary, following *in vitro* differentiation of *Miwi2*-knockdown ES cells, I observed a premature peak in *Nanog* at Day 3 and a premature decline in *Oct4* expression at Day 4, but no overt change was observed in *Sox2*. Furthermore, although *Miwi2* expression was significantly knocked down in R1-pRSIT-Miwi2 cells relative to scrambled controls, R1-pRSIT-Miwi2 cell lines did not exhibit significant changes in expression dynamics of select differentiation genes. Overall, pluripotency and lineage marker gene expression was largely unaffected following *Miwi2* knockdown in mouse ES cells and cells appear to differentiate normally.

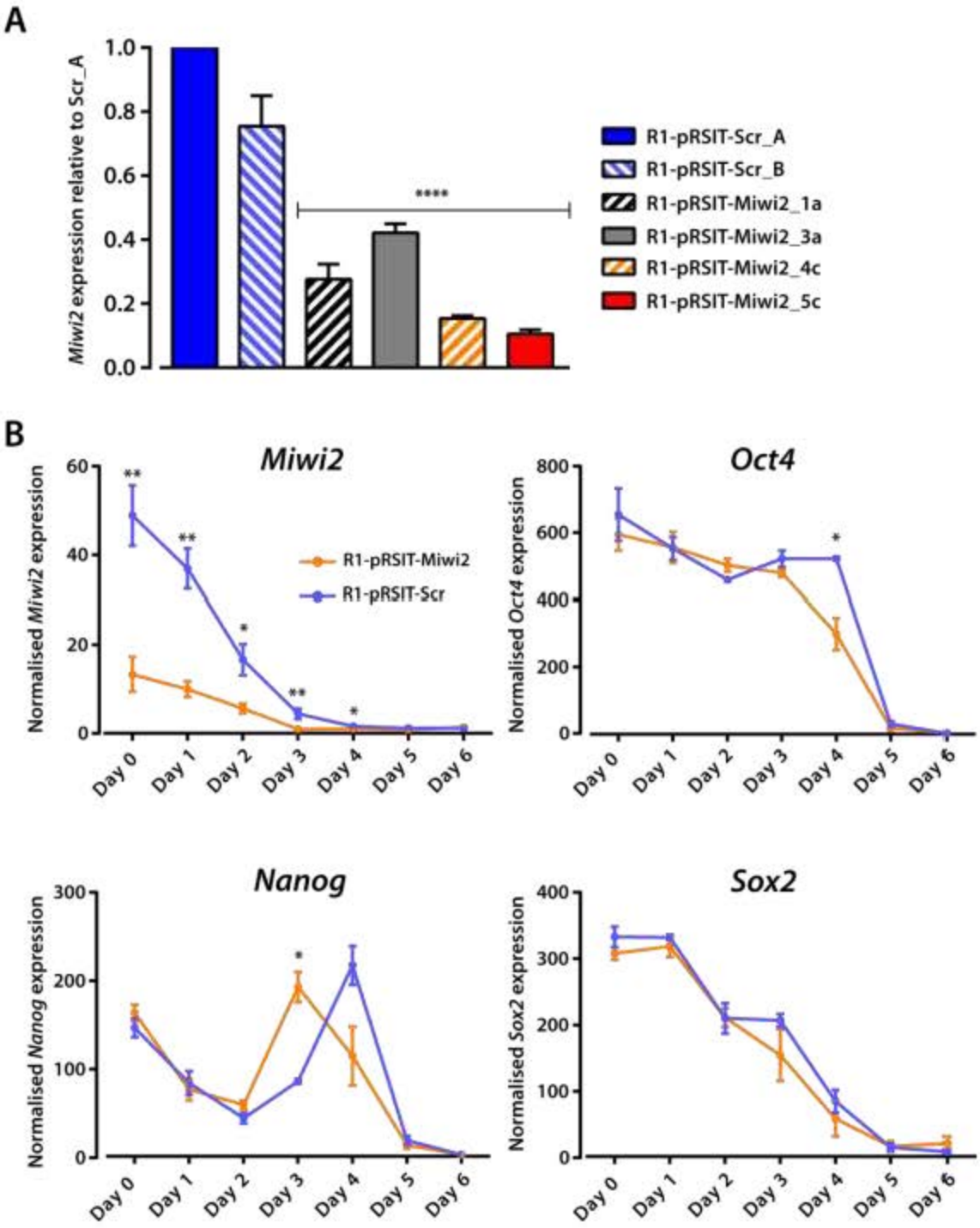

A
Miwi2 expression relative to Scr_A
1.0
0.8
0.6
0.4
0.2
0.0
****
R1-pRSIT-Scr_A
R1-pRSIT-Scr_B
R1-pRSIT-Miwi2_1a
R1-pRSIT-Miwi2_3a
R1-pRSIT-Miwi2_4c
R1-pRSIT-Miwi2_5c
B
Miwi2
Normalised Miwi2 expression
R1-pRSIT-Miwi2
R1-pRSIT-Scr
Oct4
Normalised Oct4 expression
Nanog
Normalised Nanog expression
Sox2
Normalised Sox2 expression
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6

Figure 3-21: *Miwi2* and pluripotency gene transcript expression during *in vitro* differentiation of *Miwi2*-knockdown mouse embryonic stem cells

Quantitative RT-PCR data for gene expression in scrambled control versus *Miwi2* knockdown cells. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant differences in gene expression between scrambled control and *Miwi2* knockdown cell lines (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). (A) Baseline *Miwi2* expression at Day 0, 48h after doxycycline induction of shRNA. Y-axis values are the average of triplicate PCR reactions and are presented relative to R1-pRSIT-Scr_A. Error bars represent standard error of the mean of replicate PCR reactions. (B) *Miwi2, Oct4, Nanog* and *Sox2* expression over 6 days *in vitro* differentiation. Y-axis values are averaged between scrambled control (blue line, n=2) versus *Miwi2* knockdown cells (orange line, n=4) and are normalised expression values derived from Comparative C_T analysis. Raw C_T values were normalised to *Canx* and *Ube2z* reference genes. Values are different on each axis as transcript abundance varies for each gene. Error bars represent standard error of the mean of replicate cell line expression.

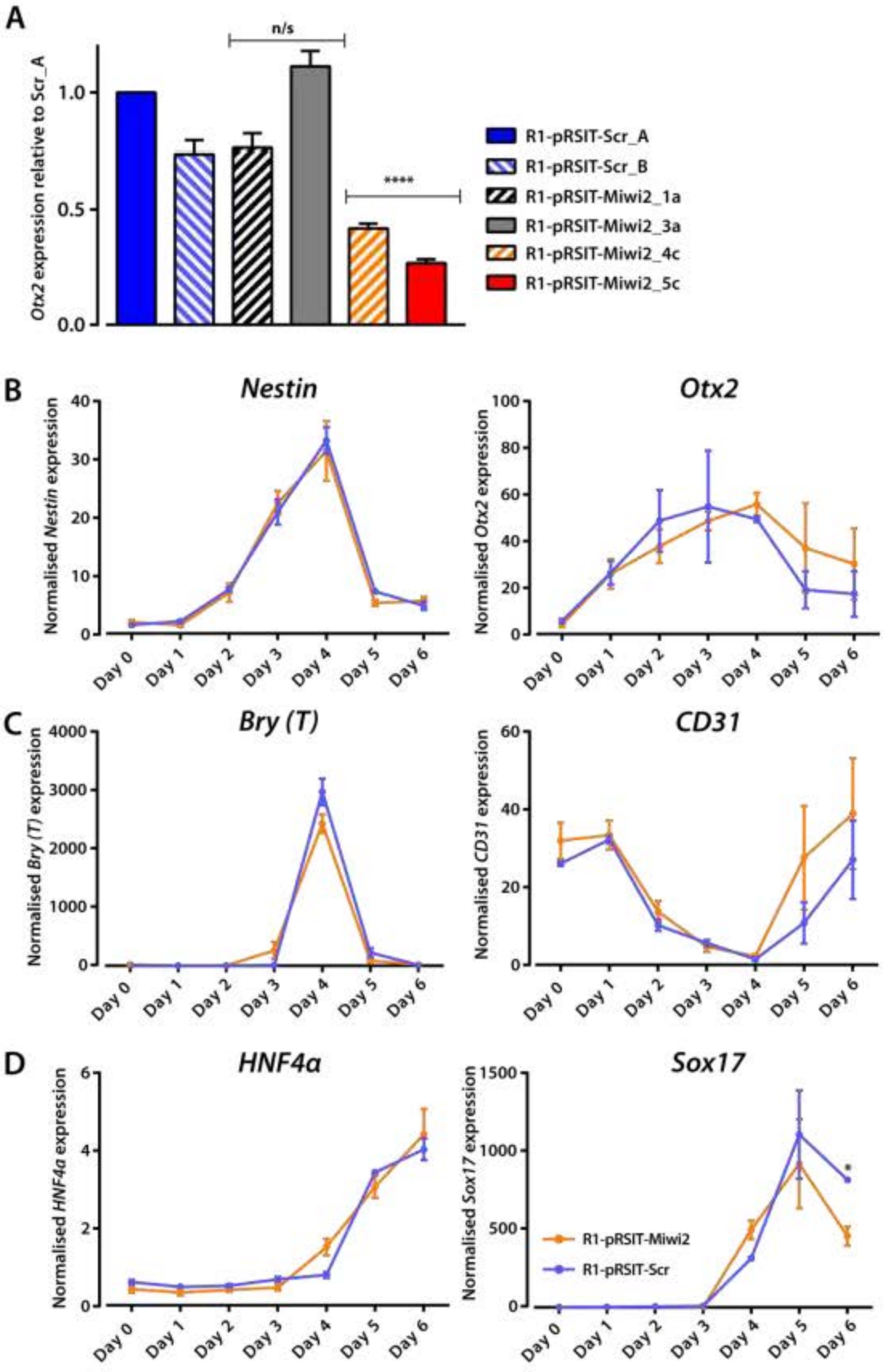

A
n/s
****
Otx2 expression relative to Scr_A
1.0
0.5
0.0
R1-pRSIT-Scr_A
R1-pRSIT-Scr_B
R1-pRSIT-Miwi2_1a
R1-pRSIT-Miwi2_3a
R1-pRSIT-Miwi2_4c
R1-pRSIT-Miwi2_5c
B
Nestin
Normalised Nestin expression
Otx2
Normalised Otx2 expression
C
Bry (T)
Normalised Bry (T) expression
CD31
Normalised CD31 expression
D
HNF4α
Normalised HNF4a expression
Sox17
Normalised Sox17 expression
R1-pRSIT-Miwi2
R1-pRSIT-Scr
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6

Figure 3-22: Primitive germ layer marker transcript expression during *in vitro* differentiation of *Miwi2*-knockdown mouse embryonic stem cells

Quantitative RT-PCR data for gene expression during in scrambled control versus *Miwi2* knockdown cells. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant differences in gene expression between scrambled control and *Miwi2* knockdown cell lines (* $p < 0.05$, **** $p<0.0001$). (A) Baseline *Otx2* expression at Day 0, 48h after induction of shRNA by doxycycline. Y-axis values are the average of triplicate PCR reactions and are presented relative to R1-pRSIT-Scr_A. Error bars represent standard error of the mean of replicate PCR reactions. (B-D) Differentiation gene expression over 6 days *in vitro* differentiation. Y-axis values are averaged between scrambled control (blue line, n=2) versus *Miwi2* knockdown cells (orange line, n=4) and are normalised expression values derived from Comparative C_T analysis. Raw C_T values were normalised to *Canx* and *Ube2z* reference genes. Values are different on each axis as transcript abundance varies for each gene. Error bars represent standard error of the mean of replicate cell line expression. (B) Ectodermal markers *Nestin* and *Otx2*. (C) Mesodermal markers *Brachyury (Bry/T)* and *CD31 (PECAM)*. (D) Endodermal markers *HNF4α* and *Sox17*.

Discussion

In this chapter, I examined murine Piwi expression in somatic tissues and embryonic stem cells. While my data revealed low-level to negligible expression of *Miwi* and *Mili* in a range of somatic tissues and ES cells, *Miwi2* transcript is expressed ubiquitously in somatic tissues, and at high levels in mouse ES cell lines; it is also translated into protein in somatic tissues and cell lines. Knockdown of *Miwi2* did not result in an explicit effect on ES cell self-renewal and pluripotency, but did induce subtle changes in gene expression as well as a range of small RNA expression changes. These changes in gene and small RNA expression did not, however, overtly affect the *in vitro* differentiation potential of these ES cells following reduction of *Miwi2* function, at least over a six day time course. The significance of these molecular changes and the precise role of Miwi2 in stem cell biology are, thus, still undetermined.

One murine Piwi gene, Miwi2*, is widely expressed in somatic tissues*

In contrast to previous reports that *Miwi2* is only expressed in a brief developmental window in developing germ cells (Aravin and Bourc'his, 2008), I have demonstrated that *Miwi2* expression in testis persists into adulthood (Figure 3-1A and Figure 8-1, Appendix); this is likely representative of the presence of GSCs that enable ongoing spermatogenesis throughout the lifetime of a mouse. According to ENCODE data, *Miwi2* transcript abundance is low but widespread in somatic tissues (Figure 3-2). Low-level expression may reflect the steady state in most cells, but another possibility is that expression of *Miwi2* may be high only in a specific subset of cells, such as stem-like cells. This idea is supported by Sharma and colleagues, who reported increased expression of a human Piwi gene, *Hiwi,* in $CD34^+$ haematopoietic stem cells (HSCs); *Hiwi* expression decreases concurrently with differentiation of HSCs into specific blood cell lineages (Sharma et al., 2001). Studies in primitive organisms also implicate Piwi family members in maintaining stem-like features, where expression of Piwi orthologues is increased in (or necessary for) adult somatic cells during tissue regeneration (Reddien et al., 2005, Rinkevich et al., 2010, Zhu et al., 2012). Consistent with the findings of Sharma et al, Haifan Lin's group recently reported an enrichment of *Miwi2* expression in HSCs and multipotent progenitors of the mouse bone marrow relative to their differentiated counterparts (Nolde et al., 2013). However, the *Miwi2*-deficient HSCs appear to function normally in response to myeloablative injury (Nolde et al., 2013) and even knockout of all three murine Piwi proteins showed normal haematopoietic development. It is possible that Miwi2 is not functional in somatic stem cells, but is merely a remnant from the pluripotent stem cell

stage. This is certainly the case for the pluripotent stem cell marker, Oct4, which is expressed in specific cells within somatic compartments (Lengner et al., 2007). The hypothesis that Oct4 is also involved in maintaining potency of multipotent adult stem cells was overturned as tissue-specific knockout mouse models for Oct4 show normal tissue function in terms of homeostasis and regenerative capacity. Whether Piwi proteins are functional in stem cell compartments other than the bone marrow remains unstudied.

Miwi2 is highly expressed in mouse embryonic stem cells and decreases upon differentiation

The observation that *Miwi2* is enriched in highly potent ES cells selected for the pluripotency marker SSEA1 compared to bulk ES cells is unprecedented (Figure 3-1B). This observation could suggest that – like SSEA1 – *Miwi2* expression is cell cycle-dependent. Alternatively, pluripotent ES cell populations could be heterogeneous. This is supported by reports that *in vitro* mouse ES cell cultures contain a rare, transient population similar to cells from two-cell (2C) embryos; all ES cells in a population possess the ability to cycle in and out of this 2C state (Macfarlan et al., 2012). 2C cells lack Oct4, Sox2 and Nanog and are totipotent: they are able to generate both embryonic and extra-embryonic tissues. Given that *Miwi2* expression decreases with differentiation, it is conceivable that *Miwi2* expression could be enriched in these totipotent 2C-like cells; investigation of transcript expression in isolated 2C-like versus pluripotent ES cells could address this possibility.

Miwi2 expression in ES cells declines in concert with their *in vitro* differentiation and lineage specification; these observations are consistent with Miwi2 playing some role in stem cell maintenance. Miwi2 protein is localised predominately in the cytoplasm in mouse embryonic stem cells (Figure 3-3). In the murine germline, Miwi2 is expressed in the nucleus of male germ cells, except in the absence of Mili (Aravin and Bourc'his, 2008) and I find that *Mili* is virtually absent in mouse embryonic stem cells. Klenov et al mutated the nuclear localisation signal of *Drosophila* Piwi, restricting the fully functional protein to the cytoplasm of female germ cells (Klenov et al., 2011). As expected, the transposon-suppression function of Piwi was removed, however germline stem cells continued to self-renew and differentiate normally. This observation demonstrates that Piwi has multiple functions that are dependent on cellular localisation. It is also consistent with the idea that cytoplasmic Miwi2 in mouse embryonic stem cells could be involved in regulating pluripotency.

Although it seems likely that Miwi2 must be expressed in the nucleus – as it is in germ cells – to silence repeat elements at the level of transcription, I have not ruled out a role for cytoplasmic Miwi2 in an alternate transposon silencing pathway in ES cells. As an example, in the germ cells of the nematode *C. elegans*, the Piwi ortholog PRG-1 is expressed in the cytoplasm in P-granules and complexes with 21nt piRNAs with a 5' U (Batista et al., 2008). Unlike in mammalian germ cells where the slicer domain of Piwi proteins is implicated in cleavage of transposable elements (Aravin et al., 2008), PRG-1 does not cleave its targets in *C. elegans* (Bagijn et al., 2012). Rather, the piRNAs silence endogenous transcripts – including the DNA transposon Tc3 – by triggering a downstream endogenous small interfering RNA (endo-siRNA) response in the cytoplasm (Das et al., 2008, Bagijn et al., 2012). It is possible that cytoplasmic Miwi2 could also induce a silencing cascade in ES cells by a similar mechanism, as endo-siRNAs have been shown to exist in mice, including ES cells (Babiarz et al., 2008), oocytes (Watanabe et al., 2008, Tam et al., 2008) and male germ cells (Song et al., 2011). Identification of the piRNA partners of Miwi2 will go some way to addressing this hypothesis; this is addressed in Chapter 5 of this thesis.

Miwi2 knockdown does not overtly affect maintenance of the stem cell state but induces perturbations in ESC gene expression

Given that *Miwi2* expression naturally declines during *in vitro* differentiation of ES cells, I investigated whether the induction of differentiation could be recapitulated by knockdown of Miwi2 using lentiviral-delivered shRNA. In doing this, I found that only 2 of the 5 α-*Miwi2* hairpins are capable of knocking down the transcript by more than 80% when transduced individually (Figure 3-13C) and it is likely that the efficient knockdown achieved by the single viral preparation with all 5 shRNAs packaged together (R1-pRSIT-Miwi2_all) was mediated predominately by these two hairpin sequences. This highlights the importance of using multiple shRNAs to maximise knockdown efficiency; ultimately, the effectiveness of a shRNA can only be defined empirically.

Lentivirally-transduced ES cells with decreased *Miwi2* appear normal, with no overt changes in cell size, appearance or doubling time (Figure 3-14). Moreover, there were no obvious changes in pluripotency gene expression, nor any aberrant expression of the typical lineage markers (Figure 3-15 and Figure 3-16). This, however, does not rule out a role of Miwi2 in maintaining stemness: even knockout of the master pluripotency factors, Sox2 and Nanog, does not induce overt differentiation of embryonic stem cells (Masui et al., 2007, Chambers et al., 2007). Microarray expression profiling in *Miwi2*-knockdown cells revealed gene expression changes that may provide

some insight into the role of *Miwi2* in ES cells. Pathway analysis of the genes altered by *Miwi2* knockdown (Figure 3-18) demonstrated perturbations in gene regulatory networks related to cellular development, growth and proliferation. These perturbations can be attributed specifically to knockdown of *Miwi2* as changes in R1-pRSIT-Miwi2 cells are directly compared to those cells expressing a lentivirally-transduced non-targeting hairpin (R1-pRSIT-Scr). Furthermore, the analysis was performed on cells targeted by multiple hairpins to rule out any off-target effects of the shRNAs (Jackson and Linsley, 2010).

Unlike immortalised cell lines – such as HEK293FT – ES cells can be difficult to transduce at high efficiency with a viral multiplicity of infection (MOI, or the number of infectious viral particles per cell) of less than 10 (Kosaka et al., 2004); I transduced mouse ES cells at a MOI of less than 10 to minimise the chance of multiple viral integrations per cell. Murine ES cells divide every 10-14 hours *in vitro* (Pauklin et al., 2011), hence, if Miwi2-depleted cells exhibited a growth defect – or even a complete exit from the pluripotent state – then the naïve population would quickly outcompete the transduced cells when transduction efficiency is low. To circumvent this issue I used a doxycycline-inducible system. Lentivirus mediates stable genomic integration of the vector, which encompasses both α-Miwi2 shRNA and puromycin resistance, therefore I was able to enrich for positive cells with antibiotic treatment. However, I observed mosaicism of GFP expression in puromycin-resistant populations (Figure 3-13A). There are several explanations for why this might occur. First, it is possible that some cells had silenced GFP expression but retained puromycin resistance. Second, given the three-dimensional nature of ES cell colony formation it is possible that puromycin-resistant cells on the outside of a colony could “shield” naïve ES cells at its core. Third, it is likely that while some cells received one copy of the shRNA vector, others had multiple integrations and thus produce more GFP. If GFP expression is directly representative of α-*Miwi2* shRNA expression, then GFP-negative cells will confound the analysis, as it is possible that any overt phenotypic effects of Miwi2 knockdown could be masked by rapidly dividing ES cells with no shRNA expression. This is a potential confounder in the crude phenotypic analysis; transduced mosaic cells were exposed to doxycycline without prior sorting for GFP. For microarray analysis however, R1-pRSIT-Scr and R1-pRSIT-Miwi2 cells were FACS-sorted based on GFP expression to mitigate the analysis of cells that did not contain an active hairpin. A caveat of this approach is that cell sorting generates limited material and this prevented me from assessing Miwi2 protein expression via western blot; it is possible that despite an >80% reduction in transcript expression,

Miwi2 protein may have a half-life that allows it to remain after 72 hours. This possibility, along with the observed mosaicism following shRNA-mediated knockdown of *Miwi2,* are potentially-confounding variables in analysis of its function; hence a different approach may be necessary to ascertain function of this pathway in ES cells, such as *in vitro* genomic disruption at the Miwi2 locus using RNA-directed endonucleases, also known as CRISPR (Cong et al., 2013).

Cells with reduced *Miwi2* function lack an obvious phenotype but undergo changes in expression of genes clustered in related ontologies; it is therefore possible that these genes somehow act to compensate for the loss of *Miwi2* to maintain the stem-like state. I also considered that Piwi proteins may function in a redundant manner and therefore Miwi and Mili could compensate for *Miwi2* knockdown. However, no significant changes in *Miwi* or *Mili* gene expression were identified by either microarray or qPCR analysis. Given that key pluripotency genes and lineage markers are regulated by small RNAs in ES cells (Houbaviy et al., 2003, Marson et al., 2008, Rosa et al., 2009), I hypothesised that the gene expression changes revealed by microarray could be indicative of perturbations at the small RNA level.

Overall length distribution of small RNAs in Day 0/3 embryoid bodies and R1-pRSIT cells are unique

Unexpectedly, the overall length distribution and categorical annotations of the total small RNA population of R1-pRSIT-Scr cells (Figure 3-19) are quite different to those of Day 0 embryoid bodies (Figure 3-7) even though both samples represent pluripotent R1 ES cell populations (albeit under different culture conditions). The unusual length distribution of small RNAs in R1-pRSIT cells is likely not indicative of RNA degradation, as the RNA integrity is high in all samples (Figure 8-6, Appendix). Rather, it may be a reflection of enrichment of a unique ES cell population: Ohnishi et al profiled small RNAs in 8- to 16-cell stage (8-16C) pre-implantation embryos and the contents of the ES-cell containing inner cell mass (ICM) of mouse blastocysts and reported an enrichment of pi-like RNAs and endogenous siRNAs and an underrepresentation of miRNAs in 8-16C embryos relative to ICM (Ohnishi et al., 2010). In my dataset, the total small RNA length profile of R1-pRSIT-Scr cells is closer to that of 8- to 16-cell embryos than contents of the ICM as observed by Ohnishi and colleagues. Others have also reported the transient reversion of cultured ES cells to a 2C-like state (Macfarlan et al., 2012); it is therefore possible that in sorting for high GFP fluorescence, I inadvertently enriched for cells that had adopted this early-stage transcriptional profile. I could address this possibility by repeating this experiment without FACS sorting prior to small RNA sequencing.

No miRNAs change in common during in vitro differentiation and following Miwi2 knockdown

I profiled miRNA expression in Day 0 (ES cell stage) versus Day 3 embryoid bodies to validate my *in vitro* differentiation trajectory. As deep sequencing is a semi-quantitative method (Creighton et al., 2009), I was able to determine relative small RNA sequence abundance based on read count. I compared my dataset (Table 3-1 and Table 3-2) with that of another group who profiled mouse ES cells (similar to my Day 0 sample) and 4 day *in vitro* differentiated ES cells (similar to my Day 3 embryoid bodies) (Houbaviy et al., 2003). Several miRNAs that were down-regulated >10-fold in my *in vitro* differentiation system were present in the ES cell-specific miRNA cohort presented by Houbaviy and colleagues: these were members of the miR-20b-5p, and miR-106a-5p families (Table 3-1). As expected, members of the ES cell-specific miR-302/367 family (Lin et al., 2008) also exhibited large changes in expression from Day 0 to Day 3 embryoid bodies. However, only some of the ES cell- and none of the embryoid body-specific miRNAs reported by Houbaviy et al exhibited expression changes in my *in vitro* differentiation system. It is possible that some miRNA changes were missed in my analysis with the stringent 10-fold expression change cut-off. Furthermore, unlike in my *in vitro* system, Houbaviy and colleagues differentiated their ES cells for an additional 24 hours and they also used retinoic acid (which I did not). These technical differences could account for differences among the small RNA datasets, and highlights the importance of thorough gene and small RNA characterisation of individual *in vitro* differentiation systems.

miRNA profiles changed dramatically during *in vitro* differentiation of ES cells and while the miRNA changes in *Miwi2* knockdown ES cells were more modest, there were some commonalities, although in opposite directions. There were increases in members of the miR-199 and miR-669 families in Day 3 embryoid bodies relative to Day 0 embryoid bodies and decreases in R1-pRSIT-Miwi2 cells compared to R1-pRSIT-Scr cells. miR-199a is a bone morphogenic protein 2-responsive miRNA which regulates chondrogenesis via direct targeting of the *Smad1* gene (Lin et al., 2009). Furthermore, members of the miR-669 family regulate the switch between cardiac and skeletal muscle lineages in postnatal cardiac progenitors via targeting of the master myogenic regulatory factor, MyoD (Crippa et al., 2011) and miR-669a has been suggested to promote MAPK signalling in primitive endoderm to maintain the multipotency of cells within the tissue (Spruce et al., 2010). The magnitude of the changes of these miRNA families in each scenario is far from equivalent: up

to 70-fold during differentiation and approximately 2-fold following *Miwi2* knockdown. Thus it is unlikely that *Miwi2*-knockdown cells exhibit a more stem-like miRNA profile.

Using deep sequencing, I was unable to validate the change in the ES cell-specific miR-302c revealed by the microarray. Furthermore, sequencing analysis revealed the up-regulation of six and down-regulation of two miRNAs in response to Miwi2 knockdown (Table 3-5 and Table 3-6) that was not detected by the microarray. There are several possible explanations for this discrepancy. First, next-generation sequencing is, arguably, a more sensitive method of detection (Malone and Oliver, 2011, Zhao et al., 2014). Second, probes antisense to one of the miRNAs, miR-199a, is not included on the microarray. Third, the miRNA probesets on the microarray cover the length of the primary transcript, hence will detect both the primary hairpin precursor and mature form of the miRNA (see the Affymetrix® probeset aligned with the miR-302/367 cluster in Figure 8-11, Appendix). If changes observed in miRNA expression are the result of increased precursor cleavage to the mature form rather than increased primary transcription, then the microarray may not detect these changes.

There is an increase in shorter sequences and a decrease in piRNA-sized small RNAs in Miwi2-knockdown ES cells

Following *Miwi2* knockdown, up-regulated RNA species were short (mostly 19-20nt), whereas down-regulated sequences were longer (23-32nt, Figure 3-20); this finding accounts for the changes in overall length distribution observed in R1-pRSIT-Miwi2 cells versus R1-pRSIT-Scr cells (Figure 3-19). In the testis, Miwi2 predominately binds piRNAs from 27-29nt, and small RNAs within this size range were down-regulated in ES cells with Miwi2 knockdown. It is therefore possible that Miwi2 is involved in the biogenesis of these pi-sized RNAs in ES cells. In the foetal testis, Miwi2 cleaves secondary piRNAs from endogenous transposon transcripts at position 10 as part of the ping pong mechanism (Aravin et al., 2008). If Miwi2 is involved in a ping pong-like piRNA biogenesis and transposon transcriptional silencing pathway in mouse ES cells, one would expect to observe a bias towards an A at position 10 in a subset of sequences, particularly those that map to transposable elements. However I find no overt preference for a 10A in the down-regulated RNAs that are derived from i) known piRNA clusters or ii) repeat elements (Figure 8-7, Appendix). This suggests the biogenesis of piRNAs in ES cells is ping pong-independent, much like in the somatic cells of the *Drosophila* ovary (Malone et al., 2009). Mili is required for homotypic ping pong biogenesis in murine pre-pachytene male germ cells (Aravin et al., 2007b, De Fazio et

al., 2011) and expression of the transcript is negligible in ES cells (Figure 3-1), hence it is unlikely that ping pong signatures would exist in the small RNAs of my dataset. If piRNAs do not depend on Miwi2 for ping pong-like biogenesis, but do require physical interaction with the protein to effect a function in ES cells, it would be expected that the small RNAs are depleted in the absence of Miwi2. This is observed in Miwi2-null foetal testes, where 29% of piRNAs are almost 10-fold down-regulated relative to wild type testes (Kuramochi-Miyagawa et al., 2008). The hypothesis that Miwi2 binds these pi-sized RNAs in ES cells is addressed in the Miwi2-pulldown analysis in Chapter 5 of this thesis.

The small RNAs that were up-regulated in *Miwi2* knockdown cells relative to scrambled controls are mostly 19-20nt in length. The significance of this peculiar size distribution and the origin of these small RNAs – mostly LTR and LINE repeats and introns – is at this point, unclear. Berninger and colleagues observed 19nt piRNA-like sequences in the adult mouse testis on the opposite strand to primary piRNAs (Berninger et al., 2011). The 5' end these 19mers is at a defined 19bp distance immediately upstream of the 5' end of the corresponding secondary piRNA; this group hypothesised these novel RNAs are processed from longer, primary piRNA-targeted transcripts concurrently with secondary piRNAs during ping-pong biogenesis. Furthermore, these 19nt small RNAs were detected in immunoprecipitates of Miwi and Mili from the adult testis, suggesting a physical interaction between the small RNA and Piwi proteins in the mouse testis. While Berninger et al observed the generation of these 19nt piRNA-related small RNAs from non-repeat sequences, Oey et al employed a similar method and observed many of these short RNA sequences – which they termed "pr19RNAs" – mapping to IAP LTR and LINE repeat elements (Oey et al., 2011). In fact, this group reported that pr19RNAs are the predominant species of all 19nt small RNAs in mouse spermatogenic tubules. It is possible that pr19RNAs are present in ES cells, and are up-regulated to compensate for *Miwi2* knockdown. However, given that i) these small RNAs are produced during ping pong biogenesis and ii) ping pong amplification is unlikely in mouse ES cells (owing to the near-absence of *Mili*), it is unlikely that these are *bona fide* pr19RNAs. A more comprehensive bioinformatic analysis of sequences downstream of these up-regulated 19nt RNAs would be required to address the possibility.

Small RNAs derived from the 5' ends of tRNA genes exhibit expression changes during in vitro differentiation and following Miwi2 knockdown

Intriguingly, I observed small RNAs mapping to transfer RNAs (tRNAs) in Day 0 and Day 3 embryoid bodies and a small number in R1-pRSIT cells (Figure 3-7 and Figure 3-19). tRNA-derived small RNAs comprise 12% of >10-fold down-regulated sequences during *in vitro* differentiation (Figure 3-8) and a smaller proportion of total down-regulated sequences (2%) after *Miwi2* knockdown. Furthermore, there are decreases in small RNAs derived from some of the same tRNAs (HisGTG and GlyGCC) during differentiation and following *Miwi2* knockdown. This is an intriguing finding, as recent studies have uncovered a new class of small RNAs – termed "tRNA fragments" or tRFs – cleaved from the 5' and 3' ends of tRNA by Dicer (Cole et al., 2009, Li et al., 2012). A 3' tRF, designated tRF-1001, has been shown to be required for prostate cancer cell proliferation (Lee et al., 2009). Despite weakly binding to Argonaute proteins (Cole et al., 2009), human 5' tRFs do not function like miRNAs or siRNAs: these small RNAs repress translation by directly associating with polysomes (Sobala and Hutvagner, 2013). This group has also shown that the tRFs do not regulate expression of their parent, full-length tRNA. Direct translational inhibition of gene expression by 5' tRFs has also been reported in the archeon *Haloferax volcanii* (Gebetsberger et al., 2012). The existence of 3' – but not 5' – tRFs has previously been identified in mouse embryonic stem cells (Babiarz et al., 2008). Given that Miwi2 is cytoplasmic in mouse ES cells, it is entirely possible that it could participate in the regulation of translation via the binding of pi-like tRFs. In fact, a function for piRNAs in translation has already been demonstrated (Lee et al., 2011). Miwi/piRNA complexes exhibit punctate expression in the cytoplasm of murine hippocampal neurons and antisense suppression of specific piRNAs results in a loss of dendritic spine area, a feature that is indicative of changes in neuronal translation.

It is possible that in complex with Miwi2, some of these 5' tRFs – such as those that are commonly down-regulated in differentiation and in *Miwi2* knockdown ES cells – act in a similar manner to human tRFs and directly affect translation to regulate gene expression in differentiating ES cells. It follows that these sequences would be down-regulated in response to the normal *Miwi2* diminution during lineage specification, which is observed in my *in vitro* differentiation trajectory. This possibility will be further addressed by deep sequencing of Miwi2-bound RNAs in wild type ES cells in Chapter 5 of this thesis.

Small RNAs derive from LTR and LINE retrotransposons in ES cells

During genomic transcription in zygotes, many transposable elements are transiently expressed until after the 2-cell (2C) stage, including LTR endogenous retroviruses (LTR_ERVs). It has been suggested that endogenous retroviruses invade the germlines of animal hosts via viral infection of germ cells where they stably integrate (Smit, 1993). Murine 2C embryos – but not ES cells – highly express LTR_ERV transcripts with a leucine tRNA primer (LTR_ERVL) (Evsikov et al., 2004, Peaston et al., 2004). Macfarlan and colleagues reported that LTR_ERVLs act as functional promoters for hundreds of 2C stage genes, for example the lineage markers *Gata4* and *Tead4,* via a 3'LTR-5'exon fusion (Macfarlan et al., 2012). The mammalian apparent LTR retrotransposon (MaLR) "superfamily" of repeat elements share structural similarities to LTR_ERVLs, however, do not produce viral proteins such as reverse transcriptase (Smit, 1993). Small RNA profiling of murine 2C embryos and blastocysts has revealed enrichment of small RNAs produced from ERVL elements after the 2C-stage, which are involved in silencing of the ERVL transcript itself (Ohnishi et al., 2010). Furthermore, small RNAs map to LINE and LTR repeats in the murine foetal testis (Kuramochi-Miyagawa et al., 2008). In fact, piRNAs aligning with intracisternal A particle (IAP) LTR retrotransposons and LINE_L1 and SINE_B1 elements have been demonstrated to directly associate with Miwi2 in the murine foetal testis (Aravin et al., 2008). In all samples, I observed close to 90% of all repeat-derived small RNAs aligning with similar sequences, in particular with MaLR and LINE_L1 repeats (Figure 3-7 and Figure 3-19). In fact, many of these small RNAs underwent >10-fold expression changes during normal differentiation (Figure 3-8) and >2-fold changes following *Miwi2* knockdown in ES cells (Figure 3-20B).

It is plausible that these small RNAs are involved in silencing of their parent repeat elements as lineage specification proceeds, either in complex with Miwi2 or as part of a separate silencing pathway. If Miwi2 and its bound piRNAs have a transposon silencing function, it follows that depletion of Miwi2 and these piRNAs would result in aberrant expression of these types of repeat elements as observed in the testis (Kuramochi-Miyagawa et al., 2008, Aravin et al., 2008). Investigation into transcript expression of murine LTR and LINE repeat elements could address this possibility. In Chapter 5 of this thesis, I address whether Miwi2 binds transposon-derived piRNAs in mouse ES cells.

Small RNAs in ES cells map to remnant transposable elements within introns

In all samples, the majority of gene-derived small RNAs align with the introns of protein-coding genes. This is not a novel discovery: hundreds of miRNAs are derived from intronic regions and are termed mirtrons (Ladewig et al., 2012). For example, miRNA clusters involved in growth and proliferation of mouse ES cells are derived from an intron of *Sfmbt2* (Zheng et al., 2011); several of the *Sfmbt2*-derived miRNAs were present in my datasets, such as members of the miR-669 family. The functions of others have been studied in less detail. One group has characterised a novel class of small RNAs in the *Xenopus tropicalis* embryo: small intronic transposable element RNAs (siteRNAs) (Harding et al., 2014). These small RNAs align in clusters to intronic regions of protein-coding genes and are derived from remnants of transposable elements. Harding and colleagues reported that genes giving rise to siteRNAs are enriched in repressive histone modifications, hence are lowly expressed. In *Xenopus*, these siteRNAs are derived from remnants of transposons present within introns (Harding et al., 2014). The group also observed diversity in the siteRNA clusters, with a number of similar but non-identical sequences mapping to common introns. Intronic reads make up 64-73% of gene-derived small RNAs in Day 0/Day 3 embryoid bodies and R1-pRSIT cells; these also comprise up to 25% of small RNAs with expression changes in both experiments. Intriguingly, many of these intronic RNAs also map to repeat element fragments, indicating that these could be *bona fide* siteRNAs. There is also overlap between genes giving rise to putative siteRNAs in Day 0/3 embryoid bodies and those that are down-regulated in R1-pRSIT-Miwi2 cells relative to scrambled controls. This is a novel finding, as the time of writing, there were no reports of siteRNAs in the mouse.

Small RNAs derived from unannotated regions of the genome in ES cells

Small RNAs mapping to unannotated intergenic regions comprise 15-25% of total small RNA reads in all samples and these increase in proportion both during normal ES cell differentiation and following *Miwi2* knockdown. Almost half of all up-regulated sequences in Day 0 and 3 embryoid bodies and a smaller proportion of sequences with expression changes in R1-pRSIT cells are derived from unannotated regions. This is intriguing, as the function of these small RNAs is currently unknown. On the other hand, the role of large intergenic noncoding RNAs (lincRNAs) has been fairly well characterised in ES cells: hundreds are found in ES cells and other progenitor cells and have an average mature spliced size of 3.2kb (Guttman et al., 2010). LincRNAs have an indispensable role in pluripotency and self-renewal of mouse ES cells: they are directly targeted by

important pluripotency-associated transcription factors and regulate gene expression via interaction with chromatin regulatory proteins (Guttman et al., 2011). Furthermore, knockdown of ES cell-specific lincRNAs results in spontaneous differentiation of mouse ES cells. It is possible that the unannotated intergenic small RNAs with dynamic expression changes in my Day 0 and Day 3 embryoid body datasets have similar diverse roles in pluripotency and lineage specification; at this point, it was unclear whether any of these unannotated sequences associate with Miwi2.

Miwi2 knockdown in mouse ES cells does not impact on in vitro differentiation

Many of the genes that were revealed by microarray to be down-regulated in response to *Miwi2* knockdown were involved in development; furthermore, the ectodermal marker, *Otx2,* is significantly down-regulated in two knockdown ES cell lines (R1-pRSIT-Miwi2_4 and R1-pRSIT-Miwi2_5, Figure 3-16 and Figure 3-17). It is important to note that lineage marker genes (such as *Otx2*) are expressed at a low level in ES cells, which begs the question of whether the change in gene expression is biologically relevant. However, lineage specification is a highly regulated process and changes in lowly expressed genes at the ES cell stage might still have significant downstream effects. This is particularly relevant for *Otx2* which – although traditionally described as a factor for neuroectodermal specification – is also required for i) maintenance of the ESC state and ii) for the transition of ES cells to epiblast stem cells during early stages of differentiation, prior to the emergence of primitive germ layers (Acampora et al., 2013). It is possible that a more directed differentiation protocol – such as retinoic acid-induced neural differentiation – may have uncovered statistically significant gene expression changes in lineage markers such as *Otx2*.

There was no overt phenotype of embryoid bodies derived from *Miwi2*-knockdown and Scr control ES cells. On the whole, there were no significant deviations in the expression patterns of selected pluripotency and lineage marker genes between scrambled controls and knockdown cell lines over the 6 days, indicating proper formation of primitive germ layers (Figure 3-21 and Figure 3-22). There were significant differences between scrambled control and knockdown cells in expression of *Oct4* and *Sox17* at single time points during *in vitro* differentiation, but I predict that these are stochastic rather than biologically relevant effects of *Miwi2* knockdown. One of the *Miwi2*-knockdown clones (R1-pRSIT-Miwi2_5c) displays aberrant expression patterns of *Nanog, Sox2* and *Oct4* and all differentiation genes, although this is not paralleled by its replicate cell line (R1-pRSIT-Miwi2_5c_rpt, Figure 8-9 and Figure 8-10, Appendix). This highlights the importance of targeting a gene of interest with multiple unique shRNAs (Kaelin, 2012) to rule out that any large

perturbations in gene expression – such as the complete absence of *Brachyury* expression – are a result of off-target effects of an individual hairpin.

There are, however, limitations to a targeted gene expression approach; there may well be changes in expression of other genes involved in lineage specification that my analysis has overlooked. Another approach that I could have taken is to perform microarray on R1-pRSIT-Scr versus R1-pRSIT-Miwi2 embryoid bodies at various time points, which would provide a broader overview of gene expression dynamics during *in vitro* differentiation. The caveat of mosaic hairpin expression also remains in this experiment, as R1-pRSIT cells were not sorted for GFP expression prior to embryoid body formation as sterility cannot be guaranteed during flow cytometry. In a different approach to ascertaining Miwi2 function, I attempted to overexpress the protein in mouse ES cells, with no success: transduction efficiency with the Miwi2 ORF-containing vectors was very low in ES cells and antibiotic selection for positive clones was not possible in this scenario.

In summary, *Miwi2*-knockdown cell lines appear to differentiate normally and exhibit correct expression of selected lineage markers over a 6 day *in vitro* differentiation time course. While the altered gene expression and small RNA profiles in Miwi2 knockdown ES cells suggest Miwi2 has some role in ES cells, it is likely not an absolute requirement for their pluripotency.

Conclusion

Here, I have demonstrated that Miwi2 is expressed in mouse embryonic stem cells and that expression diminishes concurrently with increased transcript expression of ectodermal and mesodermal germ layer markers and numerous changes in the small RNAome during *in vitro* differentiation. Several of these small RNA changes are also observed in ES cells following *Miwi2* knockdown, although the changes are much more subtle. Furthermore, several changes in global gene expression are observed in *Miwi2*-knockdown ES cells, although overall, this does not induce overt effects on stem cell self-renewal or differentiation. The evidence I have presented in this chapter suggests that Miwi2 is dispensable for embryonic stem cell self-renewal and pluripotency; however, given that its expression is enriched in pluripotent cells relative to differentiated cells, I am yet to rule out a function of the Piwi/piRNA pathway in stem cells. In Chapter 5 of this thesis, I will investigate the small RNA binding partners of Miwi2 in embryonic stem cells in an attempt to further elucidate the function of the Piwi/piRNA pathway in somatic cells.

4. Miwi2 expression and function in cancer cells

Introduction

There is evidence that neoplastic cells aberrantly adopt stem cell-like qualities in the process of malignant transformation: cancer cells proliferate in an undifferentiated state, relinquish Polycomb silencing of tumour suppressor regions (Widschwendter et al., 2007), aberrantly activate Wnt signalling (Rubinfeld et al., 1996), and exhibit an imbalance between self-renewal and differentiation based on microenvironmental cues (Li and Neaves, 2006, Malanchi et al., 2012). Furthermore, there is evidence that stem-like populations with the propensity for asymmetric division reside within tumours, and that these populations contribute to metastasis (reviewed in (Visvader and Lindeman, 2008)). It has also been suggested that certain cancers, such as chronic myeloid leukaemia, arise from adult multipotent stem cells present in healthy tissue via a multi-step clonal process to generate stem-like cells with uncontrolled proliferation (Sloma et al., 2010). Hence, many of the molecular pathways involved in the maintenance of stemness in pluripotent and multipotent stem cells are likely also to be active in cancer cells. The critical role of the Piwi/piRNA pathway in self renewal of GSCs makes it a candidate stem cell pathway to examine in cancer cells.

The Piwi proteins and piRNAs have a documented role in cancer: the *Drosophila* Piwi transcripts, *Piwi* and *Aub,* are up-regulated in *Drosophila* larval brain tumours (Janic et al., 2010), and Piwi proteins and piRNAs have been identified in MDAMB231 human breast and HeLa cervical cancer cell lines, respectively (Lee et al., 2010, Keam et al., 2014, Lu et al., 2010,). Others have identified enrichment of *Hili*, the human orthologue of Mili, in a putative cancer stem cell subset of 231 human breast cancer cells; knockdown of *Hili* leads to perturbations in anti-apoptosis and proliferation pathways in these cells (Lee et al., 2010).

In this chapter, I address the hypothesis that Piwi proteins are involved in maintaining self-renewal in cancer cells, and cancer stem cells. I utilised a syngenic panel of mammary carcinoma cell lines known as the 4T1 panel. It comprises five cell lines: 67NR, 168FARN, 4TO7, 66cl4 and 4T1 (Aslakson and Miller, 1992). The 67NR, 168FARN, and 4TO7 cells are non-metastatic, whereas 66cl4 and 4T1 spontaneously metastasise to varying degrees. Following implantation into the mouse mammary fat pad, 67NR cells do not leave the site of injection, 168FARN migrate to the

lymph nodes only, and 4TO7 disseminate via the blood to the lungs but are unable to proliferate (Aslakson and Miller, 1992). Aslakson and colleagues also reported that metastatic 66cl4 cells circulate mostly via the lymphatic system to establish tumours in the lungs, and 4T1 cells spread to the lungs and liver predominately via the venous route. The ability of some of the cell lines to create metastatic colonies may be related to their stem-like properties.

The specific aims of this chapter are to:

i. Determine whether members of the Piwi family are expressed in the 4T1 panel of murine mammary carcinoma cell lines.
ii. Investigate the effects of *Miwi2* knockdown in metastatic 4T1 cancer cells.
iii. Investigate the phenotypic effects and the mRNA and small RNA expression changes in non-metastatic 4TO7 cancer cells following ectopic *Miwi2* expression.

I examined the transcript level of the three murine Piwis, *Miwi, Mili*, and *Miwi2*, in the 4T1 cancer panel using qPCR. Based on these results, I focused further investigation on Miwi2. I investigated Miwi2 protein expression and localisation in 4T1 cells via western blot and immunofluorescence. I then attempted lentiviral shRNA knockdown of *Miwi2* in 4T1 cells and I introduced ectopic expression of *Miwi2* in lowly-expressing 4TO7 cells to observe phenotypic changes. I also investigated gene expression changes in *Miwi2*-expressing 4TO7 cancer cells via microarray, and small RNA populations via next-generation sequencing.

Results

4.1 Murine Piwi transcript expression in syngenic murine mammary carcinoma cell lines

Murine Piwi expression in the 4T1 panel

If metastatic cancers contain a cancer stem cell subpopulation, and Piwi genes and proteins are expressed in stem cells, I expected that the murine Piwi gene expression would correlate with metastatic potential within the syngenic 4T1 panel. Thus, I analysed *Miwi, Mili* and *Miwi2* expression in highly metastatic 4T1 cells by qPCR (Figure 4-1). Similar to ES cells and other murine somatic cells, 4T1 cells express negligible levels of *Miwi* and *Mili* relative to Day 19 (D19) testis. However, *Miwi2* expression is over two-fold higher in 4T1 cells than D19 testis.

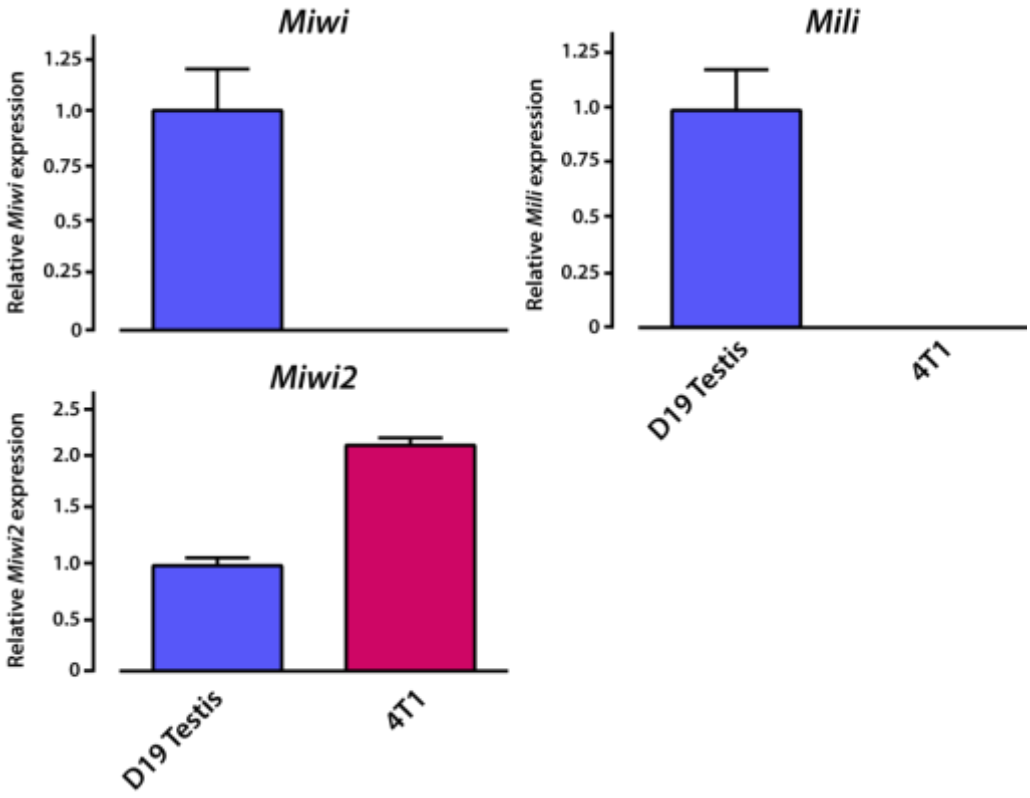


Figure 4-1: *Piwi* gene expression in metastatic 4T1 mammary carcinoma cells

Quantitative PCR data showing transcript expression of *Miwi, Mili* and *Miwi2* in 4T1 cancer cells relative to D19 testis. Error bars represent standard error of technical triplicates within each qPCR analysis.

I then investigated whether *Miwi2* transcript is expressed in other members of the 4T1 panel and if so, whether *Miwi2* levels correlate with the reported propensity of each cell line for tumorigenesis and metastasis. Figure 4-2 demonstrates an imperfect correlation between *Miwi2* expression and reported metastatic potential (Aslakson and Miller, 1992). The non-metastatic cell lines 67NR and 168FARN express *Miwi2* at a similar level to ES cells and adolescent testis, respectively. The 4TO7 non-metastatic cell line – which is able to migrate to a distant location but unable to establish a secondary tumour – has negligible expression of *Miwi2*. The metastatic cell line, 66cl4, expresses *Miwi2* at a level approaching that of ES cells, over two-fold less than what is observed in the 4T1 cells. As the 4T1 cells express higher levels of *Miwi2* transcript relative to other members of the murine mammary carcinoma panel, and 4TO7 cells express negligible levels of *Miwi2*, I focused upon these cell lines for further descriptive and functional investigation of Miwi2 in cancer cells.

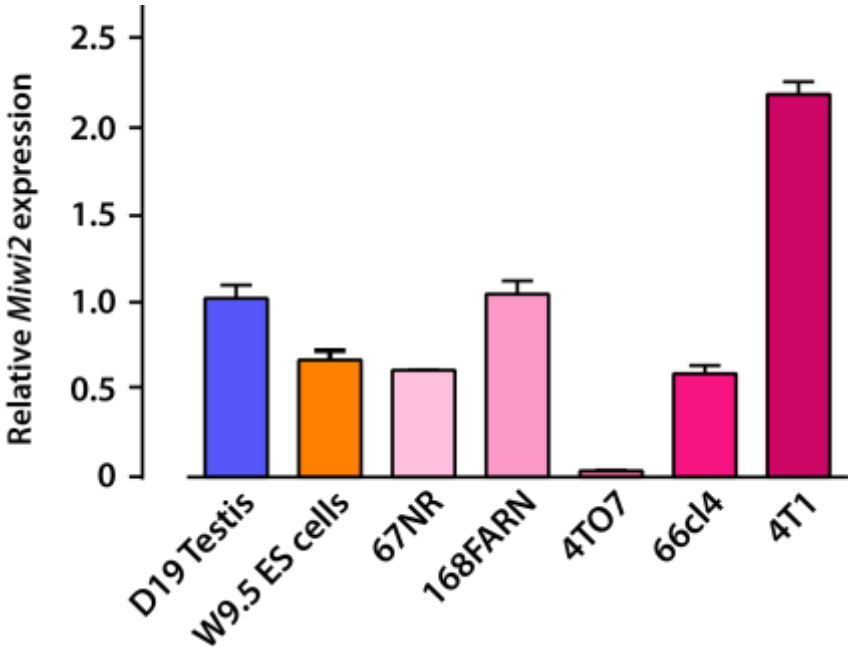


Figure 4-2: *Miwi2* transcript expression in syngenic 4T1 panel of mammary carcinoma

Quantitative PCR data showing transcript expression of *Miwi2* in D19 testis (blue bar), W9.5 ES cells (orange bar) and members of the 4T1 panel of cell lines (pink bars, depth of colour indicative of ability for metastasis). Transcript expression of all samples is depicted relative to testis. Error bars represent standard error of technical triplicates within each qPCR analysis.

Miwi2 expression in putative 4T1 tumour-initiating cells

Given that stem cells are thought to be required for tumour metastasis and *Miwi2* is enriched in highly invasive and metastatic 4T1 cells, I hypothesised that this Piwi protein may function to promote a stem-like phenotype in a discrete subset of 4T1 cells. I have previously discussed reports of the Inhibitor of DNA-binding (ID) proteins as markers for putative cancer-initiating cells. In 4T1 cells, ID1 expression induces a mesenchymal-to-epithelial transition in cancer stem-like cells to promote 4T1 metastatic colonisation *in vivo* (Stankic et al., 2013). I obtained samples of ID1-expressing and non-expressing 4T1 cell total RNA from the Swarbrick laboratory (Garvan Institute of Medical Research). ID1-positive and –negative cells can be separated via FACS from a 4T1 subline carrying a GFP reporter downstream of the ID1 promoter (Mellick et al., 2010). Notably, only a very small proportion (<5%) of 4T1 cells are positive for ID1 (Alexander Swabrick, personal correspondence). I used quantitative PCR to determine *Miwi2* transcript expression in unsorted 4T1 cells, ID1-GFP positive 4T1 cells and ID1-GFP negative cells (Figure 4-3). *Miwi2* levels are slightly enriched in ID1-expressing cells, but not depleted in ID1-negative cells.

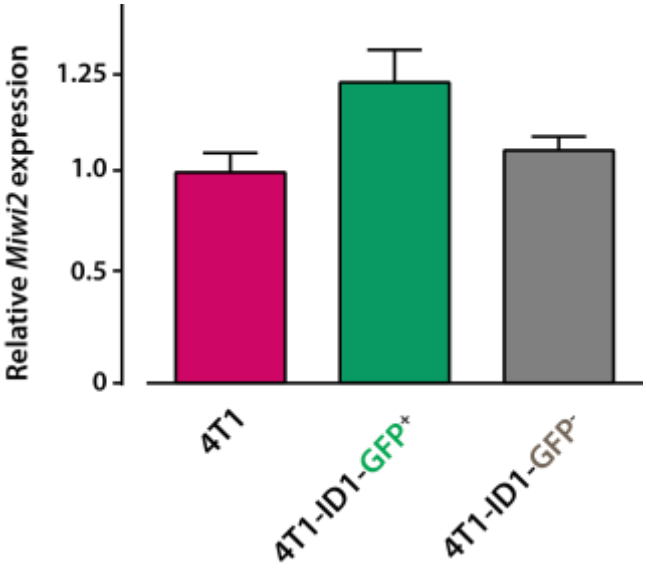


Figure 4-3: *Miwi2* transcript expression in 4T1 cells enriched for ID1 expression

Quantitative PCR data showing *Miwi2* expression in unsorted (pink bar), 4T1-ID1-GFP$^+$ (green bar) and 4T1-ID1-GFP$^-$ cells (grey bar). Transcript expression of all samples is depicted relative to 4T1. Error bars represent standard error of technical triplicates within each qPCR analysis. Statistical testing was not performed as n=1 for all samples.

Miwi2 protein expression and localisation in 4T1 mammary carcinoma cells

I then investigated whether *Miwi2* transcript is translated into protein in 4T1 cancer cells, and confirmed that it is (Figure 4-4A). Figure 4-4B (left panel) also confirms the presence of the human orthologue Hiwi2 in 231 human breast cancer cells; this sample was included as a positive control as Hiwi2 expression in these cells has been reported previously (Lee et al., 2010, Keam et al., 2014). I demonstrated the specific interaction of the rabbit polyclonal (ab21869) antibody with Miwi2 and Hiwi2 via immunoblotting with peptide-blocked antibody (Figure 4-4B, right panel) and immunoblotting of IP products (see Figure 8-3, Appendix).

Other groups have reported that Miwi2 is restricted to the nucleus of embryonic germ cells (Aravin et al., 2008), but I have established that Miwi2 is predominately cytoplasmic in mouse ES cells. Using immunofluorescence with fluorescent dyes for nucleic acid and cytoskeletal actin filaments, and the polyclonal antibody against Miwi2 (Abcam ab21869), I find that Miwi2 is expressed in cytoplasmic clusters in 4T1 cancer cells (Figure 4-4C). Intriguingly, there is heterogeneity in 4T1 expression of Miwi2: a proportion of cells within the population displays little or no Miwi2 staining (white arrows, Figure 4-4C).

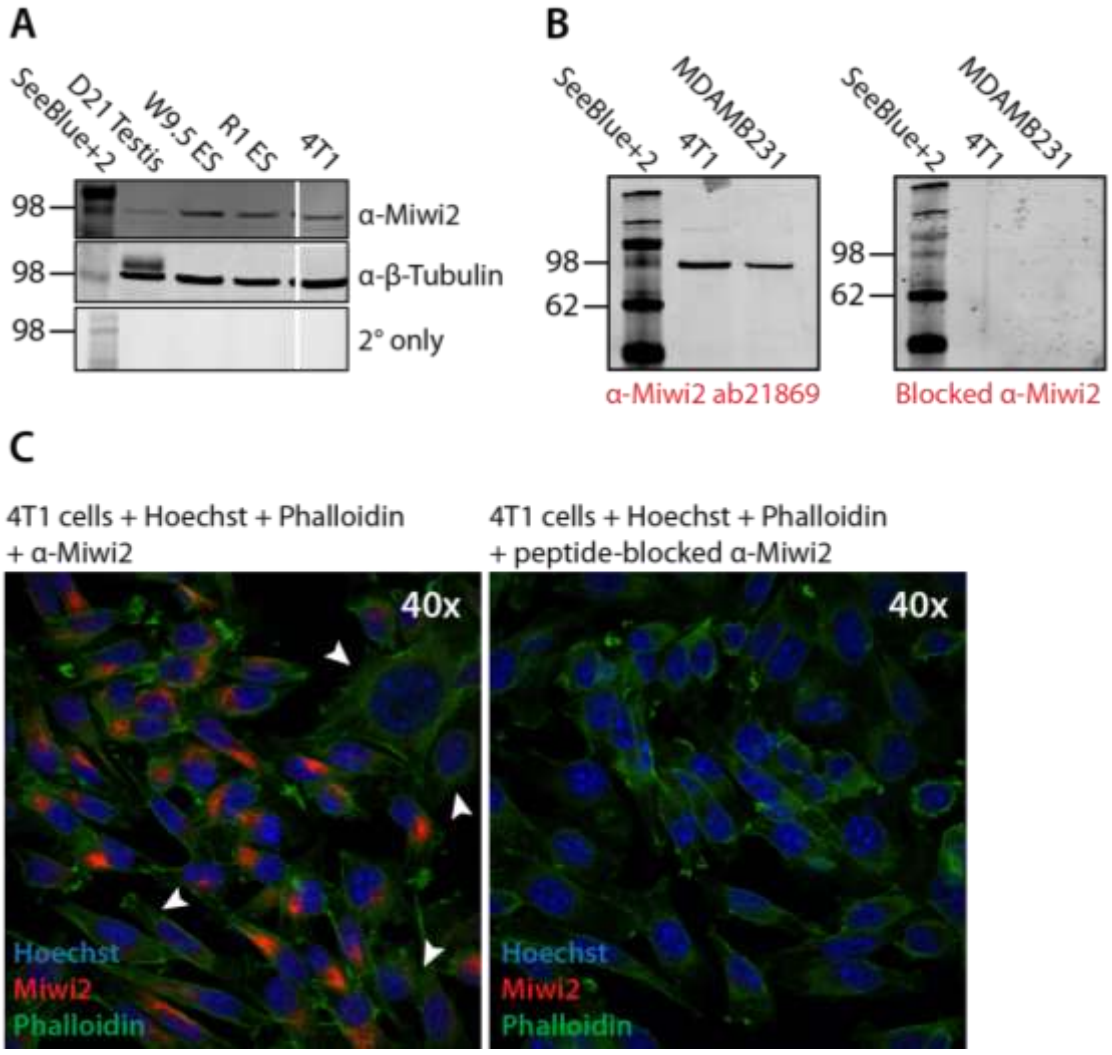


Figure 4-4: Miwi2 protein is expressed in the cytoplasm of 4T1 cancer cells

(A) Western blot showing Miwi2 expression (top panel) in whole testis and 4T1 cell lysate using polyclonal α-Miwi2 antibody (Abcam ab21869). B-tubulin and β-actin (bottom panel) were used as loading controls. (B) Western blot demonstrating antibody specificity for Miwi2 using i) α-Miwi2 polyclonal antibody and ii) α-Miwi2 polyclonal antibody blocked with Miwi2-specific peptide. (C) Immunofluorescence showing cytoplasmic localisation of Miwi2 in W9.5 ES cell colonies (left panel). Cells with negligible Miwi2 expression are labelled with white arrows. Antibody specificity for Miwi2 is demonstrated using negative control ES cells stained with a peptide-blocked α-Miwi2 antibody (right panel). Imaged on a Zeiss LSM Upright 700.

4.2 *Miwi2* knockdown in metastatic 4T1 cancer cells

Miwi2 is highly expressed in 4T1 murine mammary carcinoma cells; given that loss of *Piwi* inhibits tumour formation and invasion in *Drosophila* (Janic et al., 2010), here I postulated that knockdown of *Miwi2* in 4T1 cells could induce expression changes of genes involved in tumorigenesis and metastasis. Earlier in this thesis, I discussed the optimisation strategy for lentiviral packaging and transduction, including knockdown of *Miwi2* using 5 different shRNAs (see Section 3.4, Chapter 3). When transduced individually, only 2 of the 5 hairpins (pRSIT-Miwi2_4 and pRSIT-Miwi2_5) were capable of knocking down *Miwi2* expression by at least 80% relative to scrambled controls (see Figure 8-5, Appendix). Therefore, I packaged together both vectors encoding these shRNAs into infectious lentiviral particles to achieve maximum knockdown in 4T1 cells.

Optimisation of 4T1 cancer cell line stable selection

The pRSIT puromycin resistance marker enables selection of 4T1 cells with stable shRNA expression. It was first necessary to determine the optimal concentration for selection: I exposed 4T1 cells to puromycin over 8 days. I chose a range of concentrations based on other studies, which have reported puromycin concentrations of 2µg/ml up to 10µg/ml required for 4T1 sensitivity (Kim et al., 2010, Tao et al., 2008). The optimal antibiotic concentration is the minimum dose that results in 100% cell death after several days. Hence, I selected for stably-transduced 4T1 clones after lentiviral transduction using 6µg/ml puromycin (Figure 4-5).

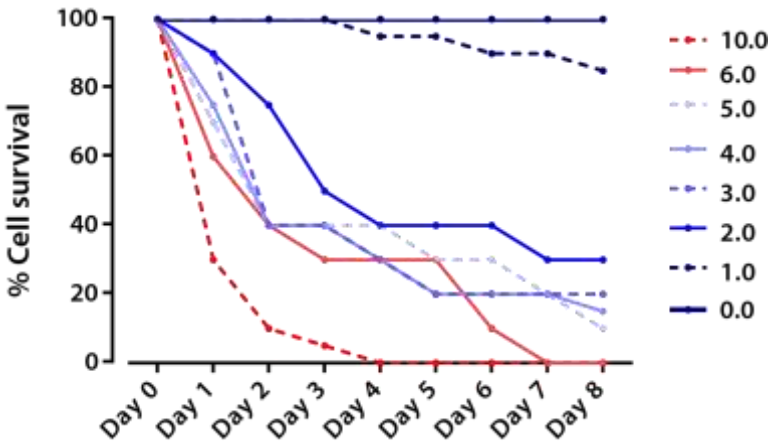


Figure 4-5: Puromycin dose response in 4T1 mammary carcinoma cells

Concentrations of puromycin (in µg/ml) inducing 100% cell death are depicted in shades of red; concentrations resulting in incomplete cell death are shown in shades of blue.

Miwi2 knockdown in 4T1 cells

Following puromycin selection at 6µg/ml for 8 days (until 100% cell death was observed in the non-transduced control), resistant colonies displayed mosaic GFP expression (Figure 4-6A). If GFP expression is indicative of hairpin expression, this implies that there will also be mosaic expression of the integrated shRNA-containing element. Thus, I sorted the cells on the basis of GFP expression to enrich for a 4T1 population with the potential for high shRNA expression. It is important to note that unlike with R1-pRSIT ES cells in Chapter 3, here I sorted for GFP expression in 4T1-pRSIT *prior to* inducing shRNA expression with doxycycline as cancer cells can be readily expanded without a feeder layer, which could have confounded the ES cell analysis. Figure 4-6B indicates that most transduced cells are GFP positive, although a small proportion are GFP-negative and some are also dead cells positive for 7AAD. The brightest 58-61% of GFP-positive 4T1 cells were collected ("GFP POS", Figure 4-6B) and replated prior to induction of α-*Miwi2* shRNA and subsequent transcript expression analysis. After 72h doxycycline treatment, the four biological replicate knockdown cell lines (4T1-pRSIT-Miwi2_4/5a) exhibited an average of 40% knockdown relative to the three scrambled controls (4T1-pRSIT-Scr_1, Figure 4-6C). Thus, unlike ES cells that exhibited more than 80% *Miwi2* depletion with the pRSIT lentiviral system, 4T1 cells appear to be largely resistant to *Miwi2* knockdown.

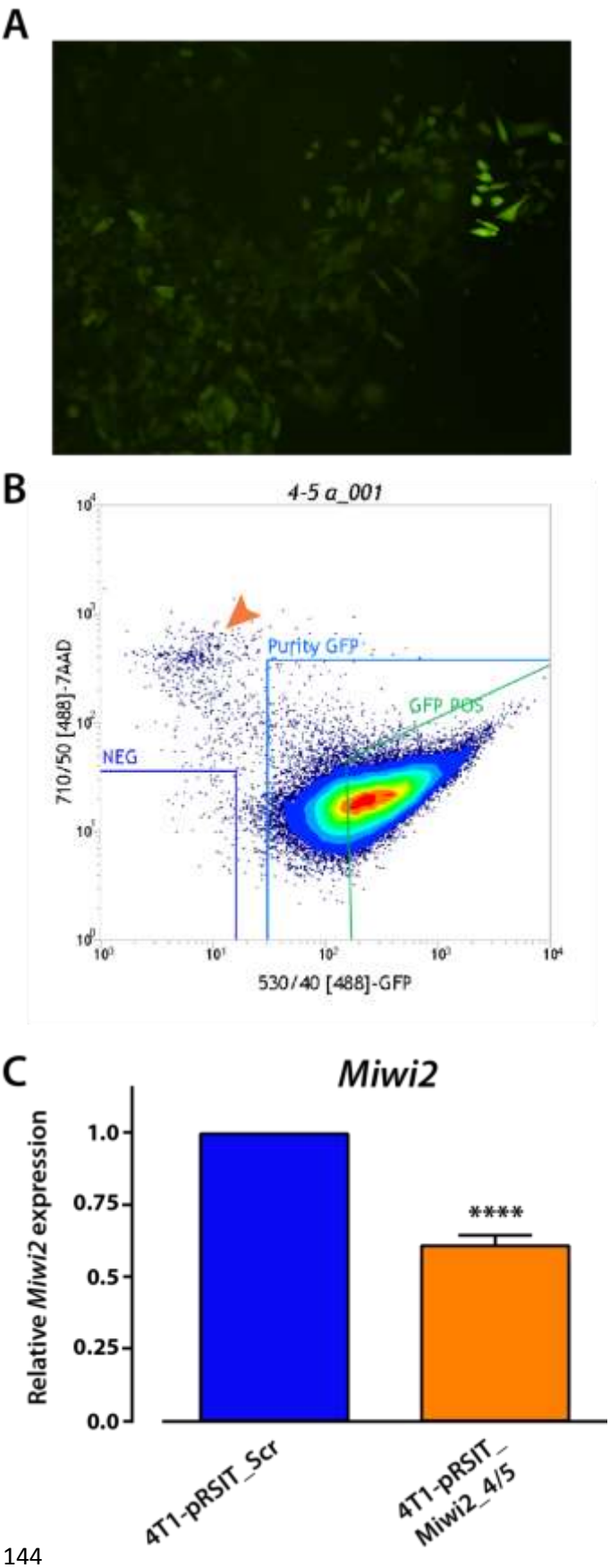


Figure 4-6: Quantitative PCR detection of *Miwi2* transcript in GFP$^+$ 4T1 cells after induced shRNA expression

(A) Puromycin-resistant 4T1 cancer cells display mosaic GFP expression. Imaged on a Nikon Ti-S Research Inverted Fluorescence Microscope. (B) Scatter plot depicting gating strategy to isolate GFP-expressing 4T1 cells. Dead cells are marked by 7AAD (y-axis, orange arrow); α-Miwi2 shRNA-expressing cells are marked by GFP (x-axis, GFP POS) (C) Quantitative PCR data showing *Miwi2* expression in knockdown 4T1 cells (4T1-pRSIT_Miwi2_4/5) relative to scrambled controls (4T1-pRSIT_Scr). Error bars represent standard error of biological replicates within each qPCR analysis (4T1-pRSIT-Scr n=3, 4T1-pRSIT-Miwi2 n=4). P-values were determined via student's t-test (two-tailed, unpaired) analysis (****, p<0.0001).

4T1 escape doxycycline-induced knockdown Miwi2 knockdown

To investigate the apparent resistance of 4T1 cells to *Miwi2* knockdown, I repeated doxycycline induction with a single 4T1-pRSIT-Miwi2 cell line from the previous analysis and monitored *Miwi2* expression daily over 96h relative to triplicate scrambled controls with 24h doxycycline treatment. Quantitative PCR reveals a knockdown of 50% relative to controls after 24-48h doxycycline treatment, then *Miwi2* levels rise to approximately 65% after 72h (my previous time point for analysis) and after 96h, *Miwi2* transcript expression reverts back to basal levels (Figure 4-7A).

It has been reported that 4T1 cells exhibit doxorubicin resistance due to the expression of P-glycoprotein (P-gp, or MDR1), an ATP-binding transporter implicated in acquired multidrug resistance of tumour cells (Bao et al., 2011). Doxycycline is known to induce expression of P-gp in the metastatic human breast cancer cell line, MCF7 (Mealey et al., 2002). I considered the possibility that that the induction of shRNA expression by doxycycline is inhibited by P-gp-mediated multidrug resistance in 4T1-pRSIT cells. P-gp can be inhibited by treatment with verapamil, an L-type calcium channel blocker (Laberge et al., 2009). As per the procedure outlined by (Bao et al., 2011) to overcome doxorubicin resistance in 4T1 cells, I treated a single 4T1-pRSIT-Miwi2 cell line with verapamil at 20µM for 2h prior to, and over the course of, 96h doxycycline induction. However, with concurrent exposure to verapamil, doxycycline failed to induce *Miwi2* knockdown at all relative to scrambled controls, even at the early time points (Figure 4-7B). The mechanism by which the 4T1 cells escape *Miwi2* silencing is still unknown, but it was clear that a different strategy was needed to gain insight into the function of Miwi2 in these cells.

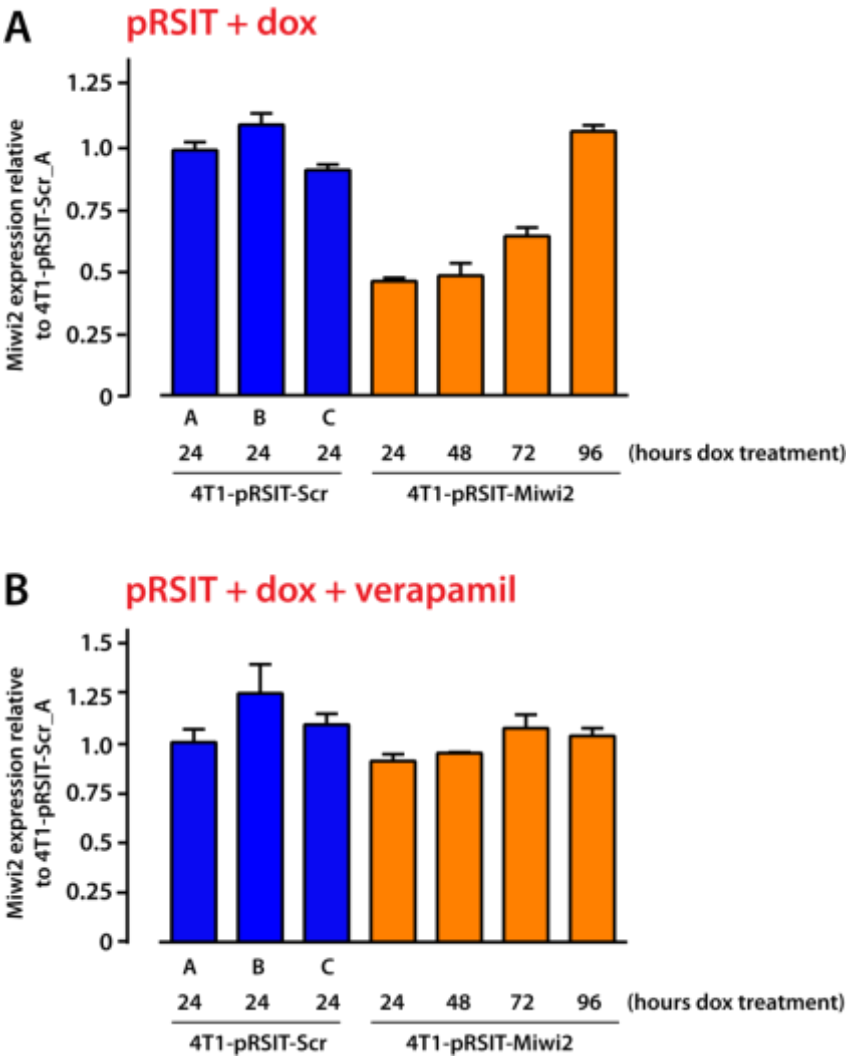


Figure 4-7: Quantitative PCR data demonstrating 4T1 evasion of doxycycline-induced *Miwi2* knockdown over 96 hours

Quantitative PCR data showing *Miwi2* transcript expression during 96h doxycycline induction in a single *Miwi2*-knockdown 4T1 cell line (4T1-pRSIT-Miwi2) relative to controls (4T1-pRSIT-Scr). Error bars represent standard error of technical quadruplicates within each qPCR analysis (n=1). (A) *Miwi2* expression in *Miwi2*-knockdown 4T1 cells during 96h exposure to doxycycline. Samples were harvested at 24h intervals. (B) *Miwi2* expression in *Miwi2*-knockdown 4T1 cells treated with verapamil 2h prior to and during 96h exposure to doxycycline. Samples were harvested every 24h.

4.3 Ectopic *Miwi2* expression in 4TO7 cancer cells

Given that 4T1 cells exhibit resistance to knockdown of *Miwi2*, I embarked on a different approach to investigate *Miwi2* function in cancer cells. Rather than knocking the protein down in the high expressing 4T1 cells, I chose to express Miwi2 in the syngenic sibling cell line, 4TO7, that has negligible expression of the transcript (refer to Figure 4-2). In this section, I describe the phenotypic, gene expression, and small RNA changes in 4TO7 cancer cells with ectopic *Miwi2* expression.

Again, I employed a lentiviral packaging and transduction approach to generate 4TO7 clones that stably expressed *Miwi2*. I constructed a lentiviral vector (pLenti-C-turboGFP, Origene) containing the Miwi2 ORF with a C-terminal GFP tag and used an empty pLenti-C-turboGFP as a control. This vector lacks a selection marker, so stable clones were established on the basis of fluorescence and cell sorting rather than antibiotic resistance. Lentivirus was packaged and infectious particles were harvested 30h post-transfection as described in Materials and Methods. For each vector, 4TO7 cells were transduced in triplicate. Several days post-transduction, GFP-positive cells were selected by flow cytometry (Figure 4-8A). Packaging and transduction of lentivirus into 4TO7 cells was more efficient using the empty vector than the Miwi2 ORF-containing vector; prior to enrichment, 29-38% 4TO7-GFP cells were GFP positive, but only 0.2-0.5% 4TO7-Miwi2 cells expressed GFP-fused Miwi2. As previously discussed, vector size is an important factor for lentiviral titre during packaging; it is likely that the smaller size of pLenti-GFP contributes to the higher transduction efficiency. After enrichment for GFP-positive cells by FACS, *Miwi2* expression was 25-fold higher in 4TO7-Miwi2-GFP relative to 4TO7-GFP (Figure 4-8B).

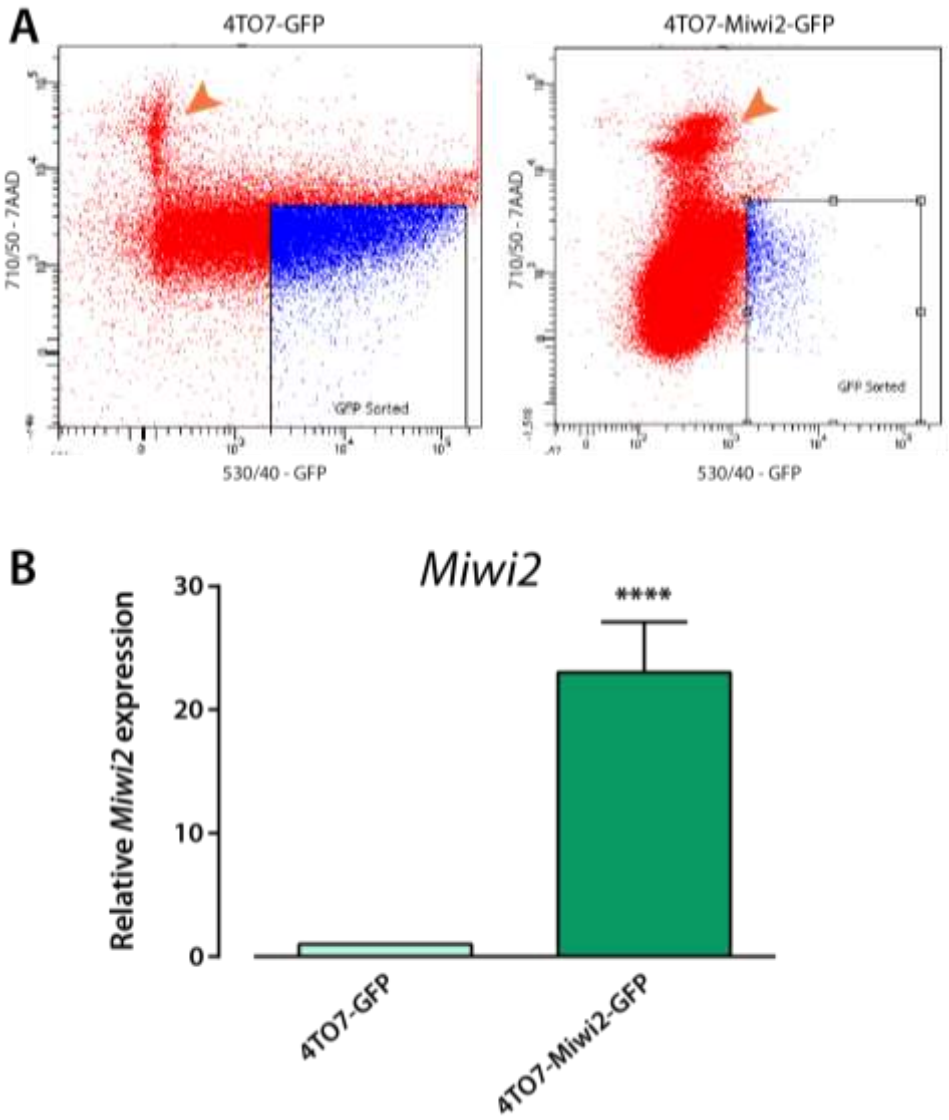


Figure 4-8: Quantitative PCR detection of *Miwi2* transcript in GFP$^+$ 4TO7 mouse mammary carcinoma cells after induced shRNA expression

(A) Scatter plot depicting gating strategy to isolate GFP-expressing 4TO7 cells. Dead cells are marked by 7AAD (y-axis, orange arrows); α-*Miwi2* shRNA-expressing cells are marked by GFP (x-axis, "GFP Sorted") (B) Quantitative PCR data showing *Miwi2* transcript levels following Miwi2-GFP expression in non-metastatic 4TO7 cells. *Miwi2* expressing cells are denoted by "4TO7-Miwi2-GFP"; control cells are labelled "4TO7-GFP". Error bars represent standard error of biological replicates within each qPCR analysis (n=3). P-values were determined via student's t-test (two-tailed, unpaired).

Effects of Miwi2 *expression in 4TO7 cancer cells: phenotype*

If ectopic *Miwi2* expression induces a more tumorigenic or metastatic phenotype in 4TO7 cells, I expected that this would be reflected by observable phenotypic changes. Visually, there is no overt difference in appearance with ectopic *Miwi2,* although it appears that there are more dividing cells (indicated by round cells with glowing edges, Figure 4-9A). Hence, I measured cell size (Figure 4-9B) and doubling time (indicated by cell density, Figure 4-9C) in 4TO7-Miwi2-GFP cells relative to controls over 8 days. At Day 2 and Day 6, there is a significant, but very small difference in cell size ($p < 0.05$, Figure 4-9A). Furthermore, 4TO7-Miwi2-GFP cells demonstrate a significantly higher doubling time at Days 2 and 4 than control cells ($p < 0.05$), although this is not seen at later time points (Figure 4-9B).

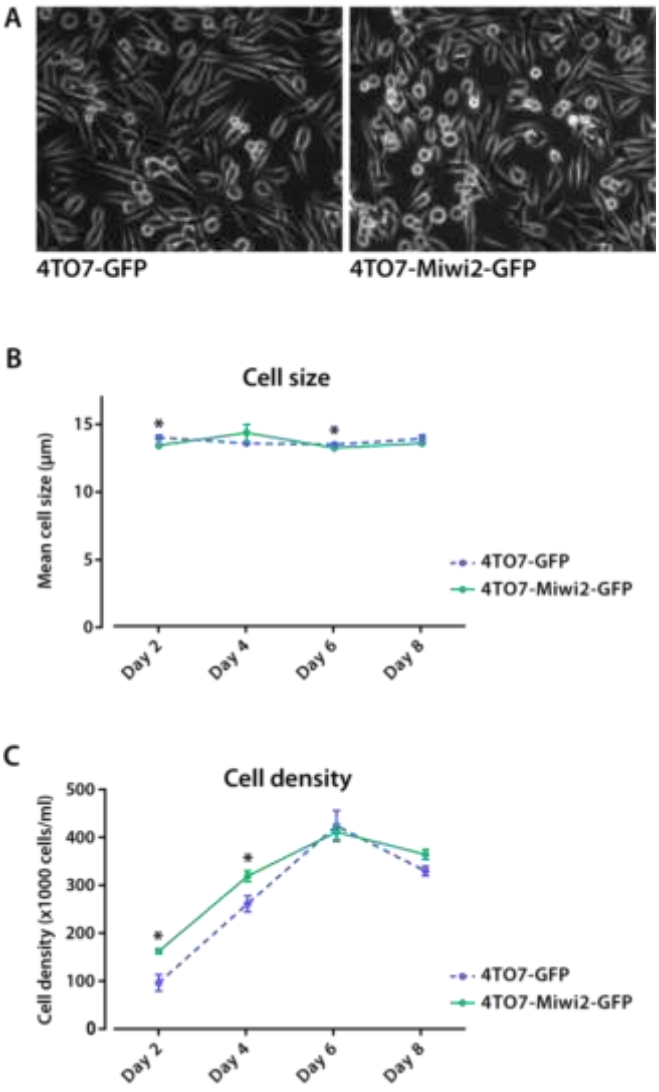


Figure 4-9: Phenotypic observation of 4TO7-Miwi2-GFP versus control cells

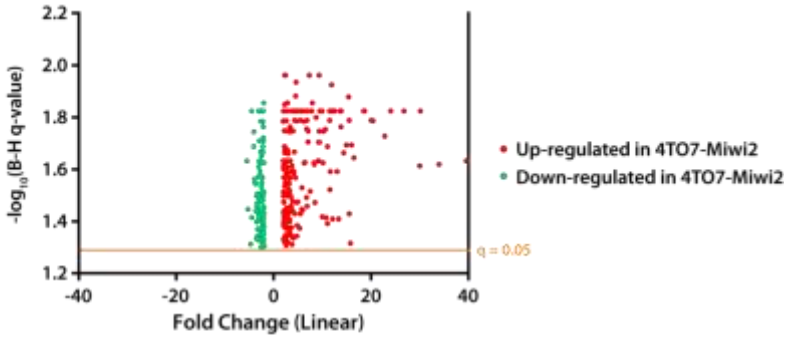
(A) Light microscope images of control (4TO7-GFP) and *Miwi2*-expressing cells (4TO7-Miwi2-GFP) after enrichment for GFP expression. Imaged on a Nikon Ti-S Research Inverted Phase Contrast Microscope, 200x magnification. (B-C) Combined phenotypic measurements from biological triplicates each of control (4TO7-GFP, blue line) and *Miwi2*-expressed 4TO7 cells (4TO7-Miwi2-GFP, green line). (B) Mean cell size of 4TO7-Miwi2-GFP compared to 4TO7-GFP cells ($n = 3 \times 10^6$). (C) Cell density of 4TO7-Miwi2-GFP compared to 4TO7-GFP cells. P-values were calculated via student's t-test (two-tailed, unpaired) analysis comparing biological triplicates of control (GFP) vs *Miwi2*-expressing (Miwi2) 4TO7 cells (*, $p < 0.05$).

Effects of Miwi2 *expression in 4TO7 cancer cells: gene expression*

There are no overt changes in cell size or appearance, and only a small change in the rate of cell division following ectopic expression of *Miwi2* in 4TO7 cells, but this does not preclude the induction of some phenotype that is more subtle, or apparent only under certain circumstances. To examine whether ectopic *Miwi2* expression causes changes to gene expression profiles, I performed transcript profiling on both 4TO7-Miwi2-GFP and 4TO7-GFP using the Affymetrix® Mouse Gene 2.0 expression array. Differential gene expression was determined using Affymetrix® Transcriptome Analysis Console (TAC) with Benjamini-Hochberg correction for multiple testing. A large number of significant gene expression changes occurred with *Miwi2* expression: 216 annotated transcripts or miRNAs were differentially expressed with a fold change of > ±2.0 and q-value <0.05; 200 species were up-regulated and only 16 species were down-regulated. Figure 4-10A is a volcano plot showing differential gene expression as a function of linear fold change versus significance; genes with the largest and most significant fold changes tended to be those whose expression increased with *Miwi2* expression. Figure 4-10B shows a heat map generated by Chipster (Kallio et al., 2011) which depicts global changes in gene expression between the datatsets. The top 10 up- and down-regulated RefSeq transcripts and miRNAs are listed in Figure 4-10C. Six of the 16 down-regulated genes are RIKEN clones or predicted genes. Four of the top ten down-regulated transcripts following *Miwi2* expression in 4TO7 cells (Figure 4-10) were microRNAs. One of these miRNAs, Let-7c, regulates the expression of several genes, including *Myc*, *Mycn, Sall4,* and *Lin28a*. *Mycn, Sal4,* and *Lin28a* were unchanged on the array, but *Myc* increased by 1.8-fold (q-value = 0.06).

I used Ingenuity Pathway Analysis to generate a list of putative gene regulatory networks affected by *Miwi2* over-expression (Table 4-1). Many of the networks encompassing multiple genes with significant expression changes are involved in immune response and interferon signalling. Figure 4-11 illustrates the top affected regulatory network: “Endocrine System Disorders, Gastrointestinal Disease, Immunological Disease”. Table 4-2 lists the genes that are enriched in both i) 4TO7-Miwi2 cells relative to 4TO7-GFP controls and ii) wild-type 4T1 cells relative to non-metastatic 67NR and 168FARN cells as described by Tao and colleagues (Tao et al., 2008).

A



B

Linear Fold Change	Top down-regulated	Linear Fold Change	Top up-regulated
↓2.52	miR-3094	↑39.59	Ifi44
↓2.46	miR-872	↑33.97	Oas2
↓2.35	Psg17	↑30.16	Usp18
↓2.3	Nap1l2	↑30.02	Oas1g
↓2.29	Ndrg2	↑26.78	Mx2
↓2.24	miR-let-7c-2	↑24.08	Irf7
↓2.17	Egln3	↑22.8	Cxcl10
↓2.13	Aldoc	↑20.04	Cmpk2
↓2.1	CD38	↑18.68	Iigp1
↓2.06	miR-5125	↑18.48	Irgm2

C

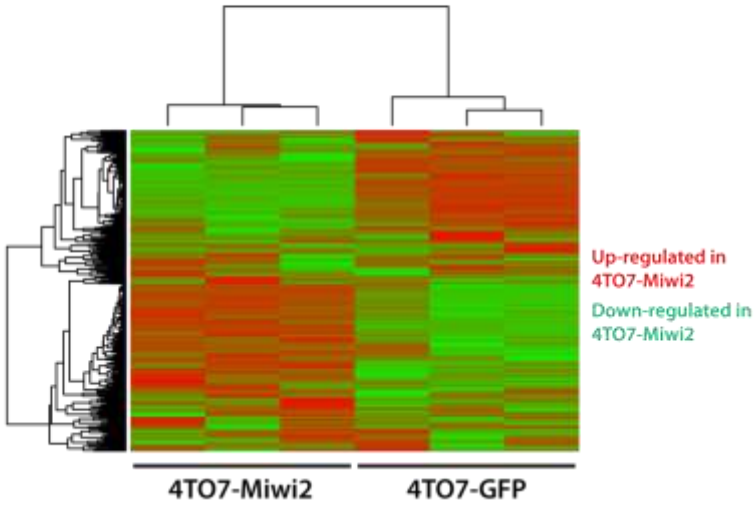


Figure 4-10: Affymetrix GeneChip® Mouse Gene 2.0 ST Array downstream analysis following *Miwi2* expression in 4TO7 cancer cells

(A) Volcano plot generated by GraphPad Prism depicting genes with significant gene expression changes. Values are plotted as a function of linear fold change (x-axis) versus significance ($-\log_{10}$q-value, y-axis). (B) Transcriptome Analysis Console (TAC)-generated list of genes with largest fold change ($q < 0.05$). (C) Heat map depicting supervised clustering generated by Chipster (Kallio et al., 2011) for genes with largest significant expression changes (fold change > ±2.0, $q < 0.05$) as revealed by microarray analysis following *Miwi2* expression.

Table 4-1: List of top predicted gene regulatory networks affected by *Miwi2* expression

List of networks identified by Ingenuity Pathway analysis involving > 10 gene/molecules identified as differentially expressed after *Miwi2* expression (fold change > ±2.0, q <0.05).

Network #	# Genes	Top functions and associated diseases
1	29	Endocrine System Disorders, Gastrointestinal and Immunological Disease
2	27	Antimicrobial Response, Inflammatory Response, Infectious Disease
3	22	Immunological disease, Infectious Disease, Protein Synthesis
4	18	Developmental disorder, Hereditary disorder, Immunological Disease
5	16	Cell-To-Cell Signaling and Interaction, Hematological System Development and Function, Immune Cell Trafficking
6	16	Cellular development, Cellular Growth and Proliferation, Cell cycle
7	15	Infectious Disease, Immunological Disease, Cell death and survival
8	13	Cell death and survival, Renal Necrosis/Death, Inflammatory response

Table 4-2: Genes enriched in both 4TO7-Miwi2 relative to 4TO7-GFP and wild-type 4T1 cells relative to non-metastatic syngenic cell lines, 67NR and 168FARN

Symbol	Gene description	NCBI RefSeq	Fold change (4TO7-Miwi2)
Atf3	activating transcription factor 3	NC_000067.6	5.52
C3	complement component 3	NC_000083.6	4.54
Ccl5	chemokine (C-C motif) ligand 5	NC_005109.3	9.27
Cxcl16	chemokine (C-X-C motif) ligand 16	NC_000077.6	3.38
F2rl1	coagulation factor II receptor-like 1	NC_000079.6	1.55
Gbp1	guanylate binding protein 2b	NC_000069.6	6.91
H2-M3	histocompatibility 2, M region locus 3	NC_000083.6	2.17
Irf1	interferon regulatory factor 1	NC_005109.3	1.99
Lcn2	lipocalin 2	NC_000068.7	2.1
Stat1	signal transducer and activator of transcription 1	NC_000067.6	11.92
Usp18	ubiquitin specific peptidase 18	NC_000072.6	30.16

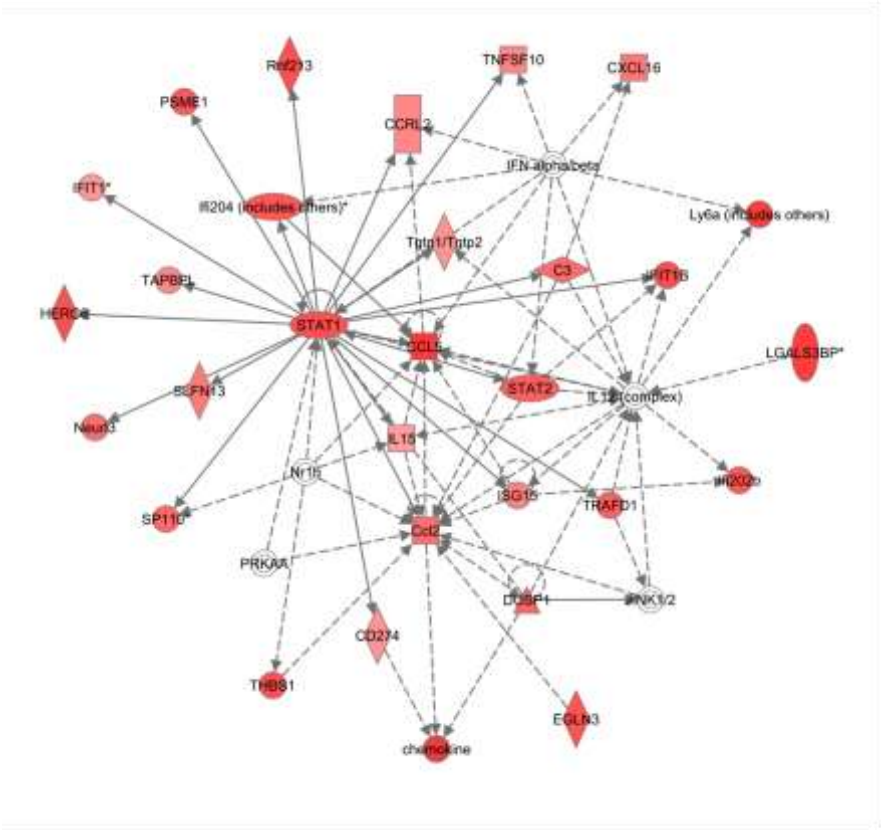


Figure 4-11: Ingenuity Pathway Analysis following *Miwi2* expression in 4TO7 mammary carcinoma cells

Top predicted network of differentially expressed genes from microarray data analysis (fold change > ±2.0, q-value < 0.05). Network is "Endocrine System Disorders, Gastrointestinal Disease, Immunological Disease". Schematic generated using Ingenuity Pathway Analysis (Ingenuity Systems). Down-regulated genes are shown in red; genes with no expression changes are shown in white. Arrows indicate direction of regulation. Dashed lines indicate a predicted relationship; solid lines indicate an empirically-validated relationship.

4.4 Small RNA expression with ectopic *Miwi2* in 4TO7 cancer cells

Following ectopic *Miwi2* expression in 4TO7 cells, I observed transcript changes and down-regulation of several miRNAs on the Affymetrix array (Figure 4-10). I hypothesised that changes in expression of other small RNAs could also be occurring, hence performed next-generation sequencing on control GFP-expressing and *Miwi2*-expressing 4TO7 cells to i) confirm altered expression of those miRNAs highlighted by the array (and any others) and ii) determine whether the small RNA population as a whole might be affected by ectopic *Miwi2*. It is possible that some of the gene expression changes that occur in response to *Miwi2* expression are mediated by microRNAs that are not represented on the array.

Characterisation of small RNAs in 4TO7-GFP and 4TO7-Miwi2 cells

I performed Illumina HiSeq® next-generation sequencing of total small RNA from biological duplicate control (4TO7-GFP) and *Miwi2*-expressing cells (4TO7-Miwi2). All findings for 4TO7-GFP and 4TO7-Miwi2 presented in this section are averaged from biological duplicate results. Primary bioinformatics analysis was performed with Paul Young (Suter Laboratory, Victor Chang Cardiac Research Institute). Following QC (see Section 2.11.2, Materials and Methods), I had approximately 25 million reads from my two control 4TO7-GFP samples and 26 million reads from my 4TO7-Miwi2 samples (Table 8-1, Appendix). Of those reads, the majority mapped successfully to the mouse genome when allowing 1 mismatch (85% for 4TO7-GFP; 78% for 4TO7-Miwi2). For 4TO7-GFP samples, I obtained 96,050 unique sequences and for 4TO7-Miwi2, 101,231 unique sequences. My sequencing dataset included many single copy reads (41% for each sample), indicating that the small RNA sequencing was not exhaustive.

Size distribution and annotations of the small RNAs in 4TO7-GFP and 4TO7-Miwi2 are shown in Figure 4-12. The majority of small RNAs from GFP control and *Miwi2*-expressing 4TO7 cells are miRNA-sized (between 21-24nt), although there is a noticeable increase in the proportion of 27-28nt sequences following *Miwi2* expression (Figure 4-12A and D). In both sets of samples, most small RNAs map to annotated regions of the genome, although the percentage of small RNAs derived from unannotated regions in 4TO7-Miwi2 is close to three times that of 4TO7-GFP. Figure 3-7 (B and E) depicts the breakdown of annotated small RNAs. As expected, the most abundant class of small RNAs in all samples is miRNA. There was no noticeable change in the percentage of miscellaneous non-coding RNA. I observed increases in the overall proportion of transposon-, gene- and tRNA-derived small RNAs and a concurrent decrease in the overall proportion of miRNAs following *Miwi2* expression. The overall increase in small RNAs mapping to transposable elements is mainly due to an increase in species mapping to

LTR repeats in 4TO7-Miwi2 relative to 4TO7-GFP; these make up 70.8% and 79% of total repeat-derived small RNAs, respectively. Figure 4-12C and F list the types of repeat elements which give rise to the greatest number of small RNAs in control and *Miwi2*-expressing 4TO7 cells: repeat-derived small RNAs map preferentially to LTR_ERVL-MaLR sequences (57.02% in 4TO7-GFP and 49.96% in 4TO7-Miwi2).

With the assistance of Paul Young (Suter Laboratory, VCCRI), I performed piRNA cluster analysis as described in 2.11.2, Materials and Methods. For 4TO7-GFP and 4TO7-Miwi2 cells, 5% and 4% of unannotated small RNAs fell within known piRNA clusters, respectively. Of those annotated, only a small percentage mapped to piRNA cluster regions (0.4% for control cells and 1.0% for *Miwi2*-expressing cells). In 4TO7-GFP cells, 6.3% of repeat-derived small RNAs and 0.4% of gene-derived reads were co-localised with the described piRNA clusters compared with 13.0% and 0.3% of small RNAs from 4TO7-Miwi2 cells, respectively.

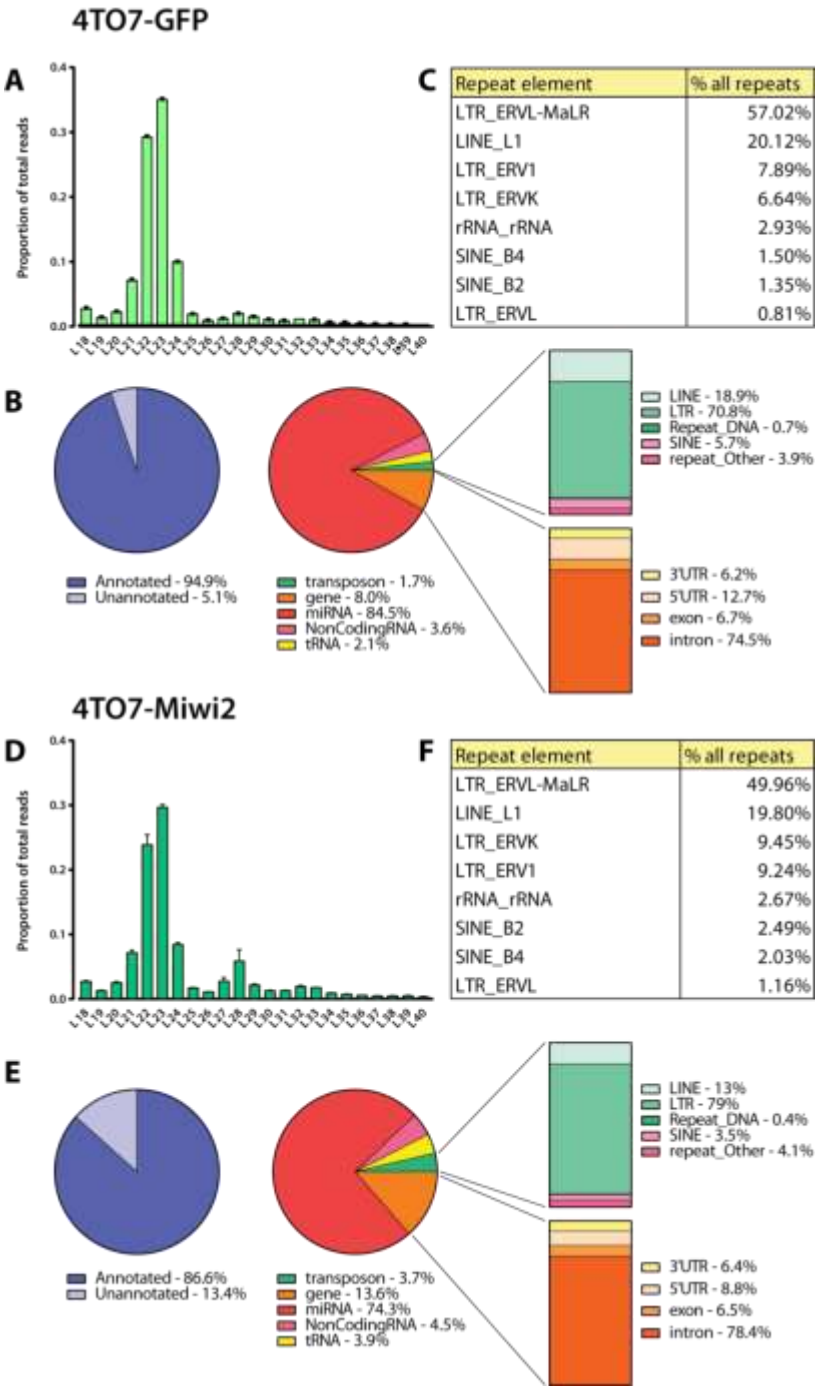


Figure 4-12: Characterisation of small RNAs in control and *Miwi2*-expressing 4TO7 cells

Small RNA sequence characteristics of (A-C) control 4TO7-GFP cells and (D-F) *Miwi2*-expressing 4TO7 cells. (A, D) Length distribution of small RNAs between 18-40nt. Values represented as a proportion of total small RNA. (B, E) Categorical annotations of small RNAs. (C, F) List of most abundant repeat elements deriving small RNAs.

Small RNA expression changes following Miwi2 expression in 4TO7 cells

In order to identify specific small RNA species with altered expression in response to ectopic *Miwi2*, data were filtered for i) >100 reads in averaged 4TO7-Miwi2 samples for analysis of up-regulated sequences or >100 reads in averaged 4TO7-GFP samples for analysis of down-regulated sequences, ii) linear expression change >±1.5-fold, and iii) p-value <0.05. Within these parameters, 112 unique sequences were up-regulated and 67 were down-regulated more than 1.5-fold following *Miwi2* expression. Of those small RNAs that were up-regulated following *Miwi2* expression in 4TO7 cells, the majority were either unannotated (23%), miRNAs (28%) or derived from tRNAs (29%); a moderate percentage of up-regulated sequences also mapped to intronic regions (9%, Figure 4-13A). Of those sequences that decreased in expression >1.5-fold, the majority (84%) were annotated as miRNAs (Figure 4-13B).

There is a global increase in 27-28nt small RNAs in response to Miwi2 expression

There was an enrichment of 27-28nt small RNAs in 4TO7-Miwi2 relative to 4TO7-GFP (Figure 4-12A and D). Given that Miwi2 preferentially binds 27-29nt RNAs (Aravin et al., 2008), I investigated whether the 27-28nt sequences in 4TO7-Miwi2 cells have piRNA characteristics. Only ten 27-28nt small RNA species were up-regulated >1.5-fold with p-value <0.05 and are listed in Table 4-3. One of these small RNAs has a 5’ U and maps to a tRNA gene within a known piRNA cluster on Chromosome 10 (tRNA90-GluCTC). Overall, these sequences exhibit no sequence bias at position 1, and at position 10, an underrepresentation of U and an overrepresentation of G (Figure 4-14A). Given that I used a small sample size (2x2) for this experiment, there is a high chance of obtaining Type II errors (false negatives) in my analysis. Hence, I also analysed 27-28nt sequences with >1.5-fold increases in expression without filtering for p-value (134 sequences, average p-value = 0.23). Of these sequences, 37% begin with a 5’ U, although there is also a slight overrepresentation of C (48%) at position 1; furthermore there is a bias towards A at position 10 (42%, Figure 4-14C). Notably, 81% map to unannotated regions of the genome and 4% map to known piRNA clusters; it is possible that these unannotated sequences could be as yet unidentified piRNAs. Furthermore, these unannotated piRNAs exhibit a clear bias towards C and U at positions 1 and 10 (Figure 4-14). There is little-to-no representation of either A or G at either position.

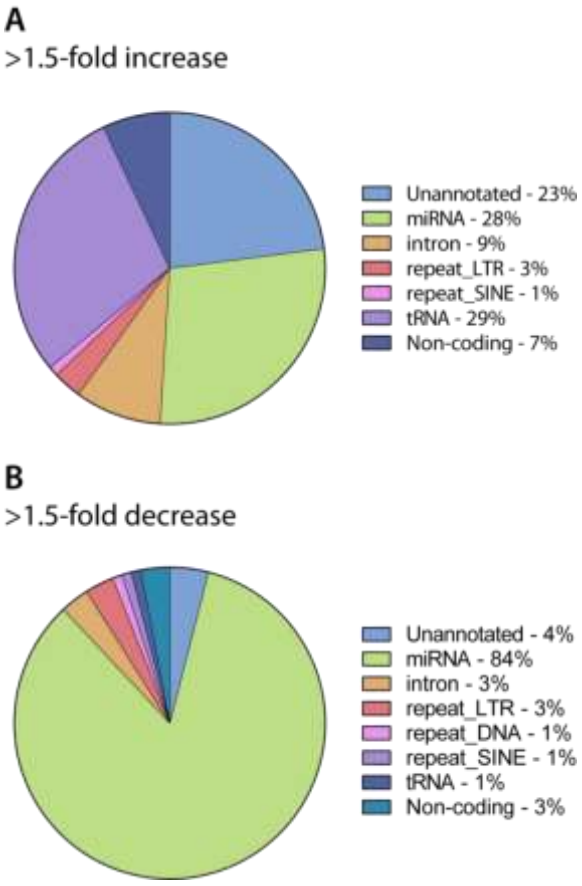


Figure 4-13: Categorical annotations of small RNAs with expression changes in 4TO7-Miwi2

Annotations of small RNAs with (A) >1.5-fold increase in expression (112 sequences) and (B) >1.5-fold decrease in expression (67 sequences).

Table 4-3: List of 27-28nt sequences with >1.5-fold increased expression in 4TO7-Miwi2

Length (nt)	Annotation	Specific annotation	Fold Change
27	NonCodingRNA	Snord53	1.52
27	tRNA|piRNA_Cluster	tRNA90-GluCTC	2.24
27	unannotated		2.25
27	tRNA	tRNA110-GlyGCC	2.77
27	unannotated		8.15
28	NonCodingRNA	Snord17	1.51
28	intron	Rpl5	1.77
28	tRNA	tRNA709-GluCTC	2.22
28	repeat_LTR_ERVL-MaLR	MTC-int	3.14
28	unannotated		7.48

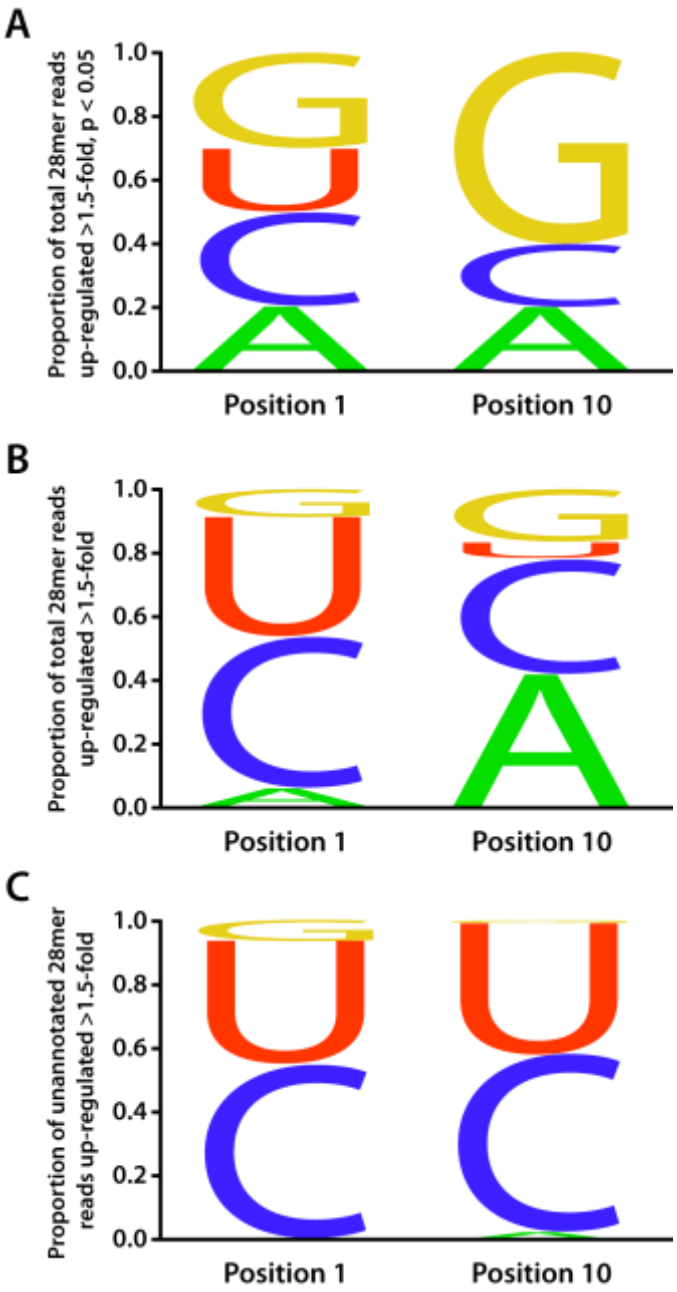


Figure 4-14: Sequence alignment of 27-28nt small RNAs >1.5-fold up-regulated in 4TO7-Miwi2 cells

(A) Positions 1 and 10 of sequences with p value < 0.05. (B) Positions 1 and 10 of sequences not filtered for p-value. (C) Positions 1 and 10 of unannotated sequences not filtered for p-value.

Small RNAs derived from unannotated regions

23% of the RNAs that were up-regulated following *Miwi2* expression in 4TO7 cells do not map to annotated regions of the genome; I interrogated the length and sequence composition for piRNA-like features. Unannotated small RNAs are predominately 23, 24 and 29nt in length (Figure 4-15A). Figure 4-15B reveals no overt sequence bias except an underrepresentation of U and a slight overrepresentation of A at position 10. Twenty seven percent of unannotated small RNA reads with increases in expression have a 5'U, but none of these map to known piRNA clusters.

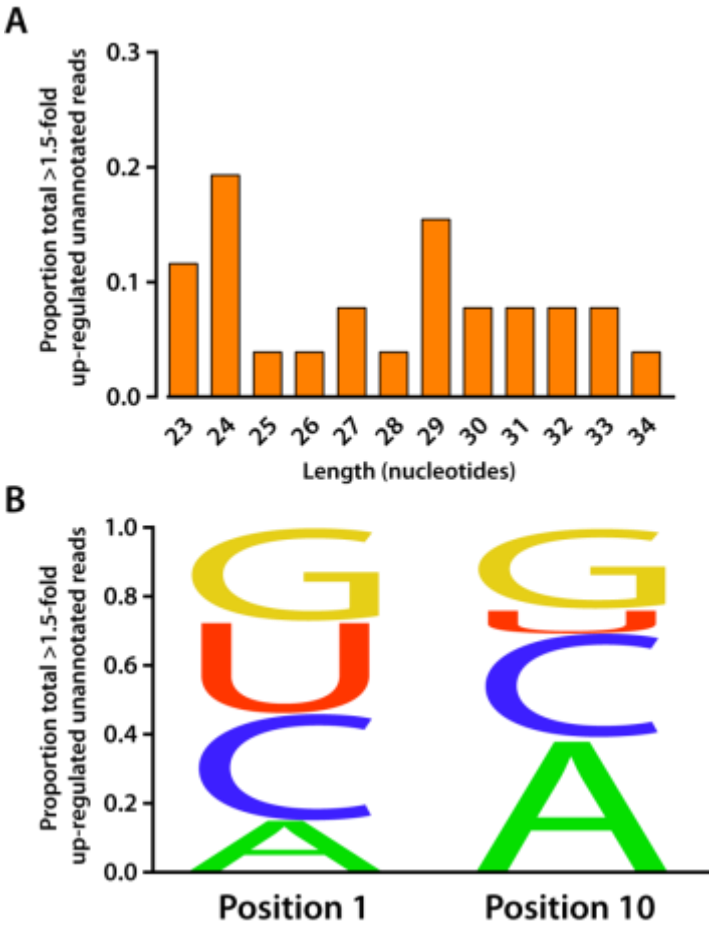


Figure 4-15: Characterisation of unannotated small RNA reads that were up-regulated in *Miwi2*-expressing 4TO7 cells

(A) Length distribution and (B) sequence bias of small RNA reads. Y-axis values representative of proportion of total unannotated up-regulated sequences.

Intragenic small RNAs

Miwi2 expression led to the increased production of small RNAs from the introns of 9 genes (Table 4-4); this comprises 9% total up-regulated reads in 4TO7-Miwi2 cells (Figure 4-13). Despite being a small sample size for analysis, I used GOrilla (Eden et al., 2009) to identify pathway associations of genes from which these small RNAs are derived, however I did not observe enrichment in any gene ontologies. Several sequences were given multiple annotations: for example, *Sfmbt2*-derived small RNAs are also annotated as miR-669a-3p as this miRNA is derived from the *Sfmbt2* intron (Zheng et al., 2011).

Table 4-4: List of genic introns giving rise to small RNAs that are up-regulated in 4TO7-Miwi2

Length	Intronic region	Putative gene function	Other annotation	Reference
20	Wwox	Mammary gland development Mammary tumour suppressor gene	miR-27b-3p	(Abdeen et al., 2013, Abdeen et al., 2011)
20	Clasp2	Microtubule dynamics Hematopoietic stem cell maintenance	LTR_ERVL_MaLR piRNA cluster	(Shahbazi et al., 2013) (Drabek et al., 2012)
21	Sfmbt2	Trophoblast maintenance and placental development	miR-669	(Miri et al., 2013)
21	Lrp1b	Signal transduction	LTR_LINE1 piRNA cluster	(Shiroshima et al., 2009)
23	Slc25a12	Myelination and maintenance of neurofilaments		(Sakurai et al., 2010)
28	Rpl5	Ribosomal protein		(Yu et al., 2006)
29-31	Tmtc1	Unknown	tRNA-GlyGCC	
33	Wdr43	Poly(A) RNA binding		NCBI Gene
34	Gnl3	Embryonic stem cell self-renewal Hematopoietic stem cell maintenance		(Qu and Bishop, 2012) (Yamashita et al., 2013)

miRNA expression changes in 4TO7-Miwi2 cells

Following *Miwi2* expression in 4TO7 cells, 12 miRNAs were up-regulated and 9 miRNAs were down-regulated >1.5-fold ($p < 0.05$) (Table 4-5 and Table 4-6). miRNAs that were down-regulated on the Affymetrix array (miR-3094, miR-872, miR-5125 and let-7c-2, Figure 4-10) did not emerge as significant candidates from next-generation sequencing. Three of these miRNAs did exhibit the expected expression trends: miR-3094-5p (1.96-fold), miR-872 (2.1-fold), and miR-5125 (4.13-fold) were down-regulated in 4TO7-Miwi2 cells, but were not flagged as significant via t-test (average p-value = 0.334).

I used DIANA mirPath (Vlachos et al., 2012) to elucidate the gene regulatory pathways targeted by the miRNAs with expression changes in my dataset. Targets were computationally predicted using DIANA-microT web server v5.0 (Reczko et al., 2012) and p-values of pathway hits were corrected for False Discovery Rate. Those miRNAs that are up-regulated >1.5-fold are involved in focal adhesion, axon guidance, and signalling pathways such as ErbB, PI3K-Akt, mTOR, TGF-beta, MAPK and Neurotrophin (Table 4-7). Down-regulated miRNAs target genes in multiple pathways: regulation of the actin cytoskeleton; cancers such as chronic myeloid leukaemia, prostate cancer, and melanoma; focal adhesion; and Wnt, PI3K-Akt, and ErbB signalling (Table 4-8).

I also used IPA to deduce the relationship between miRNA expression changes – as revealed by deep sequencing – and gene expression changes observed on the Affymetrix microarray following *Miwi2* expression. IPA MicroRNA Target Filter performs miRNA target analysis using TargetScan. IPA then pairs this data with a microarray-generated mRNA expression dataset to generate a comprehensive pathway analysis of miRNAs and target genes with significant expression changes. IPA identified 9 genes with expression changes of >2-fold in either direction, which are targeted by 2 miRNAs (miR-210-5p and miR-27b-3p) with significant changes of >±1.5-fold (Figure 4-16).

Table 4-5: List of miRNAs with >1.5-fold increase (p<0.05) after *Miwi2* expression

miRNA ID	4TO7-GFP read count	4TO7-Miwi2 read count	Fold Change
miR-5114	33	115	+4.41
mir-6236	252	634	+3.05
miR-467d-3p	60	110	+2.29
miR-27b-3p	611	1063	+2.22
miR-467a-3p	531	893	+2.18
miR-466b-3p	113	174	+2.12
miR-8112	474	700	+1.86
miR-669a-3p	293	424	+1.83
miR-3535	110	154	+1.81
miR-8114	227	295	+1.71
miR-132-3p	260	320	+1.62
miR-466c-5p	33	115	+1.53

Table 4-6: List of miRNAs with >1.5-fold decrease (p<0.05) after *Miwi2* expression

miRNA ID	4TO7-GFP read count	4TO7-Miwi2 read count	Fold Change
miR-210-5p	209	29	-6.42
miR-700-3p	954	431	-1.84
miR-674-5p	2087	957	-1.81
miR-6539	725	347	-1.73
miR-222-5p	630	306	-1.66
mir-125b-2	109	55	-1.64
miR-503-3p	196	100	-1.60
mir-27a	375	194	-1.58
miR-27a-5p	3343	1763	-1.55

Table 4-7: KEGG pathways of all miRNAs up-regulated >1.5 fold (p<0.05) in 4TO7-Miwi2

KEGG pathway	p-value	#genes	#miRNAs
Focal adhesion	2.53E-16	53	7
ErbB signalling pathway	2.75E-16	30	7
PI3K-Akt signalling pathway	8.63E-15	75	8
mTOR signalling pathway	4.66E-12	22	8
TGF-beta signalling pathway	5.47E-12	28	7
Axon guidance	8.74E-12	39	7
MAPK signalling pathway	2.75E-11	58	8
Prostate cancer	5.74E-11	26	8
Neurotrophin signalling	4.69E-10	32	7

Table 4-8: KEGG pathways of all miRNAs down-regulated >1.5 fold (p<0.05) in 4TO7-Miwi2

KEGG pathway	p-value	#genes	#miRNAs
Pathways in cancer	6.52E-10	34	6
Regulation of actin cytoskeleton	2.64E-08	23	5
Wnt signaling pathway	2.64E-08	20	6
Chronic myeloid leukemia	5.84E-07	11	5
PI3K-Akt signaling pathway	6.54E-07	30	6
Focal adhesion	2.68E-06	20	4
Prostate cancer	3.80E-06	12	4
Melanoma	3.98E-06	11	4
ErbB signaling pathway	4.93E-06	11	4

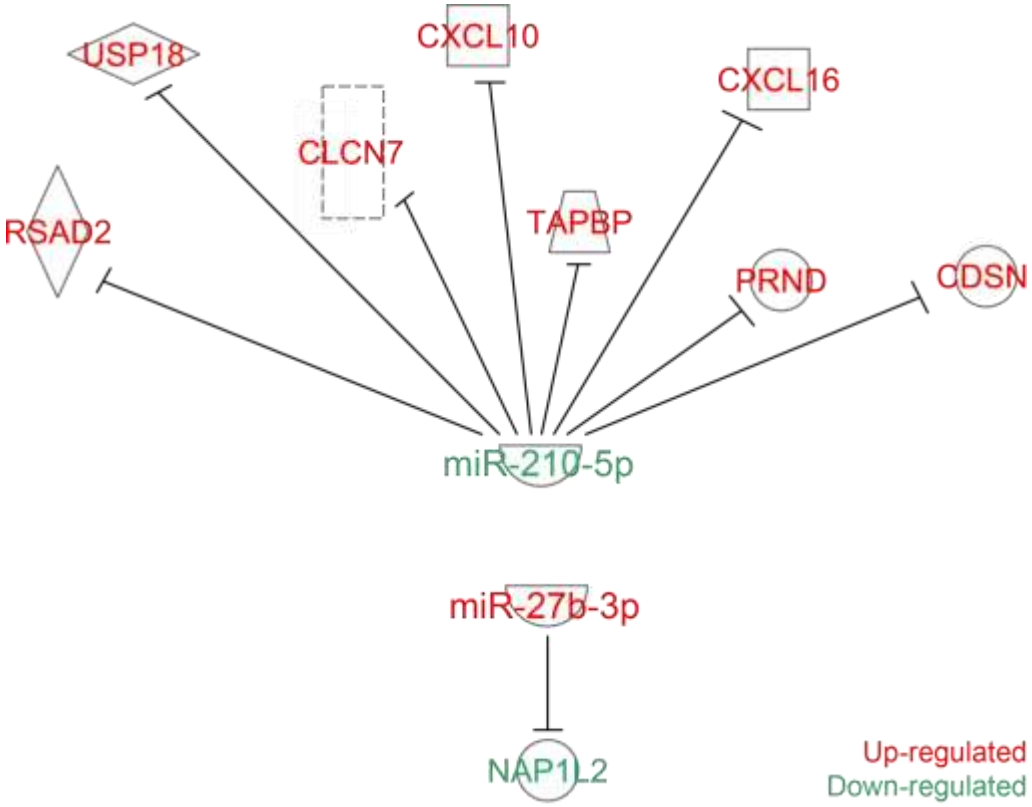


Figure 4-16: miRNAs and known targets exhibit expression changes following *Miwi2* expression

IPA MicroRNA Target Filter analysis of miRNAs with >1.5-fold expression changes and genes with >2-fold expression change in 4TO7-Miwi2 cells relative to 4TO7-GFP.

tRNA fragment expression changes following Miwi2 expression in 4TO7 cells

Our laboratory has previously described the association of processed tRNA fragments (tRFs) with the human Piwi, Hiwi2 in MDAMB231 breast cancer cells (Keam et al., 2014). In the total small RNA population in 4TO7 cells, 40 unique RNA species mapped to tRNA genes and together these made up nearly one third of all >1.5-fold ($p<0.05$) up-regulated sequences in 4TO7-Miwi2 cells (29%, Figure 4-13A). Of these up-regulated tRFs, 23% fall into previously annotated piRNA clusters. Most sequence reads are either 23nt or 30-33nt in size (Figure 4-17A). Figure 4-17B depicts the sequence alignment of tRFs and reveals an overrepresentation of G at positions 1 and 10. These tRFs preferentially map to five tRNA isotypes: GluCTC, GlyGCC, HisGTG, LysCTT and ValCAC (Figure 4-17C). Notably, one of these tRNA genes is located within a known piRNA cluster on Chromosome 10 (tRNA90-GluCTC). These are processed tRFs: visualisation of small RNA reads in WIG format using UCSC Genome Browser (GRCm38/mm10) indicates that these tRFs align with the 5' end of the full length tRNA (Figure 4-18). Intriguingly, I also observed a number of stacked reads immediately downstream of the 3' end of the ValCAC tRNA gene (Figure 4-18).

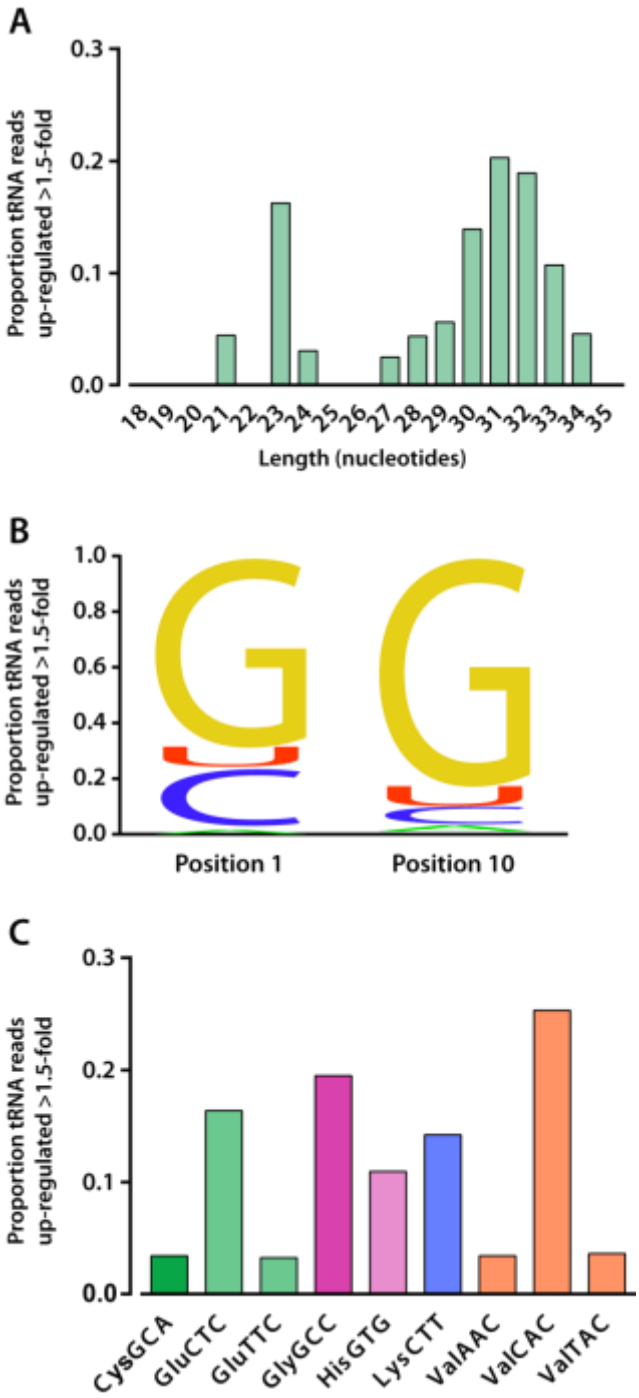


Figure 4-17: Characterisation of tRNA-derived small RNAs with increased expression in 4TO7-Miwi2 cells

(A) Length distribution, (B) sequence alignment and (C) annotations of tRNA-derived reads up-regulated >1.5-fold ($p<0.05$).

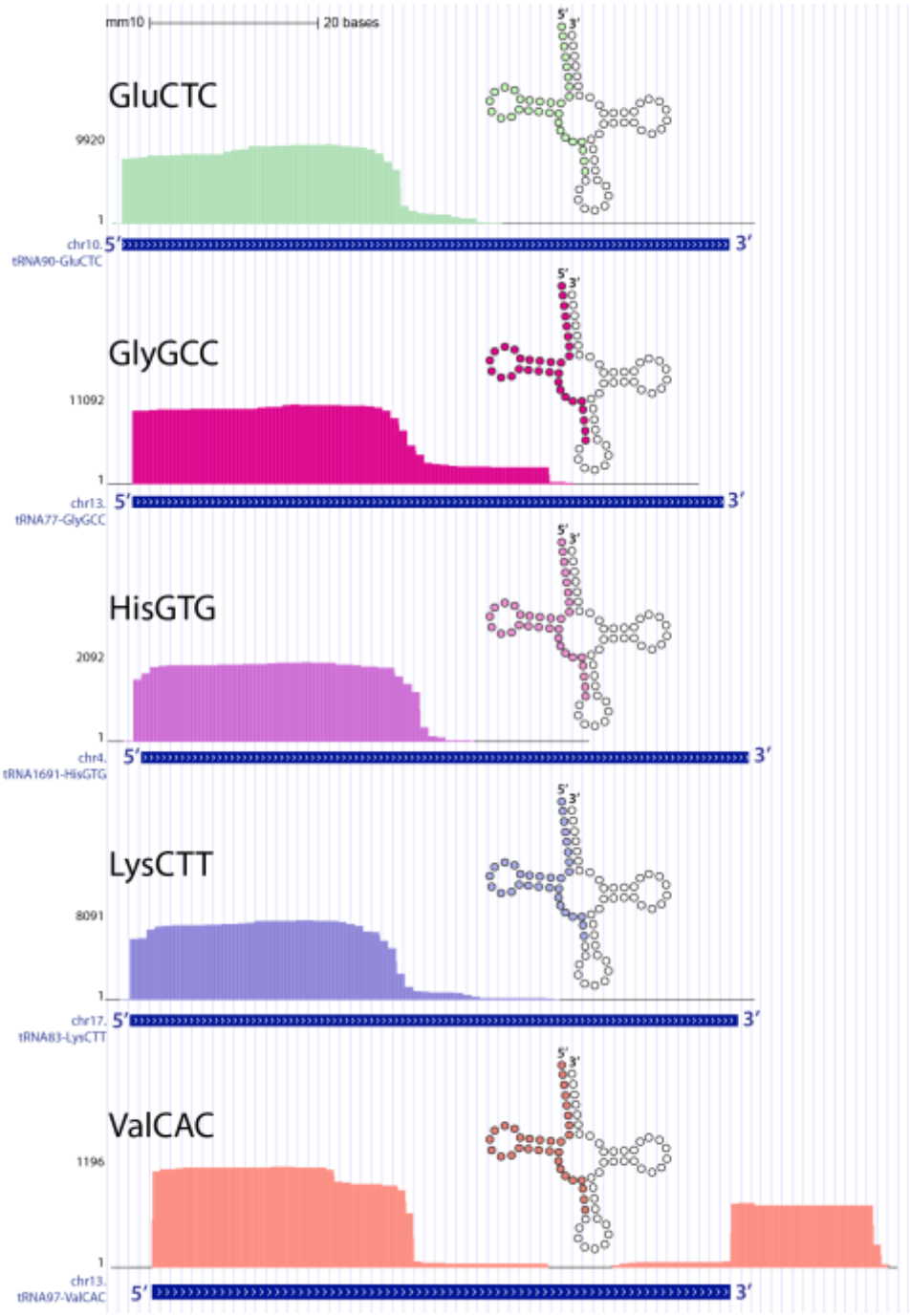


Figure 4-18: Positioning of tRNA reads on genomic loci

Graphical representation of tRNA from one replicate of 4TO7-Miwi2 reads in WIG format aligned with the mouse genome (UCSC Genome Browser, GRCm38/mm10). Secondary tRNA cloverleaf structures superimposed on browser image with coloured circles indicating the most abundantly represented bases in 4TO7-Miwi2.

Discussion

In this chapter I sought to determine whether the Piwi/piRNA pathway contributes to the stem-like properties of cancer cells using the 4T1 murine cancer cell line panel. I found that one murine Piwi gene, *Miwi2*, is expressed in many of the lines in the 4T1 panel. While *Miwi2* was most highly expressed in the most invasive and metastatic cell line, 4T1, there was an imperfect correlation with metastatic potential, and expression was only modestly enriched in putative tumour-stem cells. 4T1 cells resisted *Miwi2* knockdown, but ectopic expression of *Miwi2* in the non-metastatic 4TO7 cells induced a number of gene and miRNA expression changes. *Miwi2* expression also induced modest production of pi-sized RNAs, as well as small RNAs derived from tRNAs and unannotated regions of the genome. Togethe,r these data show that while Miwi2 is present in 4T1 cancer cells, and linked to regulatory small RNAs, it is unlikely to mark a tumour stem cell population.

Miwi2 expression in the 4T1 panel is variable and also mosaic within 4T1 clones

Similar to mouse ES cells, 4T1 cells express high levels of *Miwi2* but little to no *Miwi* or *Mili*. Miwi2 protein expression in 4T1 cancer cells is also similar to that of ES cells: expression is concentrated in distinct areas within the cytoplasm. Whether this localised expression is representative of Miwi2 enrichment in specific cell compartments or organelles is unclear, but could be resolved with dual-label immunostaining. While very strong in some cells, the expression of Miwi2 is mosaic: some cells within the clonal line exhibit weaker staining, or no staining at all. This might suggest mosaicism at a functional level and one could address this by separating the Miwi2-positive and Miwi2-negative populations for gene expression analysis and/or functional studies. This process would be technically challenging, given that Miwi2 is not expressed on the surface of the cell, but not impossible: new technologies such as the Merck Millipore Smartflare – which is endocytosed by the cell and fluoresces upon binding to its complementary mRNA *in situ* – could circumvent this issue, although this approach would not address protein expression directly.

Miwi2 expression levels also varied among lines within the 4T1 panel. While most highly expressed in 4T1 cells, it is also expressed in syngenic cell lines with more limited metastatic potential (albeit at a lower level). Only one of the four lines (4TO7) exhibited negligible *Miwi2* expression. This may provide some insight into potential roles for Miwi2 in cancer. Metastasis is a multistep process: invasion, intravasation into the bloodstream, survival during circulatory transport, extravasation into the parenchyma of a distant organ, and colonisation and growth (Fidler, 1978, Poste and Fidler, 1980). The period of time between extravasation and

colonisation is known as metastatic latency (Nguyen et al., 2009). All steps must be accomplished by a cancer cell before distant metastases are established and the 4T1 and 66cl4 cells are the only cell lines in the panel to be able to achieve this when injected into a mouse mammary fat pad (Aslakson and Miller, 1992). 4TO7 cells (which lack *Miwi2*) will migrate to distant locations after injection, but are unable to exit the latent phase and establish a secondary tumour (Aslakson and Miller, 1992). Many human cancer patients exhibit cancer cell dissemination without development of metastasis (Klein, 2003) and it would be interesting to examine these tumours for expression of Miwi2. If, like 4TO7, such tumours lack Miwi2, it would be tempting to speculate that Miwi2 is involved the processes governing entry into the final re-colonisation and growth phase of metastasis. However, some of the genes known to be required for this process, including interleukin-11, interleukin-6 and tumour necrosis factor-α (Nguyen et al., 2009), failed to show any change in response to *Miwi2* expression in 4TO7 cells. But metastasis is a complex process, and the involvement of key molecular players varies from tissue-to-tissue (Nguyen et al., 2009), so it is possible that some of the genes with increased expression in 4TO7-Miwi2 are involved in the growth phase of metastasis. In fact, it is possible that *Miwi2* is necessary – but not sufficient – for growth, and that it activates the expression of genes in the relevant pathways to mediate entry into the final phase of metastasis. Performing an *in vivo* mammary fat pad assay with *Miwi2*-expressing 4TO7 cells and observing their ability to grow a secondary tumour upon colonisation of a distant location could go some way to addressing this hypothesis.

Miwi2 expression is not enriched in putative tumour-initiating cells

Alexander Swarbrick's group at the Garvan Institute, Sydney, has observed a requirement and sufficiency for ID1 expression in 4T1 tumour propagation and metastasis; limiting dilution transplantation has revealed that $ID1^+$ cells are enriched for self-renewal, tumour-initiating and metastatic potential (A.S. personal correspondence). ID1-positive 4T1 cells showed only a slight enrichment of *Miwi2* expression, and in the ID1-negative population, no depletion (Figure 4-3 and Figure 4-4). This is inconsistent with the idea that Miwi2 is a CSC marker. While the expression of Miwi2 in 4T1 is mosaic, the majority of cells express some level of Miwi2. This is also inconsistent with a CSC marker as tumour-initiating cells within a cancer tend to exist in a quiescent state: they replicate slowly, exhibit chemoresistance, and emerge from dormancy for rapid expansion in response to stress (Visvader and Lindeman, 2008). Given that 4T1 cells undergo rapid replication, and Miwi2 is expressed in most of the cells, it is unlikely that it is enriched in a quiescent population. However, 4T1 is a highly invasive tumour cell line; it may be that the *majority* of cells within the population have stem-like characteristics with

the ability to contribute to tumorigenesis. It is also possible that many (if not all) cells within the 4T1 population exhibit cyclic expression of ID1 and other putative cancer stem cell markers, hence, any number of cells could possess the ability at various times to initiate a secondary tumour.

One limitation of this experiment is that only one biological replicate of $ID1^+$ capture was performed for each sample; only a small percentage of 4T1 cells are ID1-GFP^+ (<5%) hence limited material is available. Also, many markers for putative tumour-initiating cells in breast tumours have been reported, including CD24, CD44, CD90, CD133, and ALDH1 (Medema, 2013). Therefore, to address the hypothesis that *Miwi2* is expressed in bona fide tumour-initiating cells a much more comprehensive study of multiple CSC markers would be required.

4T1 cells resist doxycycline-mediated Miwi2 *knockdown*

I embarked on a similar approach to understanding Miwi2 function in 4T1 cancer cells as I took in embryonic stem cells – to knock down expression using α-*Miwi2* shRNA delivered by lentivirus. However, I was unable to achieve sufficient knockdown in the 4T1 despite using the same shRNAs that successfully knocked down *Miwi2* expression in ES cells. Somehow, the 4T1 cells were escaping *Miwi2* silencing despite being resistant to puromycin (which demonstrates that the cells do contain the shRNA). As the shRNAs are induced by doxycycline, I also explored whether blockade of P-glycoprotein would overcome the 4T1 resistance to Miwi2 knockdown. P-glycoprotein is a drug efflux pump which is involved in doxorubicin and doxycycline resistance in several breast cancer cell lines (Mealey et al., 2002, Bao et al., 2011), but despite its inhibition with verapamil (an L-type calcium channel blocker), *Miwi2* levels remained unchanged. This could be a technical limitation of the experiment, or could indicate that *Miwi2* expression is indispensable in 4T1 cells.

Miwi2 *expression in 4TO7 cells induces dramatic gene expression changes*

Consequently, I approached the functional study of Miwi2 in cancer cells in a different manner: by expressing it in the 4TO7 cancer cells, in which endogenous *Miwi2* is barely detectable. GFP-tagged Miwi2 was introduced into 4TO7 via lentiviral transduction. As a control I engineered GFP-expressing 4TO7 cells (without the Miwi2 ORF) to control for the presence of GFP and lentivirus.

The 4TO7 phenotype did not appear altered following ectopic *Miwi2* expression, but there were however some quite dramatic gene expression changes. These changes cluster in specific ontologies related to endocrine metabolism, and involve interferon signalling and

inflammation (Figure 4-10). Interferon activity is known to induce JAK/STAT signalling in cancer to promote cell growth, angiogenesis, recruit immune cells, and increase tumour aggression (Sansone and Bromberg, 2012). But interferon is also a hallmark of immune responses, so I considered whether the gene expression changes may have occurred simply as a result of exposure to lentiviral RNA. To illustrate this point, Bridge and colleagues exposed human lung fibroblasts to shRNA vector-containing lentiviral particles and observed a number of expression changes in interferon-related genes by microarray (Bridge et al., 2003). This group demonstrated that the shRNA, rather than the viral antigens, was responsible for the interferon response. But in my experiment both sets of cells (4TO7-Miwi2-GFP and the control 4TO7-GFP) were exposed to lentivirus, and in fact, the lentiviral titre was higher in control viral supernatant than Miwi2-GFP (owing to the smaller size of the control GFP vector). Also, exposure to the lentivirus occurred more than a fortnight before gene expression analysis: even *in vivo*, the interferon response resolves within 72 hours post-exposure to the dsRNA delivered by the lentivirus (Brown et al., 2007). Finally, the Miwi2 ORF and GFP ORF produce single stranded transcripts - not double-stranded – hence, they would not trigger a dsRNA-mediated response of this nature. Why ectopic *Miwi2* expression would result in gene expression changes in these pathways in 4TO7 cells is unknown and curious, but the fact that similar gene expression changes did not occur following exposure of ES cells to lentivirus (Section 3.4, Chapter 3) suggests that the induction of interferon-related genes in 4TO7-Miwi2 cells results specifically from exposure to *Miwi2*.

Further support for the specificity of *Miwi2*-induced changes in 4TO7 can be found in data from a study comparing transcript profiles of 4T1 cells to 67NR and 168FARN cells, which express *Miwi2* at lower levels (Tao et al., 2008). Tao et al identified 175 genes that were both i) up-regulated in 4T1 relative to the non-metastatic cell lines and ii) previously known to be involved in metastasis and/or tumorigenesis. Of these genes, 11 are also significantly up-regulated in 4TO7 cells following *Miwi2* expression (Table 4-2). Of particular note are the genes STAT1 and Usp18 that are up-regulated both my study and that of Tao and colleagues (Tao et al., 2008); both genes are involved in interferon signalling pathways. Traditionally, it was thought that STAT1 was involved in immune surveillance and acts a tumour suppressor (Dunn et al., 2006, Stephanou and Latchman, 2003), however others have reported that STAT1 expression is correlated with tumour invasiveness/metastasis, chemoresistance and tumour cell growth (Liu et al., 2013, Khodarev et al., 2012). Usp18 is an interferon-induced protein with a putative role in cancer via inhibition of apoptosis, promoting tumour aggression (Potu et al., 2010). *Miwi2*-induced Usp18 expression and that of another gene, Cxcl16, is matched by

a concurrent decrease in miR-210-5p (which targets these genes), suggesting a role for *Miwi2*-induced small RNA regulatory pathways to generate a more '4T1-like' gene expression profile in 4TO7-Miwi2 cells.

Some of the ontologies overrepresented in the 4TO7-Miwi2 gene set (Table 4-1) were in common with those reported by Tao and colleagues for 4T1 cells. This suggests that *Miwi2* expression might confer metastatic features to the otherwise non-metastatic 4TO7 cells. This is testable *in vivo* as 4TO7 cells can be used in a mouse mammary fat pad assay (Aslakson and Miller, 1992): the ability of transplanted cells to metastasise with and without Miwi2 expression would reveal whether Miwi2 does indeed confer a greater metastatic ability to cancer cells.

There is an increase in pi-sized RNAs following Miwi2 expression in 4TO7 cells

While *Miwi2* expression in 4TO7 caused many changes to small RNA profiles, perhaps the most obvious was the increase in 27-28nt small RNAs (Figure 3-7). This is particularly notable as Miwi2 typically binds 27-29nt piRNAs in the testis (Aravin et al., 2008). However, apart from length, most of the individual sequences that change significantly with Miwi2 expression do not exhibit other typical piRNA-like features. The sequence changes are diverse: there are few statistically significant changes in specific sequences, but this is not surprising as piRNAs are not abundant like microRNAs and, as a population, exhibit considerable sequence complexity. At the time of writing, 35,796 uniquely-mapping mouse piRNAs had been described (Sai Lakshmi and Agrawal, 2008) compared to 1908 mature mouse miRNAs (miRBase v20, http://www.mirbase.org/). Furthermore, 41% of all small RNAs were present as single reads, which could indicate either that the sequencing was not exhaustive, or that there are many low-abundance small RNAs in 4TO7 cells that could be piRNAs.

When considering all (including non-significant) up-regulated 27-28nt small RNAs, there was a slight preponderance to carry an adenine at position 10, which could imply the presence of a ping pong-like mechanism in 4TO7-Miwi2 cells. However, ping pong in 4TO7 cells would likely depend on the presence of Mili as it does in mouse germ cells (Aravin et al., 2008, De Fazio et al., 2011). But according to microarray, *Mili* is present at even lower levels than *Miwi2* in wild type 4TO7 cells. However, this does not rule out the possibility that, in the absence of Mili, ectopic Miwi2 acts in a homotypic piRNA amplification loop. Furthermore, at this point it remains unknown whether these pi-sized RNAs are bound by Miwi2 in cancer cells. This can be addressed by Miwi2 immunoprecipitation and deep sequencing of bound RNAs from i) 4TO7-Miwi2 cells and ii) 4T1 cells. The latter experiment is likely to have more biological relevance

than the former, as 4T1 cells naturally express Miwi2 protein. This experiment was performed and results presented in Chapter 5 of this thesis.

miRNAs dominate small RNA expression changes in Miwi2-4TO7 cells

As expected, miRNAs comprise the bulk of the total small RNA pool in 4TO7 cells (Figure 3-7), and they also make up the majority of sequences that are down-regulated in response to Miwi2 expression (Figure 4-13). This implies some relationship between *Miwi2* and miRNA expression. In 4TO7-Miwi2 cells, only a small number of miRNAs (12 up-regulated and 9 down-regulated) exhibited statistically significant expression changes, but there was a global decrease in the proportion of miRNAs (84.5% of reads in controls and 74.3% in 4TO7-Miwi2), implying perturbations in many miRNAs in response to *Miwi2*. Some of the miRNAs with the most dramatic expression changes with ectopic *Miwi2* are involved in focal adhesion pathways. Focal adhesions (FAs) form when cytoplasmic clusters of integrins recruit scaffold and signalling proteins to the inner surface of the plasma membrane, adhering the cell to its extracellular matrix (ECM) (Galbraith and Sheetz, 1998, Calderwood et al., 2000, Petit and Thiery, 2000). During cell migration, FA-mediated attachment to the ECM generates the forces required for movement. Disassembly of the FA-complexes releases the attachment as the cell progresses (Webb et al., 2002). This process is a fundamental component of tumour metastasis (Schlaepfer et al., 2004, Parsons et al., 2000, Hsia et al., 2003).

Pathways analysis demonstrated that genes targeted by the *Miwi2*-altered miRNAs are implicated in several different cancer types; this could suggest that *Miwi2* has some universal function in cancer. This is supported by reports of Piwi gene and protein expression in multiple human cancers such as breast cancer (Liu et al., 2010), colon cancer (Li et al., 2010), seminoma (Qiao et al., 2002), gastric cancer (Liu et al., 2006), epithelial ovarian cancer (Chen et al., 2013), glioma (Sun et al., 2011), pancreatic cancer (Grochola et al., 2008), oesophageal cancer (He et al., 2009), and soft-tissue sarcoma (Taubert et al., 2007). Furthermore, some of the miRNAs most affected by ectopic *Miwi2* have targets within crucial developmental signalling pathways that are often dysregulated in cancer, such as ErbB (Yarden and Pines, 2012), PI3K-Akt (Wong et al., 2010), mTOR (Laplante and Sabatini, 2012), TGF-beta (Katsuno et al., 2013), and the interacting MAPK and Wnt pathways (Biechele et al., 2012, Jeong et al., 2012). However, pathway analysis of the gene expression changes in 4TO7-Miwi2 does not overlap with those predicted from the mirPath analysis. It is possible that some of the miRNA-regulated changes were subtle, or obscured by larger gene expression changes in the interferon-related genes. It is also possible that miRNAs are not having a pronounced effect on the transcript level, as they

are well known to act in translational repression, which can occur quite independently of transcript degradation. Alternatively, these miRNAs expression changes could form part of a secondary response to compensate for the dramatic shifts in gene expression in 4TO7-Miwi2. If Miwi2 uses piRNAs as site-specific guides to mediate transcriptional or post-transcriptional gene silencing, target prediction of Miwi2-bound piRNAs would go some way to answering this question.

There were several individual miRNAs of note that exhibited statistically significant expression changes in response to *Miwi2*, such as that within the intronic *Sfmbt2* miRNA cluster. This miRNA family is expressed in embryonic stem cells and the placenta and one of its clustered mature miRNAs (miR-467a) is thought to promote growth and survival of mouse embryonic stem cells (Zheng et al., 2011). miR-467a-3p and other members of the *Sfmbt2* cluster (miR-466b-3p, miR-466c-5p, miR-467d-3p, and miR-669a-3p) are up-regulated in 4TO7-Miwi2 cells (Table 4-5), suggesting that *Miwi2* expression could impart miRNA-mediated, stem-like growth properties to the cell. The concomitant increase in the angiogenic miR-27b-3p and decrease in the expression of its target gene *Nap1l2* could also indicate a role for *Miwi2* in maintenance of a stem-like state: *Nap1l2* promotes histone acetylation during neuronal differentiation and its loss results in defects in differentiation and aberrant maintenance of the neural stem cell state (Attia et al., 2007). Interestingly, the increase in miR-27b-3p expression is accompanied by a decrease in expression of another microRNA, miR-27a-5p, which derives from the same precursor RNA (i.e. polycistron). But divergent expression of individual miRNAs from polycistronic primary transcripts, mediated by adenosine deaminase RNA editing enzymes, has been documented in *Drosophila* (Chawla and Sokol, 2014). It is possible that this process is also occurring in 4TO7 in response to ectopic *Miwi2*.

From a technical perspective, there was a greater number of miRNA expression changes in 4TO7-Miwi2 (relative to the control) revealed by sequencing than microarray. This is unsurprising, given that next-generation sequencing is more sensitive than microarray (Malone and Oliver, 2011, Zhao et al., 2014). But an unexpected confounder is that sequencing failed to validate the changes in miRNAs as revealed by microarray, for example the decrease in levels of let-7c in 4TO7-Miwi2. This inconsistency could be explained by the existence of miRNA isomers or "isomiRs" (Zhou et al., 2012). The bioinformatic analysis that we performed allows for a conservative wobble in either 5p or 3p miRNA annotations. Hence, isomiRs that differ from the miRBase entry by i) up to 3 bases at the 5' end and ii) up to 10 bases downstream of the 3' end are given the same annotation. Hence, a statistically significant change in a unique

miRNA may be masked by non-significant changes in its isomiRs. In addition, microarray probe sets for microRNAs detect mature, pre- and pri-miR expression, whereas small RNAs are size-selected for sequencing, excluding precursors.

Many up-regulated small RNAs are derived from unannotated regions of the genome

The biological significance of increases in small RNAs mapping to unannotated regions of the genome (Figure 4-12 and Figure 4-13) is currently unclear as the function of these small RNAs is unknown. They do not exhibit obvious sequence features characteristic of ping pong biogenesis – such as an overrepresentation of 1U or 10A – and do not fall into known piRNA clusters. However, given that cytoplasmic Miwi2 is likely to have a different function in cancer cells than nuclear Miwi2 in the testis, the absence of these canonical piRNA-like sequence features does not exclude the possibility that these RNAs are Miwi2-dependent. Without RNA immunoprecipitation of Miwi2-GFP from 4TO7 cells, I am unable to address the possibility that these small RNAs are bound to Miwi2. But given that *Miwi2*-expressing 4TO7 cells adopt a partial 4T1-like gene expression profile, it is possible that there is some overlap with these unannotated small RNAs and those Miwi2-bound small RNAs in 4T1 cells; this will be investigated in Chapter 5 of this thesis.

5' tRNA fragments are up-regulated in 4TO7 cells in response to Miwi2 expression

tRNA-derived small RNAs are up-regulated in 4TO7-Miwi2. This is intriguing, as our laboratory has demonstrated that the human Piwi orthologue of Miwi2, Hiwi2, associates with processed tRNA fragments in MDAMB231 breast cancer cells (Keam et al., 2014). The presence of tRFs in cancer cells is not a novel finding: Dicer-dependent 19nt 5' tRFs derived from tRNA-Lys, tRNA-Val, tRNA-Gln and tRNA-Arg in HeLa cervical cancer cells were the first tRFs to be described and are in comparable abundance to the five most highly expressed miRNAs: miR-21, let-7-f1, let-7a-1, let-7c, and let-7f-2 (Cole et al., 2009). Furthermore, 3' tRFs are more highly expressed in cancer cells than normal tissues and are involved in gene regulation to promote prostate cancer cell division (Lee et al., 2009). However, not all tRFs promote a cancer cell phenotype: CU126 is a Dicer-dependent tRF expressed in human B cells and down-regulated in lymphoma cells (Maute et al., 2013). CU126 is an unusual tRF in that it behaves as a miRNA (as it associates with all four human Argonaute proteins) and adopts a tumour suppressor role.

tRNA cleavage can also occur under stressful conditions, such as oxidative or heat stress; this highly conserved process produces half-tRNA fragments where cleavage occurs at the anticodon loop (Thompson et al., 2008, Fu et al., 2009). The presence of oxidative stress-induced tRNA halves in 4TO7 cells would not be unexpected, as cancer cells tend to have

elevated levels of oxidative stress relative to normal tissue (Schumacker, 2006). However, it is unlikely that the tRFs I observe to be up-regulated with Miwi2 expression are stress-induced tRNA halves as the 3' ends of the tRF do not terminate within the tRNA anticodon loop (Figure 4-18). However, it is possible that some of the longer reads in the WIG stacks represent stress-induced tRNA halves, particularly those that map to GlyGCC (Figure 4-18).

The presence of 30-33nt tRNA halves has been observed in mouse serum; these mostly align with valine, glycine and glutamine tRNA genes (Dhahbi et al., 2013). However, tRNA fragments had not been reported in murine cancer at the time of writing this thesis. The specific tRFs up-regulated in *Miwi2*-expressing 4TO7 cells are also present in control cells but are simply less abundant. Given that tRFs regulate gene expression by physical association with polysomes in human cells (Sobala and Hutvagner, 2013), these fragments could be up-regulated in response to *Miwi2* for the fine-tuning of translation. An intriguing possibility is that tRFs physically associate with Miwi2. This speculation is supported by other studies: Couvillion et al (Couvillion et al., 2010) reported the binding of tRNA fragments with the *Tetrahymena thermophilus* Piwi protein, Twi12. Our laboratory has shown that 5' tRFs bind cytoplasmic Hiwi2 in MDAMB231 human breast cancer cells and that Hiwi2 is physically associated with actively translating ribosomes (Keam et al., 2014). Furthermore, in this study we showed that Hiwi2 also binds piRNAs derived from the sense strand of active genes independent of the parent gene transcript levels, pointing to a role for Piwi proteins in the regulation of translation, although this mechanism is yet to be elucidated. In my dataset, I observed increased expression of 28nt small RNAs derived from tRNA-GluCTC in response to *Miwi2* expression in 4TO7 cells; these also align with a known piRNA cluster and could be *bona fide* piRNAs. Whether or not Miwi2 binds tRF-piRNAs in cancer cells will be addressed in Chapter 5.

Conclusions

In this chapter, I have confirmed Miwi2 transcript and protein expression in highly metastatic murine carcinoma cells and variable expression in its syngenic sister lines. The relationship between Miwi2 and propensity for metastasis remains to be elucidated. 4T1 cells resisted *Miwi2* knockdown, hence I expressed the transcript in non-metastatic 4TO7 cells with low-level *Miwi2* expression. A number of the genes up-regulated in 4TO7 cells following *Miwi2* expression are involved in physiological processes such as cellular movement, growth, and proliferation, as well as cell-to-cell signalling and interaction; in particular, expression of a number of genes involved in interferon signalling are perturbed. Given the role of interferon/STAT signalling in cancer metastasis it is conceivable that aberrant expression of *Miwi2* confers some metastatic properties to the cell. However, without *in vivo* studies of 4TO7-Miwi2 cell behaviour following mammary fat pad transplantation, it is impossible to draw conclusions regarding the role of Miwi2 in tumorigenesis and metastasis. Furthermore, it is possible that interferon-related gene expression changes are merely a response to viral exposure, but I have argued against this possibility. Following *Miwi2* expression, I observed changes in the small RNA repertoire of 4TO7 cells including: an increase in pi-sized RNAs; up-regulation of miRNAs that target genes in cancer-related pathways; and an increase in small RNAs that align with unannotated regions of the genome. The function of many of the RNAs identified is unknown. In this chapter I have also revealed the existence of 5' tRFs in mouse cancer cells; these small RNAs are up-regulated in response to *Miwi2* expression and could be linked to translation. The question remains whether or not these tRFs bind Piwi protein as is the case in *Tetrahymena thermophilus* and human MDAMB231 breast cancer cells. In Chapter 5 of this thesis, I will characterise the Miwi2-bound piRNAs in metastatic 4T1 cells. Together with the functional evidence I have presented in this chapter, these studies will provide further clues as to the role of Miwi2 in cancer cells. Given that this mammary carcinoma model emulates stage IV human breast cancer, it will be of great therapeutic benefit to delineate any of the mechanisms responsible for metastatic progression and tumour invasion.

5. Miwi2-bound small RNAs in embryonic stem and cancer cells

Introduction

I had now established that Miwi2 is present, but not essential, in mouse ES cells, but the means by which the protein performs its function, if any, in these cells is unknown. I have also shown that Miwi2 is present in 4T1 breast cancer cells, although its role in cancer and metastasis, at this stage, is unclear. The identification and analysis of *bona fide* piRNAs has provided major insights into effector Piwi protein function in the germ cells in complex organisms (Brennecke et al., 2007, Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008). However, studies of somatic piRNAs are comparatively limited. A diverse population of pi-like RNAs has been identified in the somatic stem cells of the planarian *Schmidtea mediterranea*; these 32nt RNAs map to hundreds of thousands of loci containing repetitive sequences, and their expression is dependent on the *S. mediterranea* Piwi proteins, Smedwi-1 and Smedwi-2 (Palakodeti, 2008). Furthermore, somatic piRNAs are responsible for memory formation in the neurons of *Aplysia*, where they mediate epigenetic silencing of *Creb2* via an interaction with *Aplysia* Piwi (Rajasethupathy et al., 2012). Somatic piRNAs have also been cloned from some tissues of more complex organisms, including *Drosophila,* rhesus macaque and the mouse (Ghildiyal et al., 2008, Malone et al., 2009, Li et al., 2009, Yan et al., 2011, Lee et al., 2011). In addition, piRNAs have been reported in a number of human cancer cell lines and solid tumours (Lu et al., 2010, Cheng et al., 2011). At the time of writing this thesis, the only report of Piwi protein-bound piRNAs in murine somatic cells was of Miwi-bound somatic piRNAs in the brain (Lee et al., 2011). This particular study found that the most highly expressed piRNAs in the mouse hippocampus are intergenic and exhibit significant sequence diversity. To date, characterisation of Miwi2-bound piRNAs has only been performed in male germ cells, where it associates with pre-pachytene piRNAs derived largely from repeat elements. It remained to elucidate whether Miwi2 binds small RNAs in ES and 4T1 cells and if so, to characterise this population to provide clues as to the role of the Piwi/piRNA pathway in these cell types.

Hence, the specific aims of this chapter are to:

a) Synthesise antibodies capable of immunoprecipitating Miwi2 protein and associated RNAs

b) Perform next-generation sequencing of Miwi2-bound small RNAs in ES cells and 4T1 cancer cells to characterise their sequence features and genomic origins.

In collaboration with Abmart, Shanghai, I developed a panel of anti-Miwi2 antibodies and using these, optimised RNA-immunoprecipitation (RIP) of Miwi2-piRNAs from mouse ES cells and 4T1 breast cancer cells. Bound piRNAs were isolated and subjected to library amplification for next-generation sequencing. I investigated piRNA sequence features and genomic alignments and compared these to publicly-available datasets in an effort to gain insight into Miwi2 function in embryonic stem cells and cancer cells.

Results

5.1 Production and screening of monoclonal antibodies for Miwi2 immunoprecipitation

Currently, there are no commercially-available antibodies validated for immunoprecipitation (IP) of Miwi2, and attempts to immunoprecipitate Miwi2 from neonatal testis using a rabbit polyclonal α-Miwi2 antibody (Abcam ab21869) were unsuccessful, despite the ability of this antibody to recognise Miwi2 by Western blot and immunostaining. In collaboration with Abmart, Shanghai, China, I developed monoclonal antibodies and then screened the antibodies for their capacity to immunoprecipitate Miwi2 and its bound piRNAs. Peptide antigens were developed prior to immunisation using Surface Epitope Antibody Library (SEAL™) technology; to be selected, antigens were not to exhibit stable secondary or tertiary protein structure in the folded state (thus are in the same conformation as the native protein). Hence, native (folded) protein may be recognised by the antibodies generated by antigenic exposure. This is an important consideration when developing an antibody for RNA-immunoprecipitation: a denaturing pull-down eliminates the RNA:protein interactions so it is essential to maintain the protein in its native state. The monoclonal antibody synthesis process is detailed in Section 2.8.1, Materials and Methods. In brief, sequences for multiple SEAL™ antigens and a number of potent adjuvants were encoded on a DNA vector and, once transcribed and translated, were expressed on the surface of the resulting immunogenic recombinant protein. Following immunisation with the peptide antigen, mouse spleen cells were fused with immortalised myeloma cells to generate cell lines with an infinite capacity for clonal generation of antibody. The fused hybridoma cells were injected into the peritoneal cavity of mice, resulting in tumours that secrete antibody-rich ascites fluid for use in immunodetection.

Testing the mouse monoclonal antibody panel for immunodetection of Miwi2

I received 12 ascites samples containing monoclonal antibodies targeting multiple peptide sequences within Miwi2. To confirm the specificity of each antibody, I tested the capability of each ascites fluid sample generated from individual hybridoma clones to detect Miwi2 on a western blot. For western blot, I loaded protein from HEK293FT wild type and HEK293FT-

Miwi2-FLAG cells. I had previously shown high Miwi2 expression in HEK293FT-Miwi2-FLAG cells using a rabbit polyclonal α-Miwi2 antibody (Abcam ab21869, Figure 3-3, Chapter 3). Of the 12 clones, only two exhibited a relatively clear band at the expected size of Miwi2 in the HEK293FT-Miwi2-FLAG sample (Figure 5-1A).

The two clones (IF5-2 and 1A10-2) were derived from immunisation with a single peptide (KPSEPVEIPQ). Given that I required the antibody to immunoprecipitate native Miwi2, it was necessary to ensure this particular sequence was not buried within the protein in its folded, RNA-binding conformation. At the time of experimentation there was no publicly available crystal structure data for Miwi2. Therefore, in collaboration with Dr Alastair Stewart from the VCCRI crystallography laboratory, I created a homology model of Miwi2 by aligning the peptide sequence to a structurally known homolog as described in Section 2.8.4, Materials and Methods. In brief, we performed a NCBI protein BLAST of the Miwi2 peptide sequence against all known protein structures derived from X-ray Crystallography, nuclear magnetic resonance spectrometry and electron microscopy contained within the Protein Data Bank (PDB). Human Argonaute 2 in complex with RNA was the best available candidate to act as a Miwi2 surrogate (Schirle and MacRae, 2012), with 86% coverage of and 24% identity to the Miwi2 protein sequence. We aligned the Ago2 and Miwi2 sequences with ClustalW2 and replaced the Ago2 sequence with the aligned Miwi2 sequence, predicting secondary and tertiary structure of unconserved residues using particular sequence features. We mapped the alignment onto the structure using CHAINSAW and visualised the resulting structure in multiple orientations using Pymol. The analysis predicted that the target peptide sequence was expressed on the periphery of the predicted protein structure (Figure 5-1B), therefore it could be expected that these antibodies had the capacity to immunoprecipitate Miwi2 in its native form in complex with RNA.

I then performed a trial IP of Miwi2 from HEK293FT-Miwi2-FLAG cells using IF5-2 and 1A10-2 ascites fluid. To give the antibodies the best possible chance of success, I used denaturing conditions as it was not necessary to preserve protein:RNA interactions for the trial. Eluted fractions were separated using SDS-PAGE and immunoblotted with α-FLAG monoclonal antibody (Figure 5-1C, top panel) to evaluate IP efficiency. Of note, multiple isoforms of Miwi2-FLAG are expressed by the Miwi2-FLAG vector - the Miwi2 open reading frame (ORF) encoded by the vector contains two potential start codons at the 5’ end. The FLAG signal is at the 3’ end of the ORF, therefore the α-FLAG monoclonal antibody will bind to the C-terminus of both Miwi2 isoforms in a Western Blot or IP (resulting in a double band on the membrane). There is

a Miwi2-FLAG signal at 98kDa in the elution for 1A10-2 but little depletion of Miwi2-FLAG from IP supernatant relative to input; this result suggested that 1A10-2 is able to immunoprecipitate Miwi2, although at relatively low efficiency. There is no 98kDa band is visible in the IF5-2 elution fraction but despite this, I confirmed that protein in both the IF5-2 and 1A10-2 elution fractions is, in fact, Miwi2 by Liquid Chromatography-Mass Spectrometry (LC-MS, Figure 5-1D). Samples were processed for LC-MS at the Bioanalytical Mass Spectrometry Facility, University of New South Wales.

I obtained the IF5-2 and 1A10-2 hybridoma cell lines from Abmart and weaned hybridoma cells onto IgG-free CD Hybridoma Media (GIBCO, Life Technologies). I allowed acclimatised clones to expand, overgrow, and apoptose. I collected antibody-containing supernatant and used Protein G to purify α-Miwi2 monoclonal IgG antibody. Figure 5-1E shows 1µg crude antibody run on SDS-PAGE from ascites and affinity-purified samples. Ascites fluid samples contain large contaminating bands, presumably from serum albumin (62kDa). I used ELISA to confirm antibody isotype, and deduced that 1A10-2 is IgG2b. IF5-2 hybridoma cell line contains a mixture of IgG1 and IgG3, indicated by multiple light chain bands (Figure 5-1E). Affinity-purified (A-P) 1F5-2 and 1A10-2 were able to more effectively immunoprecipitate Miwi2-FLAG from transiently-transfected HEK293FT cells than ascites fluid (Figure 5-1C, lower panel).

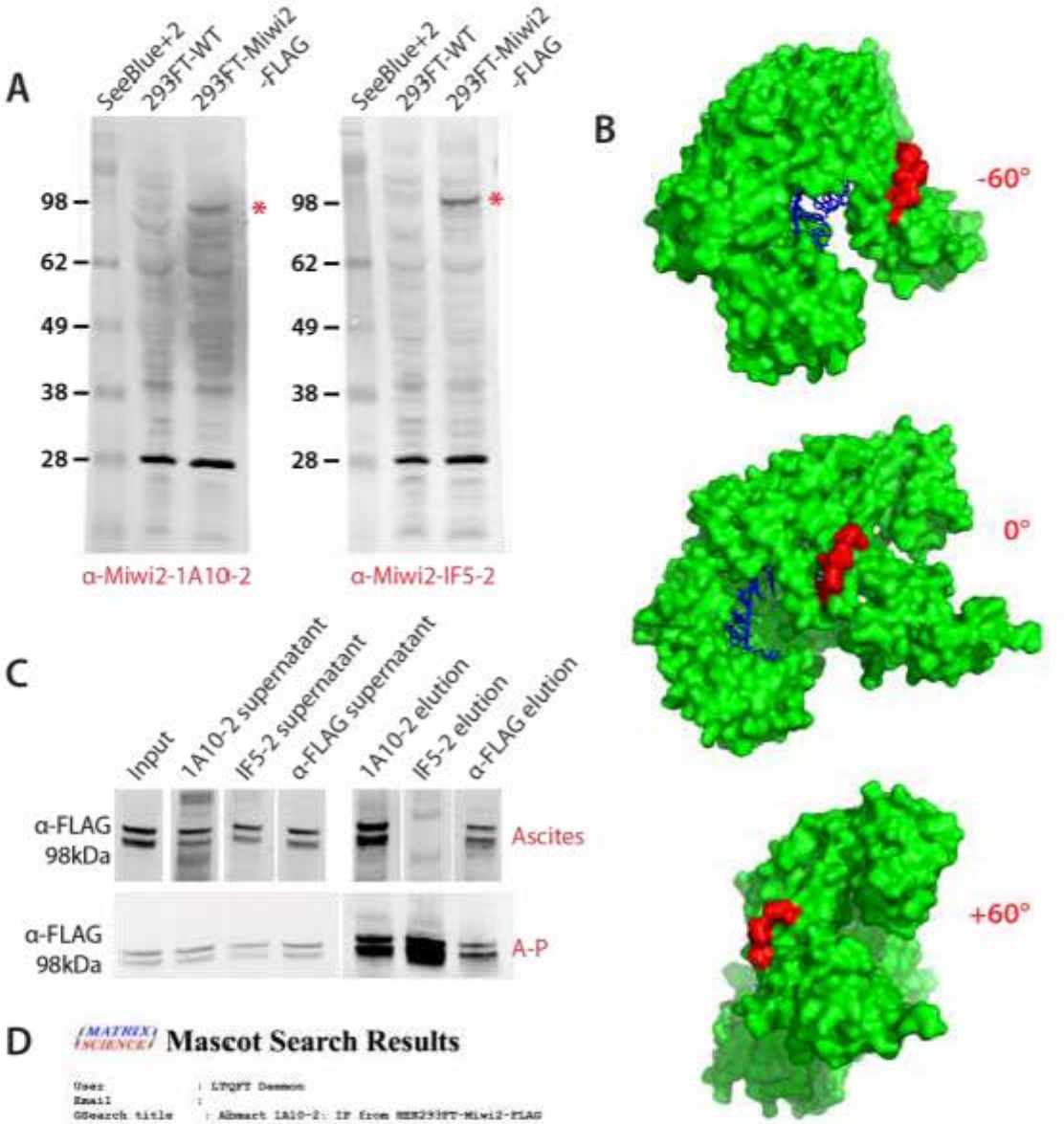

A
SeeBlue+2
293FT-WT
293FT-Miwi2 -FLAG
98
62
49
38
28
α-Miwi2-1A10-2
α-Miwi2-IF5-2
B
-60°
0°
+60°
C
Input
1A10-2 supernatant
IF5-2 supernatant
α-FLAG supernatant
1A10-2 elution
IF5-2 elution
α-FLAG elution
α-FLAG 98kDa
Ascites
A-P
D
Mascot Search Results
Mascot Search Results
Mascot Search Results
E
Seeblue+2
Ascites 1A10-2
Ascites IF5-2
A-P 1A10-2
A-P IF5-2
kDa
62
49
38
28
H-chain
L-chain

Figure 5-1: Candidate Selection for Miwi2 antibody

(A) Western Blot: 1A10-2, IF5-2 ascites fluid. Miwi2-sized band (98kDa) is indicated by red asterisk. (B) 180° view of Miwi2 homology model (green) in complex with small RNA (blue) plus peptide binding site (red) visualised with Pymol. In each image, the protein is rotated by 60°C around a vertical axis. (C) Ascites fluid (top panel, "Ascites") and affinity-purified (lower panel, "A-P") antibody IP validation in HEK293FT-Miwi2-FLAG. Input, IP supernatant and heat elution fractions were immunoblotted with α-FLAG (Sigma F1804). (D) Liquid Chromatography Mass Spectrometry (LC-MS) identification of Miwi2 present in IP fraction. (E) 1µg crude antibody separated by SDS-PAGE from ascites ("Ascites") and affinity-purified ("A-P") 1A10-2 and IF5-2 samples. Gel was stained with SYPRO Ruby and visualised at 492nm using a Fuji imager.

5.2 Analysis of total and Miwi2-bound small RNAs in R1 mouse embryonic stem cells

By using FLAG-tagged Miwi2, I was able to unequivocally demonstrate that the anti-Miwi2 monoclonal antibodies are capable of immunoprecipitating Miwi2; this gave me permission to use these antibodies to capture endogenous Miwi2 from mouse ES cells. I harvested protein lysate from pluripotent R1 ES cells at passage 23 and performed Miwi2 RNA-immunoprecipitation in duplicate using 15mg of lysate and both 1A10-2 and IF5-2 anti-Miwi2 antibodies as described in Section 2.10.2, Materials and Methods. A negative control IP was performed in parallel using a non-specific isotype-matched (IgG) antibody. Following overnight IP, Miwi2-bound RNAs were isolated and small RNA libraries prepared for next-generation sequencing according to the method described in Section 2.11.1, Materials and Methods.

Bioinformatic processing of small RNA sequencing data

I performed Illumina HiSeq® next-generation sequencing of unenriched small RNA from ES cells ("total small RNA") and small RNAs enriched by Miwi2 immunoprecipitation from ES cells ("Miwi2-bound small RNA"). The unenriched fraction was used for comparison to identify small RNAs that had been specifically enriched by the IP. All data for Miwi2-bound small RNAs in this section are averaged from biological duplicates. Bioinformatics analysis was performed with Paul Young (Suter Laboratory, VCCRI). Following quality filtering (see Section 2.11.2, Materials and Methods), I had approximately 4.6 million reads for the total small RNA sample and a total of 5.5 million reads for my Miwi2-bound small RNA samples (Table 8-1, Appendix). Of those reads, the majority mapped successfully to the mouse genome when allowing 1 mismatch (91% for total small RNA; an average of 54.5% for Miwi2-bound small RNA). For the total small

RNA sample, I obtained 162,712 unique sequences and for the Miwi2-bound small RNA duplicate sequencing runs, I observed a total of 499,316 unique sequences. My sequencing dataset included many single copy reads (74% for total small RNA; and an average of 82% for Miwi2-bound RNA samples), indicating that the small RNA sequencing was not exhaustive. Small RNA expression for each sample was normalised to the total number of reads within each sample.

Basic characterisation of total and Miwi2-bound small RNAs

I examined length distribution and sequence annotations of total and Miwi2-bound small RNAs in ES cells; these are shown in Figure 5-2. The dominant length of unenriched small RNAs is 21-24nt, and most of these are annotated as miRNAs (Figure 5-2A). Unsurprisingly, miRNAs comprise the vast majority of annotated total small RNA in ES cells (88%, Figure 5-2B). Of the genic reads, most are derived from introns (78%). Of the small RNAs mapping to repeats, they are most frequently produced from LTR retrotransposons (63%) and LINE repeats (24%, Figure 5-2B). The overwhelming majority of all annotated small RNA reads are sense to annotated sequence elements (Figure 5-2C).

The Miwi2-bound small RNA profile in ES cells is quite distinct to the unenriched total small RNA population. Firstly, the length distribution is very different: while there is a tendency for Miwi2 to bind short (19-24nt) RNAs in ES cells, the IP also enriched for small RNAs of up to 34nt in length (Figure 5-2D). Most of the Miwi2-bound 19-34nt small RNAs are derived from repeat elements (50%) or genes (44%, Figure 5-2D and Figure 5-2E), and a significantly larger proportion of reads map to unannotated regions in the IP than total small RNA – this is significant as these RNAs may be derived from intergenic piRNA clusters that have been described in the pachytene testis (Gan et al., 2011). Of the gene-derived Miwi2-bound small RNAs, approximately half map to introns, one quarter to exons, one fifth to 3' untranslated regions (UTRs) and one tenth to 5'UTRs (Figure 5-2E). Most of the repeat-associated small RNAs bound to Miwi2 in ES cells originate from LTR retrotransposons (75%), and a smaller percentage from LINE repeats (13%); the 11% produced from rRNA could represent some degradation in my samples. As is the case in the total small RNA population, the overwhelming majority of annotated Miwi2-bound small RNAs align with the sense strand of annotated sequence elements (Figure 5-2F).

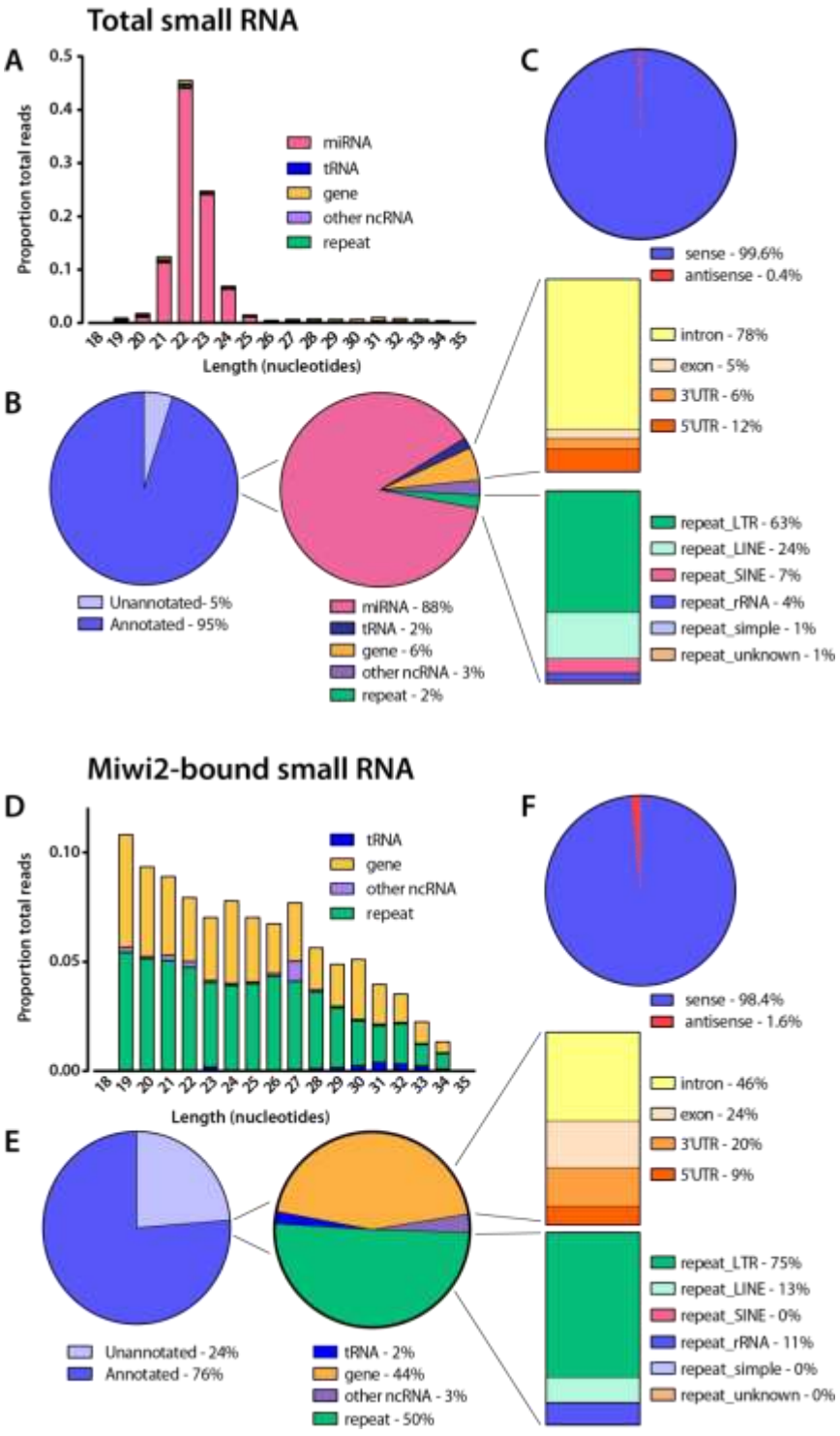


Figure 5-2: Basic sequencing statistics for total and Miwi2-bound small RNAs in ES cells

Small RNA sequence characteristics of (A-C) total small RNA and (D-F) Miwi2-bound small RNA. (A, D) Length distribution of small RNAs between 18-35nt with annotations. Values represented as a proportion of total small RNA. (B, E) Categorical annotations of small RNAs. (C, F) Sense and antisense sequences with respect to annotated sequence features as a proportion of all annotated small RNA reads.

Nucleotide bias in small RNAs from total ES cells and Miwi2-IP

To investigate whether the Miwi2-bound small RNAs exhibit piRNA-like sequence features such as 1U or 10A (indicative of ping pong biogenesis), I determined the overall nucleotide composition of each read to observe any sequence bias. At position 1, the total small RNA cohort in ES cells exhibits an over-representation of adenine and an under-representation of guanine and cytosine (Figure 5-3A). This bias for a 5'A can be explained by the fact that the 12 most abundant sequences in unenriched ES cells are miRNAs with a 5'A, which occupy 48% of total small RNA reads. There is no overt sequence bias at position 10 for unenriched small RNAs. Miwi2-bound small RNAs do tend to display 1U (Figure 5-3B), as well as those that are 27-29nt (Figure 5-3C) but do not favour 10A (Figure 5-3B and Figure 5-3C). Thus, the Miwi2-bound fraction in ES cells exhibits a bias towards primary – but not secondary – piRNA sequence features.

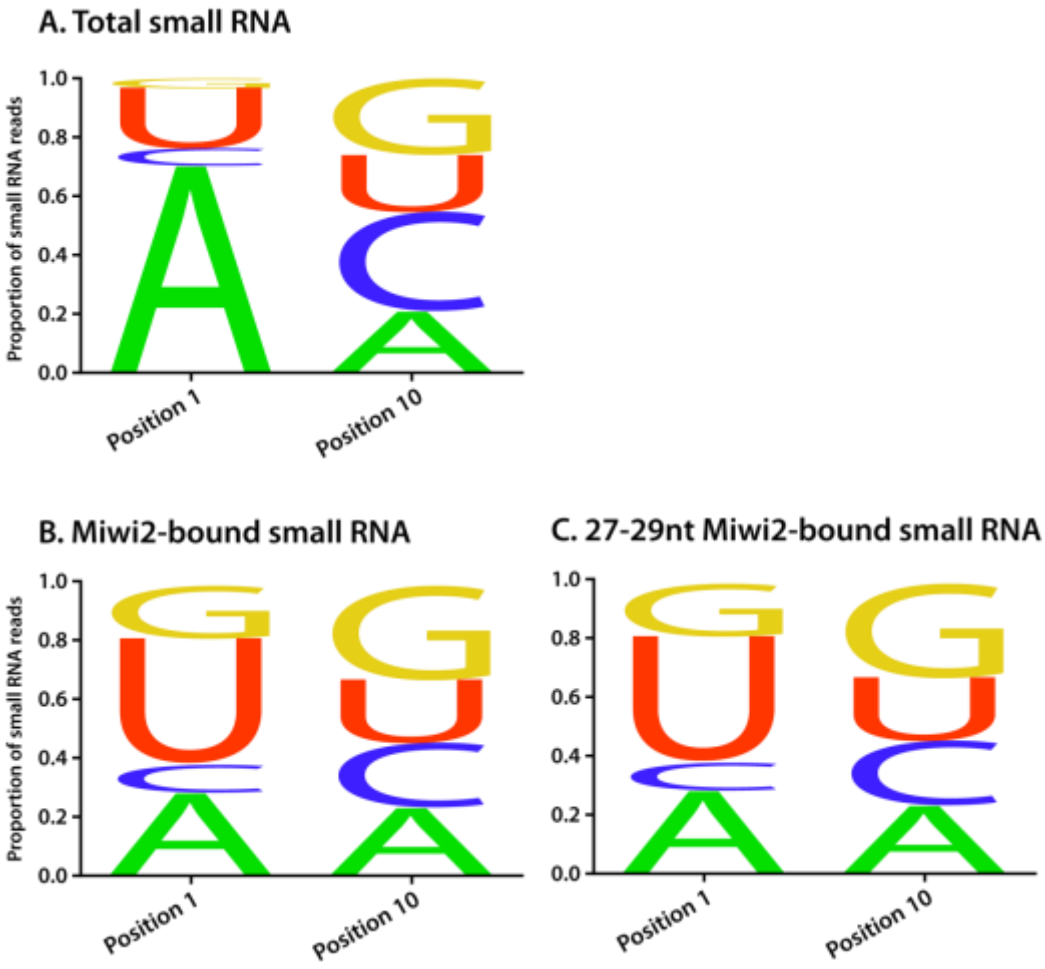


Figure 5-3: Sequence alignment of small RNAs in ES cells

Sequence alignment at positions 1 and 10 of (A) total small RNA population from whole ES cells and (B) Miwi2-bound small RNAs in Miwi2-immunoprecipitate from ES cells.

Analysis of reads overlapping known piRNA clusters

I then examined whether Miwi2 IP enriches for small RNAs that are produced from canonical piRNA clusters. With the assistance of Paul Young (Suter Laboratory, VCCRI), I performed piRNA cluster analysis as described in Section 2.11.2, Materials and Methods. 1% of total and 8% of Miwi2-bound small RNAs fell within piRNA clusters from piRNABank. In both the unenriched and Miwi2-bound small RNA populations, the majority are derived from known piRNA cluster regions in unannotated regions of the genome (52% in total and 87% in Miwi2-bound small RNA, Figure 5-4A and Figure 5-4B). Of the 48% of small RNAs from otherwise-annotated piRNA clusters in the unenriched sample, approximately one third are miRNAs and 23% align with tRNAs, 14% with LINE repeats, 9% with introns and 8% with LTR repeats (Figure 5-4A). Conversely, Miwi2-bound small RNAs that align with piRNA clusters are mostly produced from LTR (38%) and LINE (26%) retrotransposons (Figure 5-4B). A smaller percentage of the Miwi2-bound small RNAs aligning to features within piRNA clusters also map to tRNAs and rRNA.

Given that almost one quarter of all Miwi2-bound small RNAs are produced from unannotated regions of the genome (Figure 5-2), I visually examined whether some of these might align with large intergenic piRNA clusters reported in the pachytene testis (Gan et al., 2011). Using UCSC Genome Browser, I visually compared my small RNA data with a published small RNA dataset from 20.5dpp pachytene testis (Li et al., 2013) and, with the assistance of Paul Young (Suter Laboratory, VCCRI), converted the coordinates for compatibility with the most recent version of the mouse genome for visualisation. Figure 5-5 shows a screenshot from UCSC Genome Browser depicting annotated piRNA clusters on chromosomes 15 and 17 in the testis, which are also seen in the total small RNA population, and in the Miwi2-bound fraction of ES cells. The piRNA cluster on chromosome 17 (Figure 5-5A) is a typical large bi-directional cluster (~75kb) giving rise to piRNAs from both strands, whereas that on chromosome 15 (Figure 5-5B) is a smaller single-stranded cluster less than 20kb in size; both fall into the genomic coordinates of piRNA clusters deposited in piRNABank. The bi-directional cluster on chromosome 17 is interspersed with retrotransposon sequences, although the cluster on chromosome 15 is derived from a repeat-poor region.

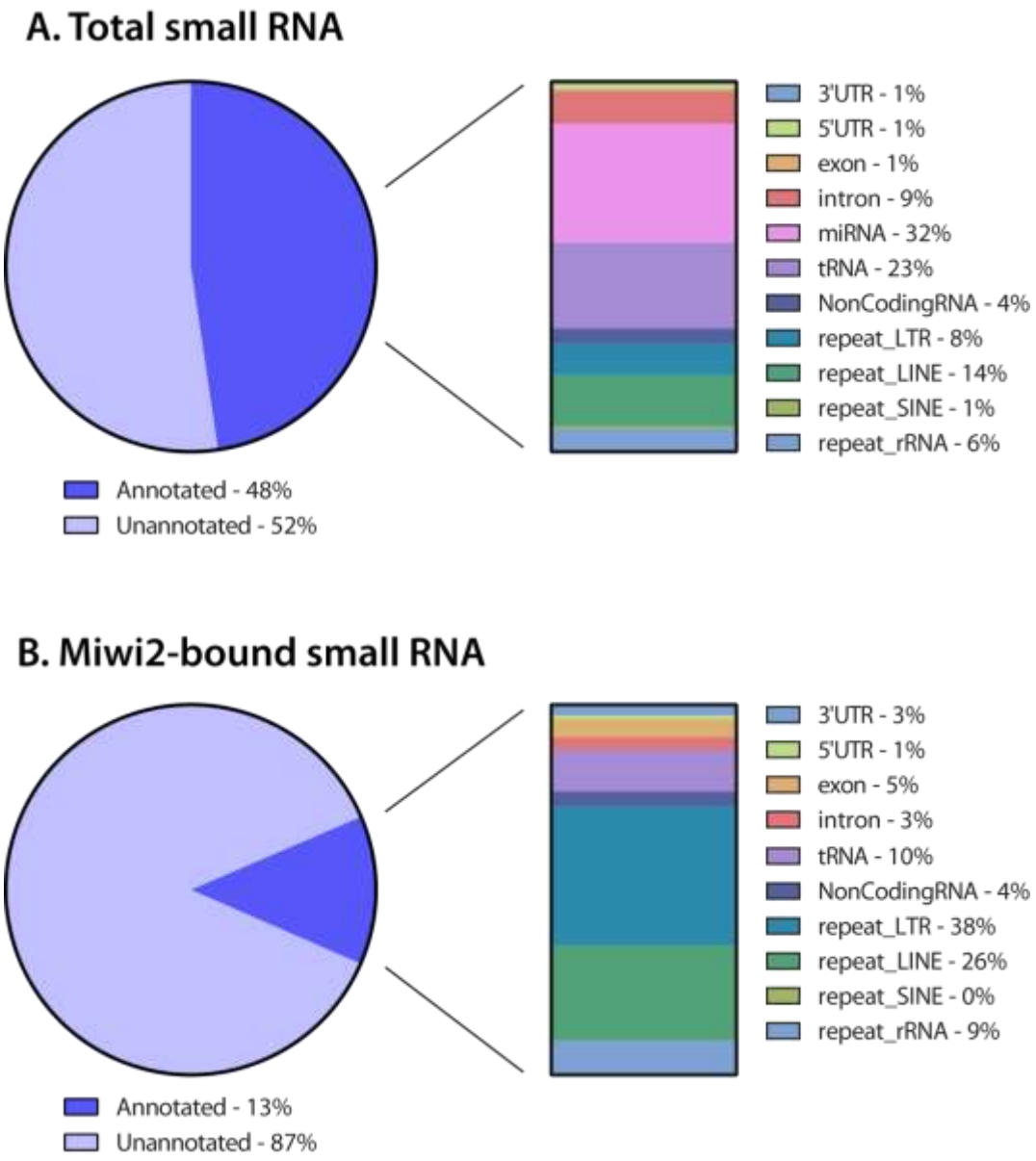


Figure 5-4: Annotations of small RNAs derived from piRNA clusters in ES cells

Proportion of annotated and unannotated reads aligning with piRNA clusters from piRNABank (left panel) and breakdown of annotated piRNA cluster reads by biotype (right panel) in (A) total small RNA and (B) Miwi2-bound small RNA fraction in ES cells.

Figure 5-5: Large, intergenic piRNA clusters in mouse ES cells and testis

UCSC Genome Browser view of small RNAs aligning in large clusters. (A) Cluster on Chromosome 17. (B) Cluster on Chromosome 15. Tracks displayed are ESC Total Small RNA, ESC Miwi2 IP and Mouse Testis 20.5dpp (Li et al., 2013). Small RNA reads from the positive strand shown in blue; reads from the negative strand shown in red. Other tracks displayed include RefSeq Genes and Repeating Elements. Arrows on genes denote 5' → 3' direction.

Analysis of repeat-derived small RNAs

Given that Miwi2 tends to associate with repeat sequences in the pre-pachytene testis (Aravin et al., 2008)(Kuramochi-Miyagawa et al., 2008), I analysed the precise annotations of the repeat-derived small RNAs in my unenriched (2%) and Miwi2-IP (50%) samples. In the total small RNA sample, 41% of repeat-derived small RNAs map to MaLR repeats, 24% to L1, 12% to LTR_ERVK (ERVK), and 6% to LTR_ERV1 (ERV1) repeats (Figure 5-6A, left panel). There is an enrichment of MaLR-derived small RNAs (68%) in Miwi2-bound population, and are smaller proportion are derived from L1 (13%), and rRNA (11%, Figure 5-6A, right panel). Only 5% of repeat-associated, Miwi2-bound small RNAs align with ERVK repeats. I then analysed the strand origin of the MaLR-associated small RNAs and observed that in both cases, the majority of small RNAs originate from the sense strand, although a larger percentage of unenriched MaLR-derived small RNAs map to the antisense strand relative to the Miwi2-bound population (8% and 2%, respectively).

Given that Miwi2 binds piRNAs derived from particular L1 elements in the embryonic testis (Aravin et al., 2008), I investigated the precise identity of these L1 repeat-derived small RNAs bound to Miwi2 in ES cells. Table 5-1 lists the breakdown of L1 annotations in the total and Miwi2-bound small RNA populations. In total ES cells, L1-derived small RNAs map mostly to L1Md_F2 and Lx9, and these also make up similar proportions of L1-derived Miwi2-piRNAs. However, there are other types of L1 repeats producing small RNAs enriched with Miwi2 pulldown, including L1MB7, Lx2, L1_Mur3, L1Md_F3, and L1M3. The L1-generated Miwi2-piRNAs only account for approximately half of all L1-derived small RNAs in ES cells (Table 5-1), the remainder of which are derived mostly from L1VL4, L2c, Lx5c, and L1M4 repeats.

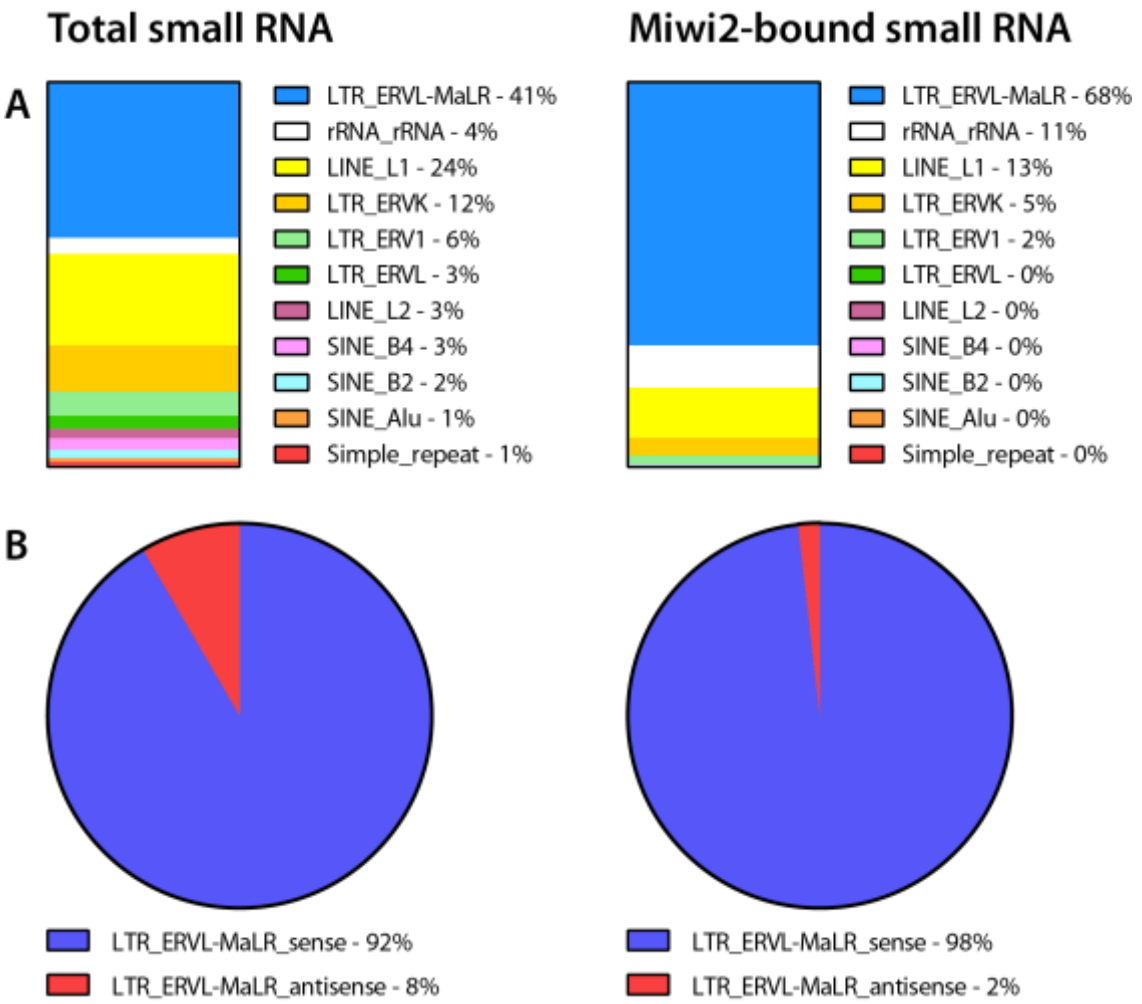


Figure 5-6: Repeat composition for total and Miwi2-bound small RNAs in ES cells

(A) Breakdown of types of repeat elements giving rise to small RNAs in ES cells. (B) Sense and antisense sequences as a proportion of all MaLR-derived small RNAs. Left-hand panel: Total small RNA. Right-hand panel: Miwi2-bound small RNA in ES cells.

Table 5-1: Breakdown of Miwi2-bound small RNAs mapping to LINE1 repeats

LINE1 annotation	% Total small RNA (LINE1)	% Miwi2-bound small RNA (LINE1)
L1MB7	2.9%	18.6%
L1Md_F2	16.2%	13.7%
Lx9	10.3%	12.8%
Lx2	1.5%	12.7%
L1_Mur3	3.7%	12.2%
L1Md_F3	3.7%	11.3%
L1M3	0.0%	7.3%
Lx8	5.1%	6.8%
L1MEg	1.5%	1.4%
L1_Mus3	0.7%	1.1%
L1MC3	0.0%	0.6%
L1MB2	0.0%	0.6%
Lx5b	1.5%	0.5%
L1MEg2	0.0%	0.2%
L1MC1	0.7%	0.2%
L2a	0.7%	0.2%
% all LINE1 reads	48.5%	100%

Analysis of gene-derived small RNAs

Our laboratory (and others) has observed piRNAs deriving from genes (Keam et al., 2014), and in my dataset, almost half of the Miwi2-bound small RNAs aligned with genic regions (Figure 5-2). Of all annotated small RNAs, these comprised 6% mapping to 495 genes and 39% mapping to 172 genes in the unenriched and Miwi2-bound RNA samples, respectively (Figure 5-2). Hence, in the Miwi2-enriched population, a greater number of small RNAs were produced per gene relative to the unenriched population. To test whether the abundance of gene-derived small RNAs was merely a reflection of degradation of the most abundant transcripts, I compared the abundance of gene-derived small RNAs to the relative abundance of their parent gene. Publically available gene expression datasets were obtained from the Gene Expression Omnibus (GEO, GSM64924 and GSM64926) and reanalysed to obtain expression values (using Affymetrix Expression Console). Figure 5-7 demonstrates that there is no correlation between gene-derived small RNA copy number and parent transcript abundance in either unenriched or Miwi2-bound small RNA populations (R^2 values are 0.00006 and 0.023, respectively). These data indicate that the gene-derived small RNAs captured by the Miwi2 IP (and extant in the ES cells without IP) are likely generated independently of the parent gene.

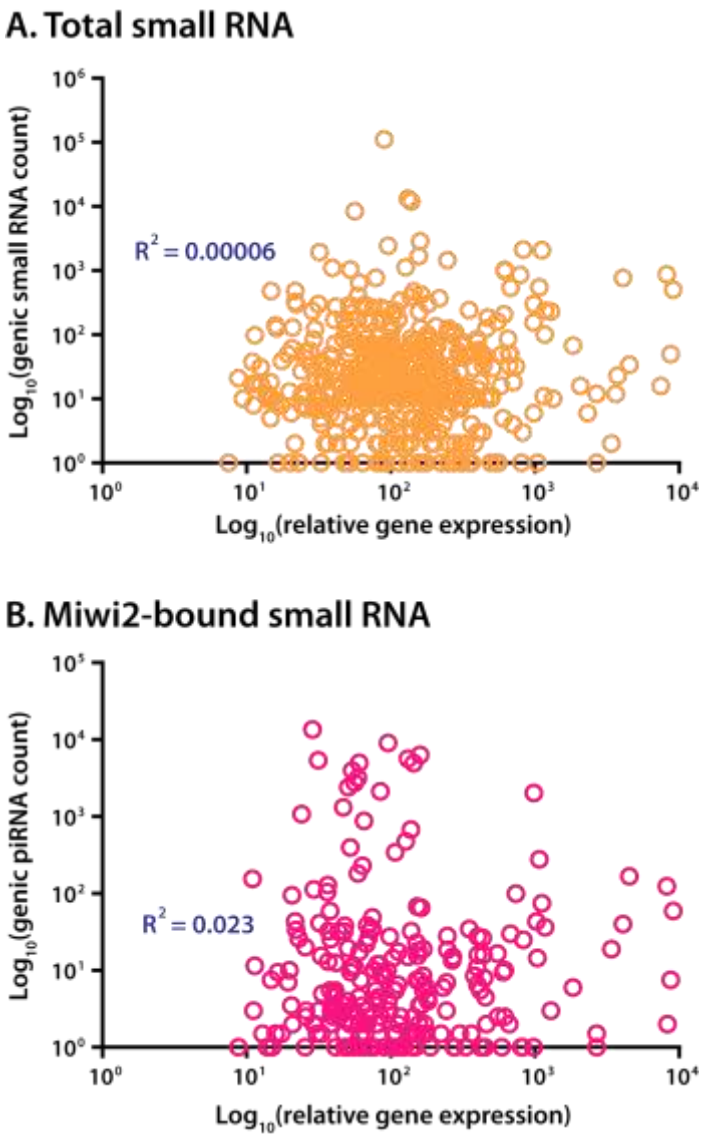


Figure 5-7: No correlation between copy number of gene-derived small RNAs and parent transcript abundance in ES cells

Scatter plot showing the number of gene-derived small RNAs compared to the expression level of the parent gene in R1 ES cells. Biological triplicate datasets for transcript abundance in wild type R1 ES cells obtained from Gene Expression Omnibus GSE2972 (Hailesellasse Sene et al., 2007). (A) Total small RNA. (B) Miwi2-bound piRNA.

Gene-derived piRNAs have previously been reported to occur in 3'UTRs (Robine et al., 2009, Keam et al., 2014) and in my datasets, small RNAs produced from 3'UTRs were enriched with Miwi2 IP relative to total small RNA in ES cells (6.8% and 0.34%, respectively). Using UCSC Genome Browser, I visually inspected whether some of these 3'UTR-derived small RNAs are produced from clusters, and – like the intergenic clusters – whether these are also present in the pachytene testis. Figure 5-8 shows expression of the 3'UTR cluster at *Abl2* in the testis as a representative example. Furthermore, Miwi2 IP enriches for 3'UTR clusters over *Abl2* and *Tfrc*, as well as scattered reads along the coding region. This pattern of small RNA mapping is seen throughout the genome over selected genes. Intriguingly, I also observed small RNAs clustered at the 3' and 5' ends and dispersed along the coding region of the master pluripotency genes, *Oct4, Sox2* and *Nanog* (Figure 5-9). These clusters are enriched by Miwi2 IP and are absent from the 20.5dpp testis dataset.

I predicted functions of small RNA-producing parent genes using Ingenuity Pathway Analysis. Genes whose UTRs or coding regions produce small RNAs with at least 10 copies were included in the analysis, and I filtered for networks incorporating more than one gene (Table 5-2 and Table 5-3). Small RNAs in unenriched ES cells align with genes predominately implicated in cellular growth and proliferation, organismal survival, and inflammatory response (Table 5-2). The most likely functions of genes that produce Miwi2-piRNAs in ES cells include cellular movement, reproductive system development and function, and cellular growth and proliferation (Table 5-3).

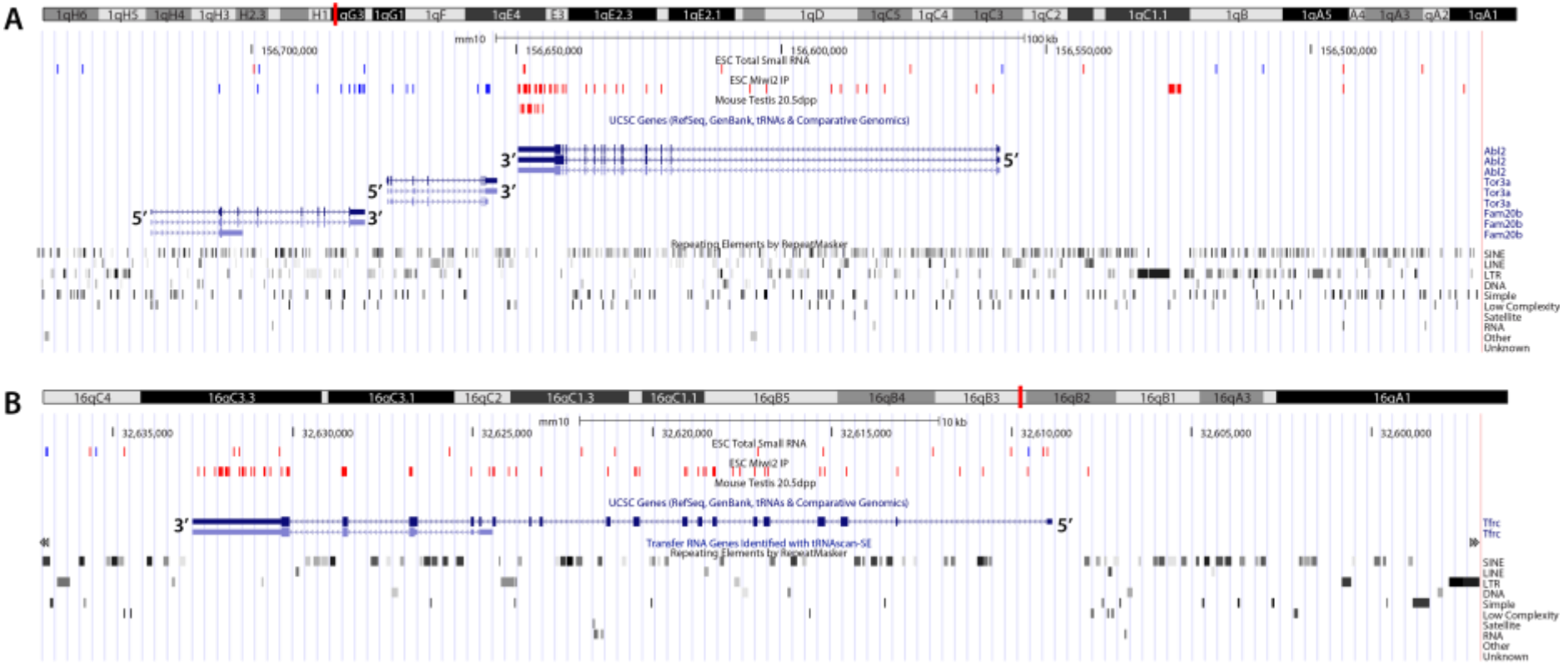


Figure 5-8: 3'UTR piRNA clusters in ES cells and the pachytene testis

UCSC Genome Browser view of small RNAs aligning with the 3'UTR, exons and introns of genes in ES cells. (A) *Abl, Tor3a* and *Fam20b* on Chromosome 1. (B) *Tfrc* on Chromosome 16. Small RNA tracks displayed are ESC Total Small RNA, ESC Miwi2 IP and Mouse Testis 20.5dpp (Li et al., 2013). Small RNA reads from the positive strand shown in blue; reads from the negative strand shown in red. Other tracks displayed include UCSC Genes, Transfer RNA Genes, and Repeating Elements. Arrows on genes denote 5' → 3' direction.

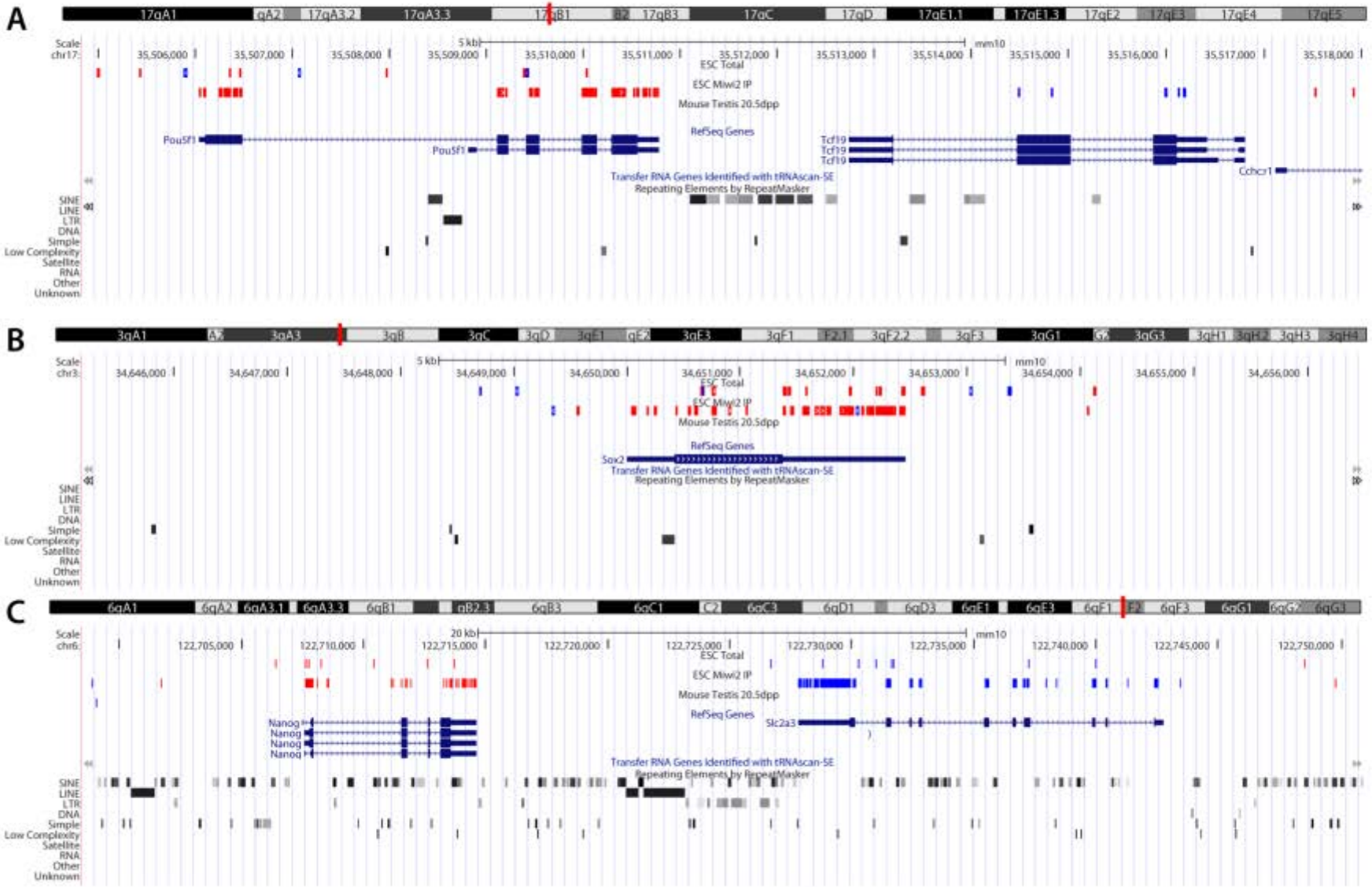

A
17qA1
17qA2
17qA3.2
17qA3.3
17qB1
17qB2
17qB3
17qC
17qD
17qE1.1
17qE1.3
17qE2
17qE3
17qE4
17qE5
Scale
chr17:
5 kb
mm10
35,506,000
35,507,000
35,508,000
35,509,000
35,510,000
35,511,000
35,512,000
35,513,000
35,514,000
35,515,000
35,516,000
35,517,000
35,518,000
ESC Total
ESC Miwi2 IP
Mouse Testis 20.5dpp
RefSeq Genes
Pou5f1
Pou5f1
Tcf19
Tcf19
Tcf19
Cchcr1
Transfer RNA Genes Identified with tRNAscan-SE
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
B
3qA1
3qA2
3qA3
3qB
3qC
3qD
3qE1
3qE2
3qE3
3qF1
3qF2.1
3qF2.2
3qF3
3qG1
3qG2
3qG3
3qH1
3qH2
3qH3
3qH4
Scale
chr3:
5 kb
mm10
34,646,000
34,647,000
34,648,000
34,649,000
34,650,000
34,651,000
34,652,000
34,653,000
34,654,000
34,655,000
34,656,000
ESC Total
ESC Miwi2 IP
Mouse Testis 20.5dpp
RefSeq Genes
Sox2
Transfer RNA Genes Identified with tRNAscan-SE
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
C
6qA1
6qA2
6qA3.1
6qA3.3
6qB1
6qB2.3
6qB3
6qC1
6qC2
6qC3
6qD1
6qD3
6qE1
6qE3
6qF1
6qF2
6qF3
6qG1
6qG2
6qG3
Scale
chr6:
20 kb
mm10
122,705,000
122,710,000
122,715,000
122,720,000
122,725,000
122,730,000
122,735,000
122,740,000
122,745,000
122,750,000
ESC Total
ESC Miwi2 IP
Mouse Testis 20.5dpp
RefSeq Genes
Nanog
Nanog
Nanog
Nanog
Slc2a3
Transfer RNA Genes Identified with tRNAscan-SE
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

Figure 5-9: piRNA clusters at the 3'UTR of key pluripotency genes in ES cells

UCSC Genome Browser view of small RNAs aligning with the 3'UTR, exons and introns of genes in ES cells. (A) *Oct4* (*Pou5f1*) and *Tcf19* on Chromosome 17. (B) *Sox2* on Chromosome 3. (C) *Nanog* and Slc2a3 on Chromosome 6. Small RNA tracks displayed are ESC Total Small RNA, ESC Miwi2 IP and Mouse Testis 20.5dpp (Li et al., 2013). Small RNA reads from the positive strand shown in blue; reads from the negative strand shown in red. Other tracks displayed include RefSeq Genes, Transfer RNA Genes, and Repeating Elements. Arrows on genes denote 5' → 3' direction.

Table 5-2: Gene networks from pathway analysis of genes producing small RNAs in ES cells

# genes involved	Top Diseases and Functions
28	Cellular Growth and Proliferation, Organismal Survival, Inflammatory Response
20	Cellular Movement, Infectious Disease, Cell Death and Survival
18	Behaviour, Lipid Metabolism, Small Molecule Biochemistry
16	Endocrine System Disorders, Endocrine System Development and Function, Tissue Morphology
15	Cancer, Neurological Disease, Cellular Development
15	Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Behaviour
14	Cardiac Dysfunction, Cardiovascular Disease, Heart Failure
13	Cell Death and Survival, Cellular Development, Cellular Growth and Proliferation
13	Cellular Development, Cellular Growth and Proliferation, Haematological System Development and Function
13	Tissue Morphology, Cell-To-Cell Signalling and Interaction, Tissue Development
13	Nervous System Development and Function, Embryonic Development, Organismal Development
13	Drug Metabolism, Protein Synthesis, Cell-To-Cell Signalling and Interaction
13	Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Cellular Assembly and Organization
12	Cellular Movement, Haematological System Development and Function, Hypersensitivity Response
12	Cell Death and Survival, Cell Cycle, Cancer
12	Cellular Movement, Immune Cell Trafficking, Haematological System Development and Function
11	Cellular Movement, Cellular Growth and Proliferation, Cellular Assembly and Organization
11	Cell Death and Survival, Organ Morphology, Renal Dilation
10	Protein Synthesis, Gene Expression, RNA Post-Transcriptional Modification

Table 5-3: Gene networks from pathway analysis of genes producing small RNAs bound to Miwi2 in ES cells

# genes involved	Top Diseases and Functions
16	Cellular Movement, Reproductive System Development and Function, Cellular Growth and Proliferation
14	Endocrine System Disorders, Gastrointestinal Disease, Metabolic Disease
13	Protein Synthesis, Cell Morphology, Connective Tissue Development and Function
13	Cell Cycle, Cellular Movement, Cell Morphology
11	Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Inflammatory Response
11	Organismal Injury and Abnormalities, Renal and Urological Disease, Cellular Development
11	Endocrine System Development and Function, Molecular Transport, Protein Synthesis

5.3 Immunoprecipitation of endogenous Miwi2 from 4T1 mammary carcinoma cells

It remained to be answered whether Miwi2 binds piRNAs in murine cancer cells; I therefore isolated endogenous Miwi2 and its bound RNAs from 4T1 mammary carcinoma cells for identification and further analysis. I have previously discussed optimisation of the α-Miwi2 monoclonal antibody panel and validation of the Miwi2 IP in Section 5.1, and I employed an identical strategy to immunoprecipitate the RNA:protein complex as described in Section 5.2. Miwi2-bound RNAs were subjected to library amplification for next-generation sequencing according to the method described in Section 2.11.1, Materials and Methods.

Bioinformatic processing of small RNA sequencing data

I performed Illumina HiSeq® next-generation sequencing of total small RNA from whole 4T1 cells ("total small RNA") and small RNAs enriched following Miwi2 immunoprecipitation from 4T1 cells ("Miwi2-bound small RNA"). All findings for Miwi2-bound small RNAs in this section are averaged from biological duplicate results. Quality filtering and adapter trimming was performed with Paul Young (Suter Laboratory, VCCRI). Following bioinformatic processing (see Section 2.11.2, Materials and Methods), I had approximately 5.57 million reads for the total small RNA sample and a total of 11.45 million reads for my Miwi2-bound small RNA samples (Table 8-1, Appendix). Of those reads, the majority mapped successfully to the mouse genome when allowing 1 mismatch (88% for total small RNA; an average of 91% for Miwi2-bound small RNA). For the total small RNA sample, I obtained 91,290 unique sequences and for the Miwi2-bound small RNA duplicate sequencing runs, I observed 59,123 unique sequences. My total small RNA sequencing dataset included single copy reads (84%), indicating that the small RNA sequencing was not exhaustive. Only an average of 8% sequences from the duplicate Miwi2 IP samples were present as single reads.

Basic characterisation of total and Miwi2-bound small RNAs

Size distributions and annotations of unenriched and Miwi2-bound small RNAs in 4T1 cells are depicted in Figure 5-10. Of the unenriched small RNA population, the dominant length is 21-24nt (Figure 5-10A). Ninety five percent of total small RNA reads align to annotated regions of the genome; the vast majority are miRNAs with a smaller proportion mapping to genes, miscellaneous noncoding RNA, repeats, and tRNAs (Figure 5-10B). Repeat-derived small RNAs mostly map to LTRs and LINEs, and gene-derived small RNAs align mostly with introns and 5' UTRs (Figure 5-10B). Of all annotated small RNAs in the unenriched population, the overwhelming majority are derived from the sense strand with respect to annotated sequecnce features (Figure 5-10C).

Conversely, Miwi2-enriched RNAs do not exhibit an overt peak in length distribution. With the exception of a slight preference for 20 and 24nt sequences, the length Miwi2-bound RNAs is similarly distributed from 19-33nt (Figure 5-10D). Miwi2 immunoprecipitation enriches for small RNAs from unannotated regions over total small RNAs in 4T1 cells (Figure 5-10E). Of the 90% of annotated small RNAs, almost half map to genes, 36% map to repeats, 12% map to tRNAs and 5% map to miscellaneous noncoding RNA (Figure 5-10E). The dominant type of repeat element giving rise to Miwi2-piRNAs is LTR, followed by LINE and rRNA. Notably, tRNA-derived small RNAs are approximately pi-sized: 24-34nt in length (Figure 5-10D). Of all annotated reads, 99% are produced from the sense strand with respect to annotated sequence features (Figure 5-10F).

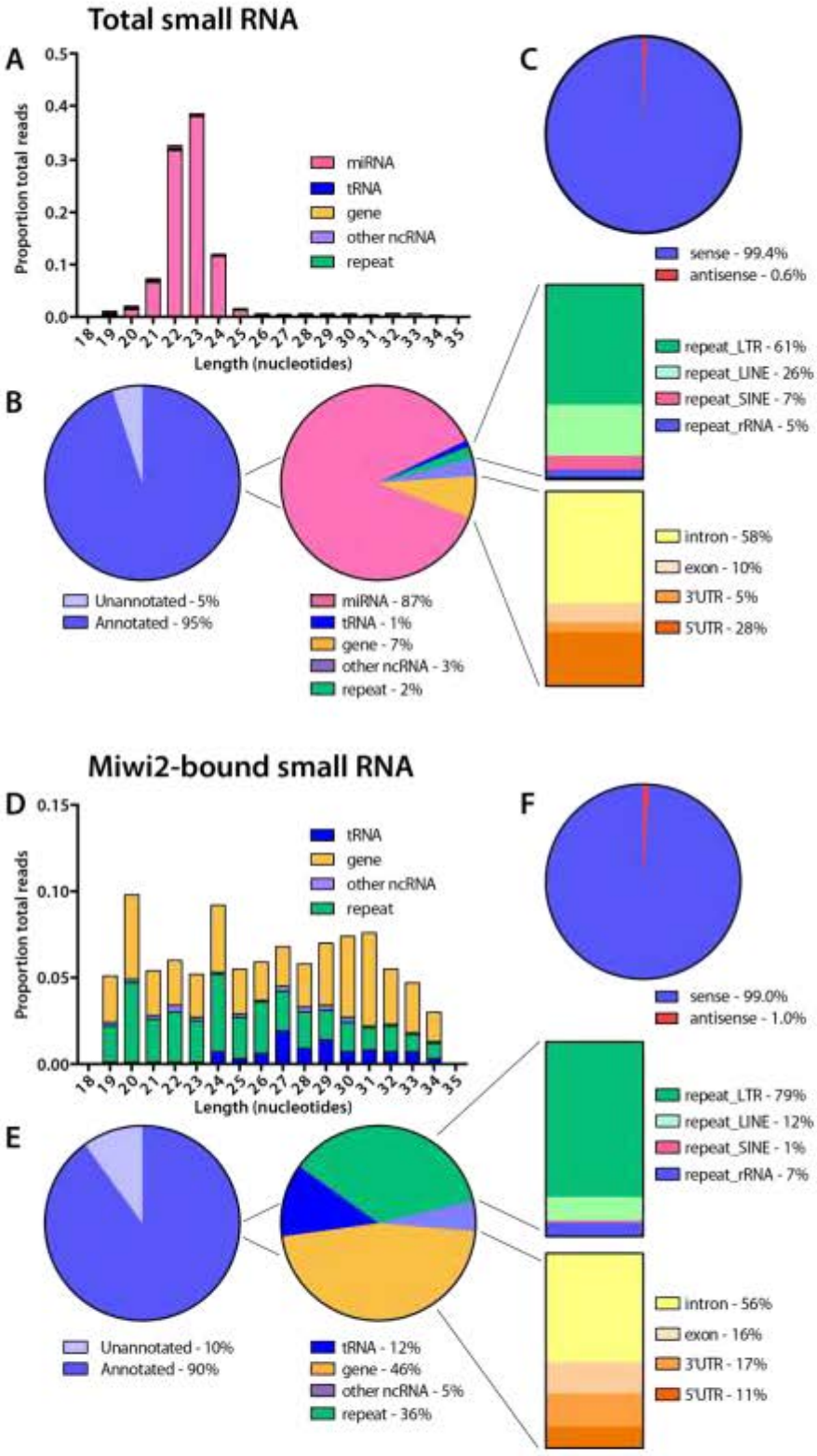


Figure 5-10: Sequence characteristics of total and Miwi2-bound small RNA in 4T1 cells

Small RNA sequence characteristics of (A-C) total small RNA and (D-F) Miwi2-bound small RNA. (A, D) Length distribution of small RNAs between 18-35nt with annotations. Values represented as a proportion of total small RNA. (B, E) Categorical annotations of small RNAs. (C, F) Sense and antisense sequences as a proportion of all small RNA reads.

Sequence alignment of total and Miwi2-bound small RNAs

To check whether the unenriched and Miwi2-bound small RNAs exhibit canonical piRNA-like sequence features (such as 1U or 10A), I analysed base preferences of 4T1 small RNA sequences at positions 1 and 10. Of the total small RNA population in 4T1 cells, the majority harbour a 5' U and at position 10, most sequences have an overrepresentation of guanine and an underrepresentation of uracil and adenine (Figure 5-11A). The 11 most abundant sequences are responsible for this bias, as they harbour a 5' U and comprise 39% total reads. Conversely, the preference for 1U is not observed in Miwi2-piRNAs, with only a small bias for adenine at positions 1 and 10 (33%) and a slight underrepresentation of cytosine at positions 1 and 10 (Figure 5-11B). Hence, unlike in ES cells, there is no evidence of enrichment of primary piRNA sequence features in the Miwi2-bound fraction in 4T1 cells; but like ES cells, there is no bias for secondary processing sequence features.

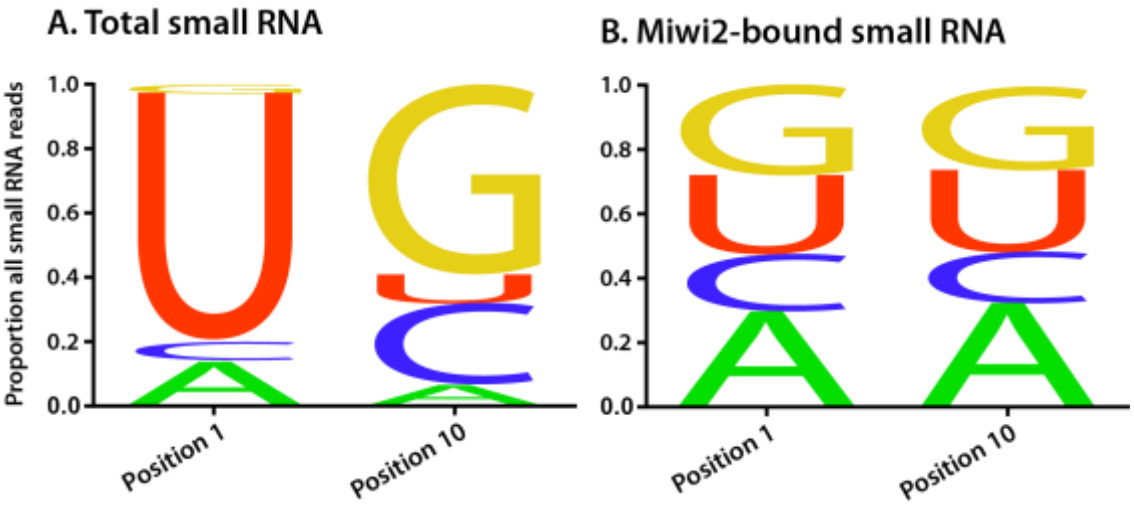


Figure 5-11: Sequence alignment for total and Miwi2-bound small RNA in 4T1 cells

Sequence alignment at positions 1 and 10 of (A) total small RNA population from whole 4T1 cells and (B) Miwi2-bound small RNAs in 4T1 cells.

Analysis of piRNA clusters in 4T1 cells

I then examined whether Miwi2 IP enriches for small RNAs that are produced from canonical piRNA clusters. With the assistance of Paul Young (Suter Laboratory, VCCRI), I performed piRNA cluster analysis as described in Section 2.11.2, Materials and Methods. One percent of total and 8% of Miwi2-bound small RNAs fell within known piRNA clusters. There is a clear enrichment of cluster-derived small RNAs mapping to unannotated regions in Miwi2 pulldowns compared to total small RNAs (38% and 20%, respectively). I then analysed the genomic origin of small RNAs generated by otherwise-annotated piRNA clusters. In whole 4T1 cells, small RNAs are produced from piRNA clusters that mostly overlap with tRNAs, introns, and miRNAs. Miwi2 binds RNAs produced by piRNA clusters overlapping predominately with introns, LTRs, and tRNAs.

Given that 10% of Miwi2-bound small RNAs are derived from unannotated regions in 4T1 cells (Figure 5-10), I visually examined my datasets for examples of large piRNA clusters at intergenic loci, as I have shown in ES cells. Using UCSC Genome Browser, I compared unenriched and Miwi2-bound small RNAs from 4T1 cells to total small RNA from the 20.5dpp testis (Li et al., 2013). As an example, annotated piRNA clusters in the pachytene testis at chromosome 13 and 15 are not prominent in 4T1 cells: several scattered reads are present within the cluster on chromosome 15 in the total small RNA sample, but are absent in the Miwi2-associated fraction (Figure 5-13). Visual inspection over large chromosomal regions using UCSC Genome Browser yielded no instances of canonical, intergenic piRNA cluster expression in 4T1 cells.

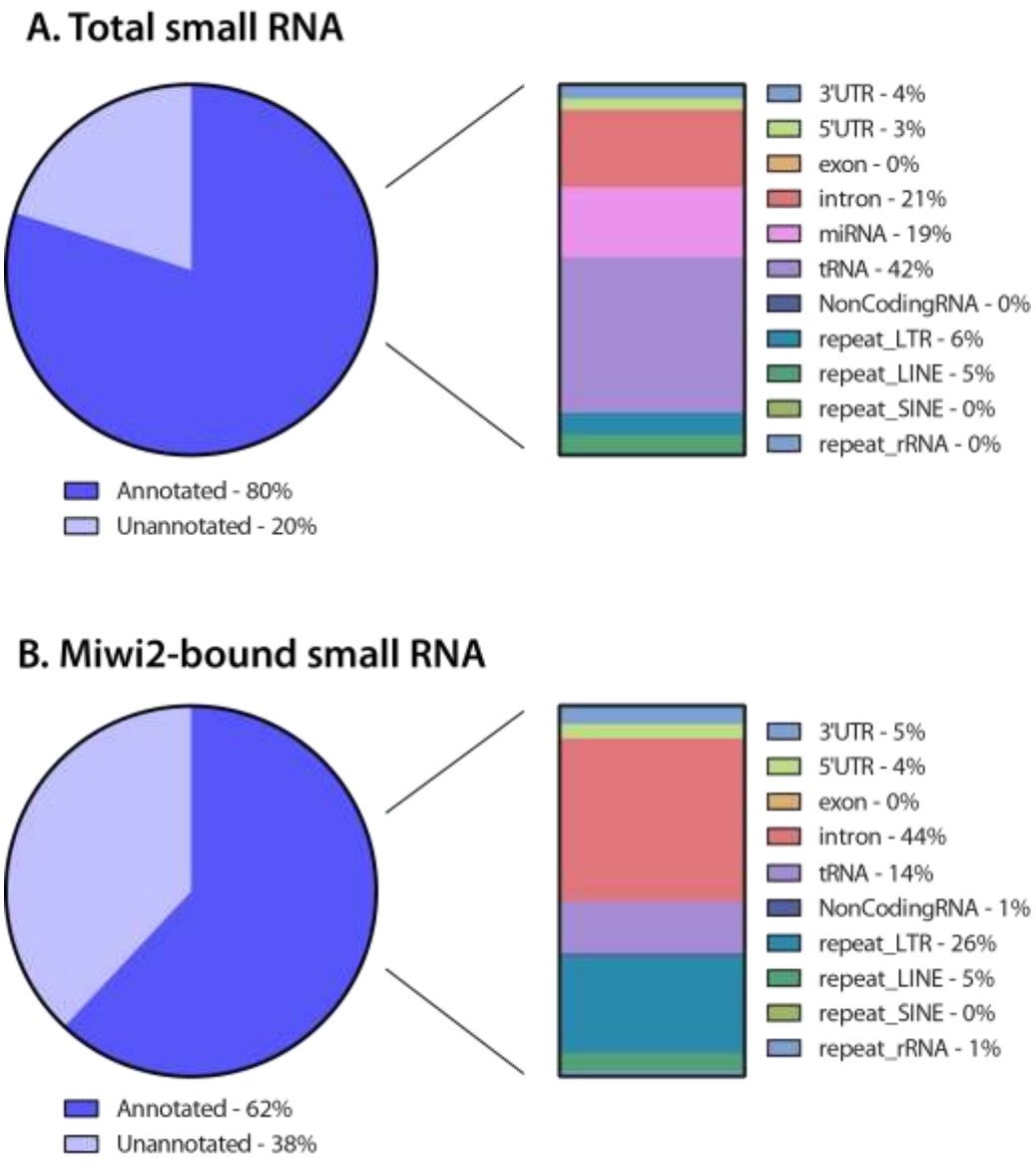


Figure 5-12: Annotations of small RNAs derived from known piRNA clusters in 4T1 cells

Proportion of annotated and unannotated reads aligning with piRNA clusters from piRNABank (left panel) and breakdown of annotated piRNA cluster reads by biotype (right panel) in (A) total small RNA and (B) Miwi2-bound small RNA fraction in 4T1 cells.

Figure 5-13: UCSC Genome Browser view of intergenic piRNA cluster loci in 4T1 cells and mouse testis

UCSC Genome Browser view of representative piRNA clusters. (A) Cluster on Chromosome 13. (B) Cluster on Chromosome 15. Tracks displayed are 4T1 Total Small RNA, 4T1 Miwi2 IP and Mouse Testis 20.5dpp (Li et al., 2013). Small RNA reads from the positive strand shown in blue; reads from the negative strand shown in red. Other tracks displayed include RefSeq Genes, Transfer RNA Genes, and Repeating Elements. Arrows on genes denote 5' → 3' direction.

Analysis of repeat-derived small RNAs

I performed a detailed comparison of the repeat-derived small RNAs in both set of 4T1 samples. Of the 2% repeat-derived small RNAs in unenriched 4T1 cells, the majority map to MaLR, L1, and ERVK repeats (Figure 5-14A). After enriching for Miwi2-bound small RNAs, it is apparent that most of the 36% repeat-derived reads come from MaLR repeats, with a smaller percentage mapping to L1 and ERVK repeats than in the total small RNA cohort (Figure 5-14A). There is, however, a slight enrichment in sequences produced from rRNA repeats in the Miwi2-bound dataset than the unenriched population. Miwi2 appears to selectively associate with small RNAs from the sense strand of MaLR repeats, a slight enrichment over MaLR-derived small RNA from the unenriched population.

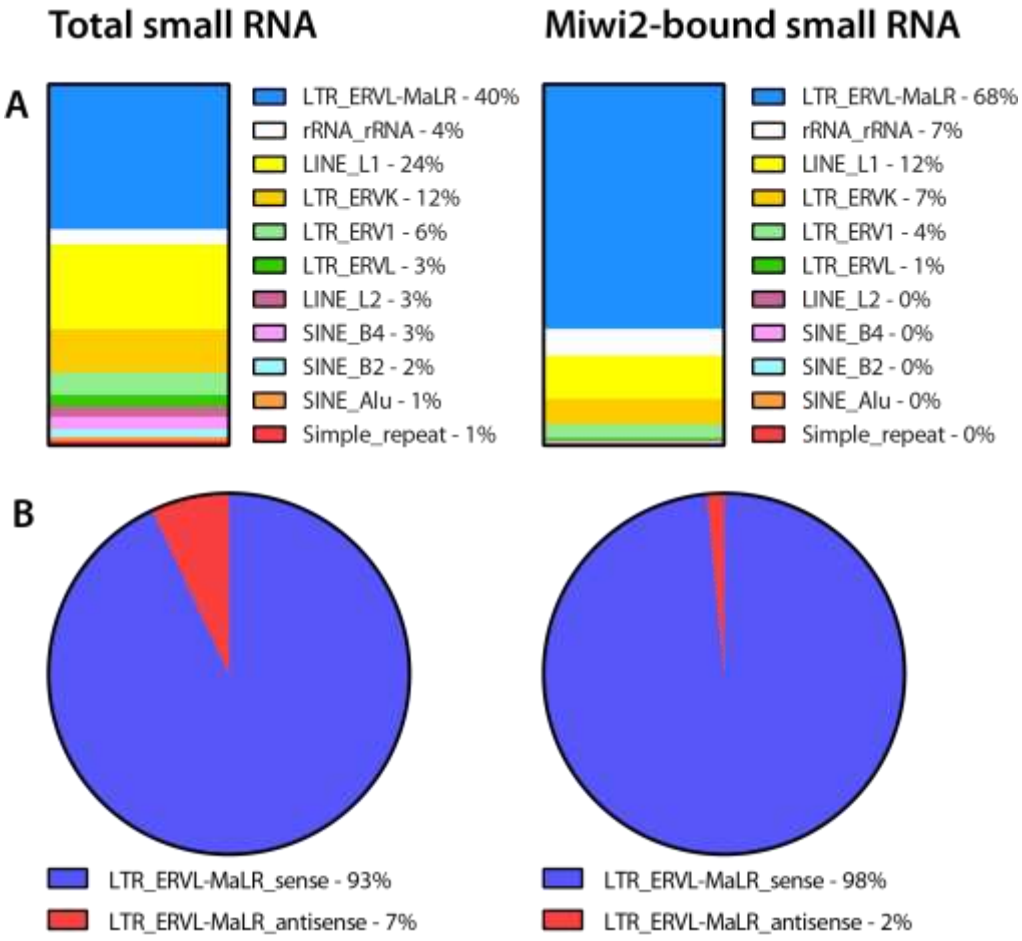


Figure 5-14: Repeat composition for total and Miwi2-bound small RNAs in 4T1 cells

(A) Breakdown of types of repeat elements giving rise to small RNAs in 4T1 cells. (B) Sense and antisense sequences as a proportion of all MaLR-derived small RNAs. Left-hand panel: Total small RNA. Right-hand panel: Miwi2-bound small RNA in 4T1 cells.

Analysis of genes giving rise to small RNAs

Like in ES cells, I examined whether Miwi2 associates with piRNAs produced from genes. I observed an enrichment of gene-derived small RNAs in the Miwi2-bound population relative to total small RNA (46% and 7% of all reads, respectively, Figure 5-10). Small RNAs with a copy number of at least 10 are produced from 38 different genes in unenriched 4T1 cells. 25 of these genes - plus 14 others - produce small RNAs that were enriched with Miwi2 immunoprecipitation and present in at least 10 copies. However, there were also gene-derived small RNAs that exhibit low copy number: 4% of all small RNA reads are present in less than 10 copies, and 1% are uniquely expressed.

I compared overall transcript abundance versus copy number of the small RNAs aligning with those genes to ask whether the most highly-expressed gene-derived small RNAs were simply degradation products: if so, there should be a correlation between small RNA copy number and parent transcript abundance. I obtained duplicate Affymetrix MoGene1.0 gene expression datasets for wild type 4T1 cells from GEO GSE25525 (datasets GSM627512 and GSM627513). I generated expression values using Affymetrix Expression Console. Figure 5-15 shows that there is no correlation between transcript abundance in whole 4T1 cells and genic small RNA copy number in either unenriched or Miwi2-bound small RNA populations (R^2 values are 0.0002 and 0.01, respectively), implying that these small RNAs are unlikely to be degradation products, and that Miwi2 is selectively binding gene-derived piRNAs.

Similar to ES cells, visual inspection of my datasets revealed instances of 3'UTR clusters of small RNAs 4T1 cells. Again, I compared the 4T1 datasets with that of the 20.5dpp testis (Li et al., 2013). As a representative example, a snapshot of chromosome 17 from UCSC Genome Browser (Figure 5-16A) shows that at this 4T1 locus – particularly in the Miwi2-bound fraction – small RNA reads cluster over genes rather the central intergenic region, and that these are mostly absent in the pachytene testis. Figure 5-16B shows small RNAs clustering across and downstream of the 3'UTR of *Ago2* in the testis and 4T1 cells; this cluster is noticeably enriched by Miwi2 pulldown. In the 4T1 Miwi2 IP sample, a smaller cluster can be seen at the 3'UTR of the adjacent gene, *Ptk2*, as its 3'UTR is shorter than *Ago2.*

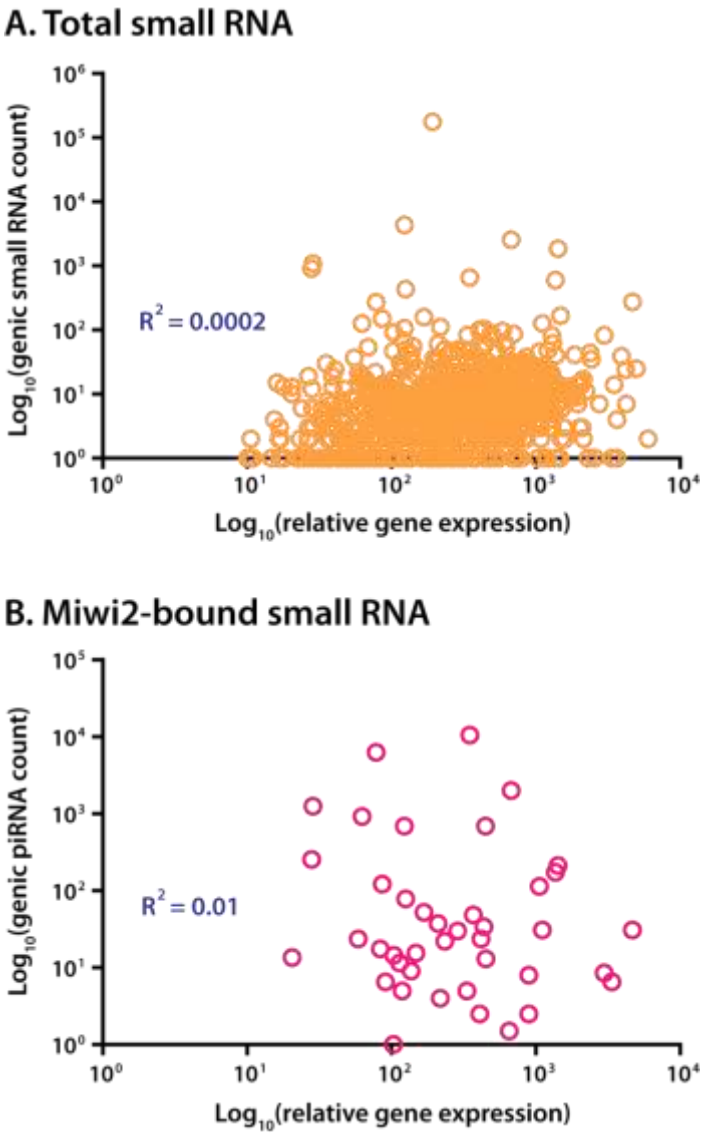


Figure 5-15: No correlation between copy number of gene-derived small RNAs and parent transcript abundance in 4T1 cells

Scatter plot showing the number of gene-derived small RNAs compared to the expression level of the parent gene in 4T1 cells. Biological duplicate datasets for transcript abundance in wild type 4T1 cells obtained from Gene Expression Omnibus GSE25525. (A) Total small RNA. (B) Miwi2-bound piRNA.

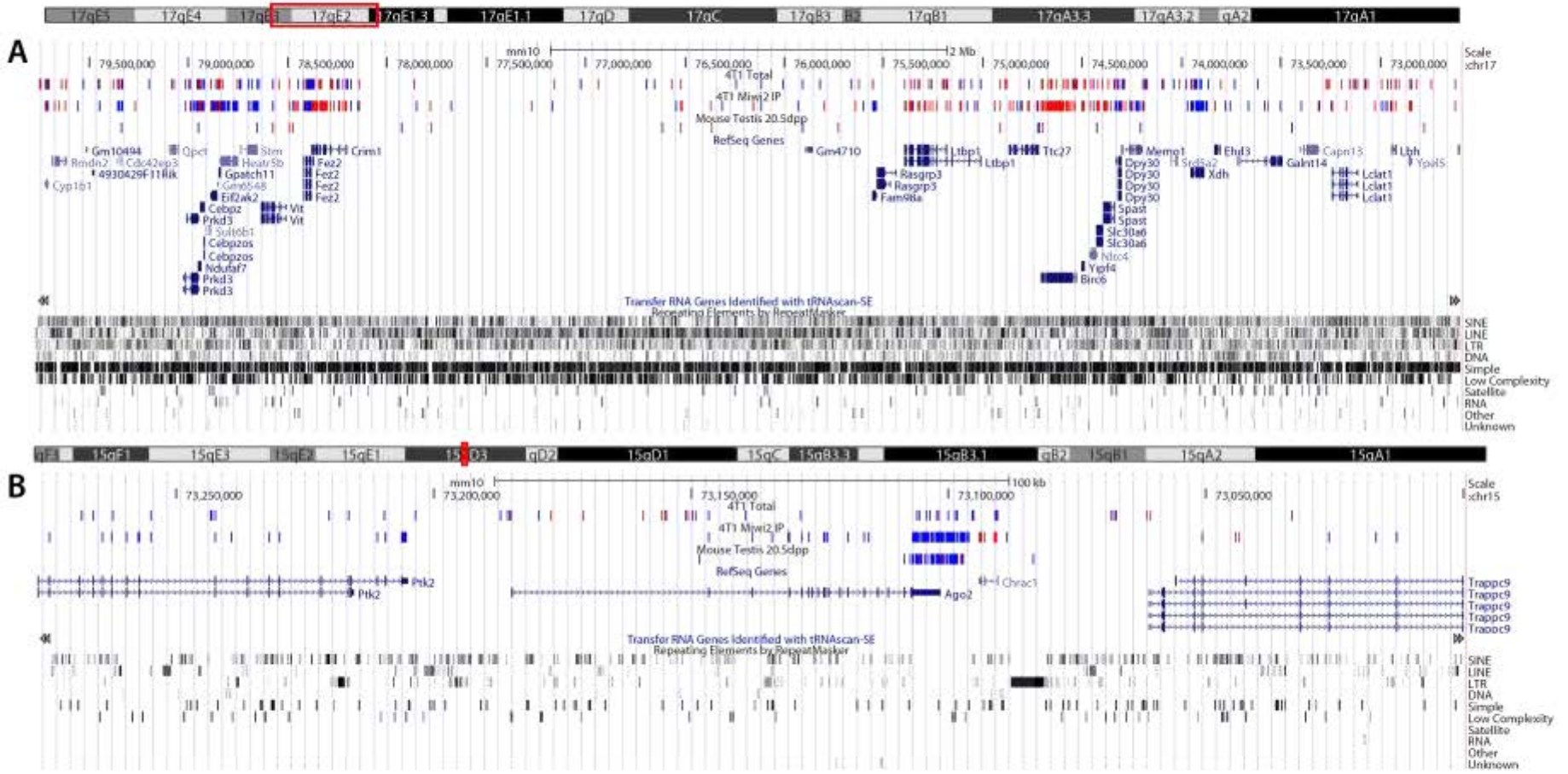

A
17qE5
17qE4
17qE2
17qE1.3
17qE1.1
17qD
17qC
17qB3
17qB1
17qA3.3
17qA3.2
17qA2
17qA1
mm10
2 Mb
79,500,000
79,000,000
78,500,000
78,000,000
77,500,000
77,000,000
76,500,000
76,000,000
75,500,000
75,000,000
74,500,000
74,000,000
73,500,000
73,000,000
Scale
chr17
4T1 Total
4T1 Miwi2 IP
Mouse Testis 20.5dpp
RefSeq Genes
Gm10494
Qpct
Stim
Crim1
Rmdn2
Cdc42ep3
Heatr5b
Fez2
4930429F11Rik
Gpatch11
Cyp1b1
Gm9548
Eif2ak2
Cebpz
Vit
Prkd3
Slit6b1
Cebpzos
Ndufaf7
Gm4710
Ltbp1
Ttc27
Rasgrp3
Fam98a
Mcmq1
Ehd3
Capn13
Lbh
Dpy30
Sgd5a2
Xdh
Galnt14
Lclat1
Spast
Slc30a6
Ndrc4
Ygpf4
Birc6
Transfer RNA Genes Identified with tRNAscan-SE
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
B
15qF1
15qE3
15qE2
15qE1
15qD2
15qD1
15qC
15qB3.3
15qB3.1
15qB2
15qB1
15qA2
15qA1
mm10
100 kb
73,250,000
73,200,000
73,150,000
73,100,000
73,050,000
Scale
chr15
4T1 Total
4T1 Miwi2 IP
Mouse Testis 20.5dpp
RefSeq Genes
Ptk2
Ago2
Chrac1
Trappc9
Transfer RNA Genes Identified with tRNAscan-SE
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

Figure 5-16: UCSC Genome Browser view of 3'UTR piRNA cluster loci in 4T1 cells and testis

UCSC Genome Browser view of small RNAs aligning with selected genic clusters. (A) Cluster on Chromosome 13. (B) Cluster on Chromosome 15. Tracks displayed are 4T1 Total Small RNA, 4T1 Miwi2 IP and Mouse Testis 20.5dpp (Li et al., 2013). Small RNA reads from the positive strand shown in blue; reads from the negative strand shown in red. Other tracks displayed include RefSeq Genes, Transfer RNA Genes, and Repeating Elements. Arrows on genes denote 5' → 3' direction.

I then predicted functions of small RNA-producing parent genes using Ingenuity Pathway Analysis. Genes whose UTRs or coding regions produce small RNAs with at least 10 copies were included in the analysis, and I filtered for networks incorporating more than one gene. Small RNA-producing genes in unenriched 4T1 cells are likely to function in cell signalling, cellular function and maintenance, and molecular transport (Table 5-4). Genes giving rise to small RNAs bound to Miwi2 in 4T1 cells are predominately involved in networks for cellular movement, cell-to-cell signalling and interaction, and embryonic development (Table 5-5). Hence, like in ES cells, it appears that Miwi2 in 4T1 cells is selectively binding piRNAs produced by genes involved in key cellular processes.

Table 5-4: Pathway analysis of genes giving rise to small RNAs in whole 4T1 cells

# genes involved	Top Diseases and Functions
10	Cell Signalling, Cellular Function and Maintenance, Molecular Transport
4	Cancer, Gastrointestinal Disease, Hepatic System Disease

Table 5-5: Pathway analysis of genes giving rise to Miwi2-bound small RNAs in 4T1 cells

# genes involved	Top Diseases and Functions
10	Cellular Movement, Cell-To-Cell Signalling and Interaction, Embryonic Development
2	Cancer, Cell Death and Survival, Cellular Development

Analysis of tRNA-derived small RNAs

Our laboratory has recently reported the association of processed tRNA fragments (tRFs) with the human Piwi, Hiwi2, in MDAMB231 breast cancer cells (Keam et al., 2014), plus I have found that 12% of Miwi2-bound piRNAs are derived from tRNAs (Figure 5-10). I found that small RNAs mapped to 41 different tRNA genes in both unenriched and Miwi2-IP samples. In unenriched 4T1 cells, there are two peaks in length distribution at 18nt and 32nt (Figure 5-17). However, the Miwi2-associated tRFs are, in a canonical sense, mostly pi-sized (24-34nt) with a preference for 27nt and 29nt fragments.

These tRFs predominately map to five tRNA isotypes: GlyGCC, LysCTT, GluCTC, GlyCCC, and HisGTG (Figure 5-17). In the unenriched small RNA population, the majority of reads map to LysCTT, followed by GlyGCC, and these are predominately 30-32nt in length. Most of the smaller (19-24nt) tRFs in 4T1 cells map to other tRNAs, including LysTTT, CysGCA, AspGTC, AlaTGC, ValCAC (Figure 5-17A). Miwi2 pulldown enriches for GlyGCC-derived tRFs; these occupy the greatest proportion of 24-31nt tRFs. The largest proportion of LysCTT-derived small RNAs in the Miwi2-bound fraction is seen in the 32-33nt tRFs, consistent with that seen in unenriched 4T1 cells. Figure 5-18 depicts stacked GlyGCC- and LysCTT-derived tRF reads in WIG format using UCSC Genome Browser (GRCm38/mm10) and in cloverleaf form, indicating that these tRFs align with the 5' end of the full length tRNA. Figure 8-12 (refer to Appendix) is a representative snapshot from UCSC Genome Broswer showing small RNA reads from the pachytene testis in WIG format: notably, 5' stacked small RNA reads align with the LysCTT and GlyGCC tRNA genes, but are less abundant than in total 4T1 cells and the Miwi2-bound fraction in Figure 5-18.

Thus, it appears that Miwi2 is selectively binding piRNAs derived from specific tRNAs, and that these are processed from the 5' end; this is in concordance with that previously described by our laboratory (Keam et al., 2014).

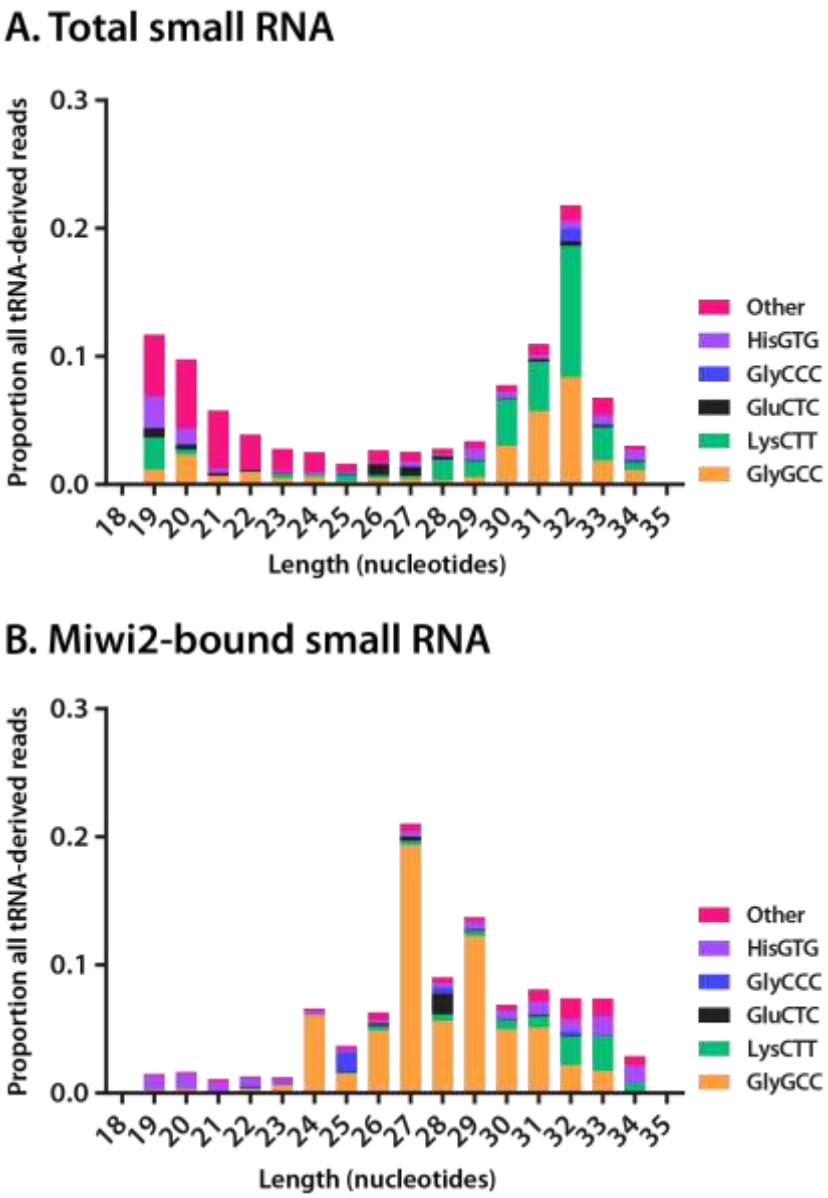


Figure 5-17: Length distribution and tRNA annotations of tRFs in 4T1 cells

tRFs are derived from specific tRNAs in (A) total and (B) Miwi2-bound small RNA population in 4T1 cells. tRF abundance represented as a proportion of all tRNA reads.

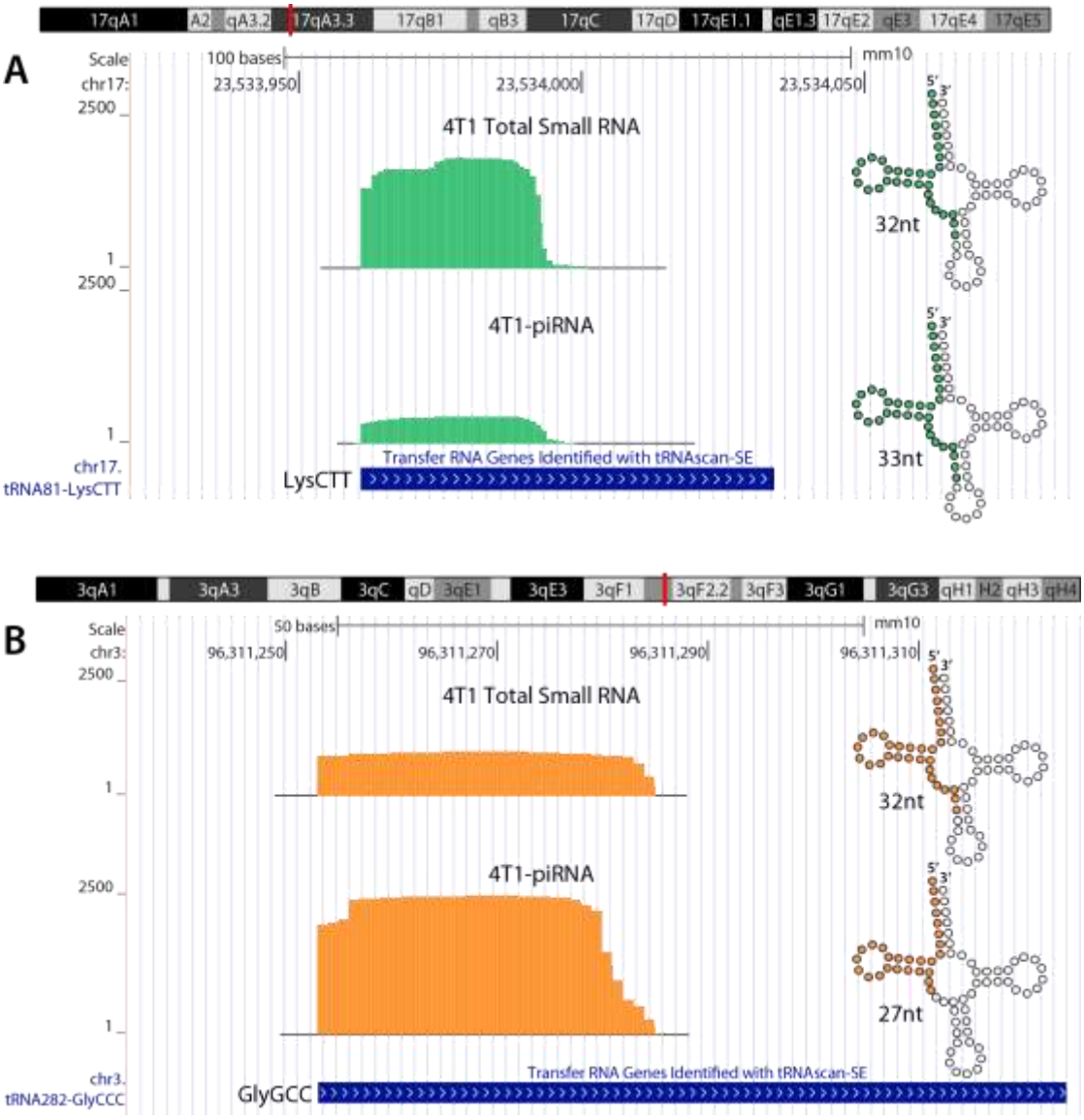


Figure 5-18: Positioning of tRFs at genomic loci in 4T1 cells

Graphical representation of the two most abundant 4T1 tRFs aligned with the mouse genome (UCSC Genome Browser, GRCm38/mm10). (A) LysCTT and (B) GlyGCC. Small RNA reads displayed in WIG format. Secondary tRNA cloverleaf structures superimposed on browser image with coloured circles indicating the most abundant tRF length

Discussion

Miwi2 can be immunoprecipitated from mouse cells

There are no commercially-available antibodies for immunoprecipitation of Miwi2. Several studies have published Miwi2 IP data based on the use of α-Miwi2 antibodies developed in-house, however, these published antibodies (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008) are polyclonal and available in finite supply. Nevertheless, I obtained both of these antibodies: the α-Miwi2 antibody from Gregory Hannon's laboratory was provided at too low a concentration for use in IP; the Anti-Miwi2-N1 antibody from Toru Nakano's laboratory was also not functional in IP – there was no band observed in the eluted fraction from ES cells. It is worth noting that neither a stained gel nor a western blot image has ever accompanied published Miwi2 IP data; the Nakano lab reports that merely a faint band is observable on a silver-stained gel following immunoprecipitation of Miwi2 from the murine embryonic testis (Satomi Kuramochi-Miyagawa, personal correspondence). Hence, I had the panel of α-Miwi2 mouse monoclonal antibodies synthesised and two of the twelve antibodies bore the capacity to immunoprecipitate transiently-overexpressed Miwi2-FLAG from HEK293FTs (Figure 5-1C,D). Both antibodies target a single peptide in the Miwi2 protein sequence: this was the only peptide sequence (out of four) against which monoclonal antibodies were raised that could immunoprecipitate Miwi2 in my hands. This highlights the difficulties in successfully selecting a peptide for immunoprecipitation of this Piwi protein. My success in immunoprecipitation of Miwi2 from 293FT-Miwi2-FLAG with these two antibodies gave me permission to use these antibodies in Miwi2-RIP.

Miwi2 binds somatic piRNAs from cluster regions in ES cells, but not in 4T1 cells.

Given that Miwi2 associates with pre-pachytene – and not pachytene – piRNAs in the testis (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008), an unexpected finding from this study is that Miwi2-bound RNAs in ES cells align with selected intergenic regions coinciding with annotated piRNA clusters in the pachytene testis. Low copy number – or even lone – reads present in my datasets can be found in the representative ES cell clusters, a characteristic shared with clustered piRNAs in the pachytene testis (Gan et al., 2011). These large, intergenic clusters in the testis are conserved at selected loci in ES cells and to a lesser extent, at selected loci in 4T1 cells; these regions tend to be repeat-poor. In ES cells, the representative clusters are enriched in the Miwi2-bound fraction, suggesting that they are *bona fide* piRNAs. Miwi2 expression is thought to be restricted to a developmental window coincident with *de novo* methylation in the pre-pachytene

testis, where it associates with a unique population of pre-pachytene piRNAs. The function of these large, intergenic piRNA clusters in ES cells is yet to be characterised, but these data suggest that they have some conserved function in the germline and somatic cells. However, this visual analysis of piRNA clusters in ES cells is limited: a full investigation using clustering algorithms such as F-Seq (Boyle et al., 2008) and bioinformatic comparison of piRNA clusters at *all* loci (rather than visual snapshots) would go some way to determining the degree of conservation in ES cells and the pachytene testis. Given that I have demonstrated that Miwi2 expression in the testis persists into adulthood, it would also be interesting to investigate whether the protein binds pachytene piRNAs in postnatal germ cells to function independently of its role in *de novo* methylation in the embryonic testis.

A smaller proportion of Miwi2-bound piRNAs align with canonical piRNA clusters in unannotated regions in 4T1 cells (38%, Figure 5-12) than in ES cells (87%, Figure 5-4). Furthermore, the large intergenic clusters from the pachytene testis – that are conserved in ES cells – appear to be absent in the Miwi2-bound fraction in 4T1 cells (Figure 5-13), suggesting that Miwi2 functions independently of these clustered piRNAs in cancer. It is unlikely that these piRNAs associate with Miwi or Mili, as transcript levels of these Piwis are negligible in 4T1 cells. Hence, the few small RNA reads from the total small RNA population that align with these cluster regions in 4T1 cells could be non-functional relics from the somatic cells from which the malignancy derived.

A large proportion of annotated Miwi2-bound piRNAs map to repeats

Retrotransposon suppression is the best-described function of nuclear Miwi2 in germ cells, so it was perhaps unexpected that a large percentage of piRNAs bound to *cytoplasmic* Miwi2 are derived from repeat elements in somatic ES cells and 4T1 cells (50% and 36%, respectively). It is known that Piwi proteins and repeat-derived piRNAs function cooperatively during distinct phases of germ cell development to mediate retrotransposon silencing at the transcription and post-transcriptional level (reviewed in (Siomi et al., 2011)). Miwi2 is thought to be highly expressed in the embryonic testis at a developmental window coinciding with *de novo* methylation following global erasure of cytosine methylation in primordial germ cells (Aravin et al., 2008); this epigenetic phenomenon also occurs in preimplantation embryos from which ES cells are derived (Sasaki and Matsui, 2008). Hence, Miwi2 could have a conserved role in methylation of repeat elements in germ cells and ES cells. In the embryonic testis, the majority (76%) of Miwi2-associating piRNAs map to ERVK and L1 retrotransposons (Aravin et al., 2008). The loss of Miwi2 in the embryonic

testis leads to the mobilisation of LTR retrotransposons, attributed to a loss of CpG methylation at the retrotransposon promoter region (Kuramochi-Miyagawa et al., 2008). In my ES cell dataset, Miwi2 associates with RNAs mostly from MaLR and L1 repeats (Figure 5-6), with none mapping to IAP elements as seen in the pachytene testis (Aravin et al., 2008); this is also the case in 4T1 cells. In mouse ES cells, Suv39 enzymes and DNA methyltransferases (DNMTs) are required for repression of L1 retrotransposons via DNA and H3K9 methylation (Matsui et al., 2010). Pezic and colleagues recently proposed a role for Miwi2 in germ cells in the deposition of repressive histone methylation marks at H3K9 at the 5'UTR of L1 elements (Pezic et al., 2014). Miwi2 could somehow be involved in the recruitment of DNA and histone methyltransferases to mediate transcriptional silencing of retrotransposons, although comprehensive analysis of methylation profiles in Miwi2-null ES cells would be required to address this possibility. Given that Miwi2-knockout mice were generated from heterozygous Miwi2 ES cells (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008), this analysis has not been performed to date, or at least has not been reported in the literature. However, Miwi2 is expressed predominately in the nucleus of male germ cells (Aravin et al., 2008), but in the cytoplasm of ES and 4T1 cells. It is expected that Miwi2 would need to be expressed in the nucleus in ES cells and 4T1 cells to mediate a transcriptional silencing function in the same manner as the fetal testis. But it is possible that Miwi2 is transiently localised to the nucleus, or more likely, somehow activates a secondary response similar to the endo-siRNA signalling cascade in *C. elegans,* which is activated by cytoplasmic Piwi proteins to mediate transcriptional silencing of transposons (Das et al., 2008, Bagijn et al., 2012). Immunoprecipitation of Miwi2 from fractionated ES cells to determine whether piRNAs are detectable in the nuclear fraction would go some way to addressing whether somatic Miwi2 has any function in transposon silencing at the level of transcription.

Miwi2 binds gene-derived small RNAs in somatic cells

There is an abundance of gene-derived small RNAs in the Miwi2-bound fractions of ES and 4T1 cells. It could be argued that these represent degraded fragments of mRNAs, although typically, mammalian mRNAs are protected from exonuclease digestion by the 5' cap (Walther et al., 1998, Sachs, 1993) and the Poly A-binding protein (PABP) associated with the polyadenylated tail at the 3' end (Sachs, 1993, Ford et al., 1997, Coller et al., 1998, Wang et al., 1999). It is unclear whether 5'-3' or 3'-5' degradation predominates in mammalian cells (Grudzien et al., 2006). In several instances, I observed a bias for small RNA clusters at the 3'UTR of genes: this could be

representative of 5'-3' degradation. However, 3'UTR piRNA clusters have been reported by others (Robine et al., 2009, Gan et al., 2011). Furthermore, I saw no correlation between the abundance of small RNAs and that of their parent genes (Figure 5-7 and Figure 5-15), indicating that these are unlikely to be degraded fragments of the most abundant genes, and that this is, in fact, a biologically-relevant finding.

Genes producing piRNAs in ES cells and 4T1 piRNAs cluster in common ontologies, particularly cellular movement, cell-to-cell signalling, and developmental processes. My data reveal the clustering of Miwi2-bound piRNAs at the 3'UTR of many genes in both ES and 4T1 cells. This finding is consistent with reports of piRNA clusters aligning with the 3'UTR of specific genes in *Drosophila* ovary somatic sheet cells, mouse testis, and *Xenopus* oocytes (Robine et al., 2009). In ES cells, 6% of total small RNAs map to a large number (495) of genes, whereas Miwi2 IP enriches for genic piRNAs (39% of annotated reads) derived from a smaller number (172) of genes; hence, the reads are more densely clustered across the genes. Notably, in the Miwi2-enriched fraction of ES cells, piRNAs cluster at the 3'UTR of the master pluripotency genes, *Oct4, Sox2,* and *Nanog* (Figure 5-9). This is an intriguing finding, as the 3'UTR is indispensable for transcript stability, localisation and translation (reviewed in (Di Giammartino et al., 2011)). The role of these 3'UTR piRNA clusters is yet to be characterised, but given that clusters are enriched in genes involved in the earliest stages of lineage specification, Miwi2 and its bound piRNAs could have some role in regulating their expression.

Despite reports that 35% of all small RNAs are derived from 3'UTRs in the mouse 10dpp testis (Aravin et al., 2008, Robine et al., 2009), my data show that selected 3'UTR clusters are enriched in ES and 4T1 cells, which are mostly absent in the pachytene testis. One exception is at the Argonaute gene, *Ago2*: Miwi2 enriches for a cluster in 4T1 cells that is also present in the testis dataset (Figure 5-16). This is an unusual piRNA cluster in that extends downstream of the 3'UTR. Unpublished cluster analysis data from David Martin's laboratory at the Children's Hospital Oakland Research Institute (CA, USA) shows a similar phenomenon in the adult mouse testis and somatic tissues, and that a subset of these are dependent on Miwi2 (David Martin, personal correspondence). There are many examples of extended 3'UTR clustering of piRNAs in 4T1 cells and these are enriched by Miwi2 immunoprecipitation. These extended 3'UTR piRNAs are defined as unannotated in my analysis, hence likely comprise some of the unannotated reads in the Miwi2-bound fraction in 4T1 cells. It is possible that these downstream intergenic regions are, in fact,

unannotated 3'UTRs. 3'UTR length is dynamic depending on cell type: they tend to be shorter in proliferating cells, such as cancer cells and the testis, and during iPS cell reprogramming (Mayr and Bartel, 2009, Smibert et al., 2012, Ji et al., 2009). This is consistent with my finding from visual inspection of UCSC Genome Browser that these extended 3'UTR piRNA clusters are not prevalent at my chosen loci in ES cells.

There is evidence that 3'UTR piRNA clusters can regulate expression of their parent genes, as expression of *Traffic Jam* – whose 3'UTR generate large numbers of piRNAs in *Drosophila* – is increased in *Piwi* mutant follicle clones (Robine et al., 2009). The mechanism for Miwi2/piRNA regulation of gene expression – if it exists – is currently unclear. My data show that Miwi2-bound piRNAs map to the sense strand of protein-coding genes in ES cells and 4T1 cells, independent of the parent transcript expression level. This observation is consistent with a role in post-transcriptional regulation, such as that which occurs during the maternal-to-zygotic transition in *Drosophila*, where Aub and Ago3 are required for the degradation of maternal mRNAs (Rouget et al., 2010).

piRNAs also could also target other genes *in trans* based on sequence complementarity to their 3'UTRs. For example, those small RNAs with potential binding sites in the 3'UTR of *Nanos* are required to regulate its expression levels during posterior patterning in the *Drosophila* embryo via recruitment of the deadenylation machinery and mRNA destabilisation (Cox et al., 1998, Lehmann and Nusslein-Volhard, 1991, Rouget et al., 2010). Another group has isolated Miwi-bound RNAs from elongating spermatids and, in addition to 29-31nt piRNAs, identified larger protein-coding transcripts (>200nt) associating with the Piwi protein. Furthermore, they identified binding sites for the pi-sized RNAs in the 3'UTRs of Miwi-bound mRNAs and provided evidence that piRNAs regulate gene expression in a miRNA-like fashion (Gou et al., 2014). This was the first study to report the association of protein-coding transcripts with Piwi proteins; this is probably because previous studies – like mine – enriched sequencing libraries for small RNAs smaller than 50nt. The 3'UTR piRNAs in my dataset tend to be derived from the sense strand of genes and are therefore unlikely to regulate gene expression in the same manner, although it is possible that some repeat-derived piRNAs in my dataset have gene targets. Target prediction software could address this, although in contrast to miRNA:target interactions, definitive parameters for piRNA targeting have yet to be defined.

Miwi2 binds tRNAs in 4T1 cells

The finding that Miwi2 binds processed tRFs suggests that this pathway could also function in translation. Studies from our laboratory have shown that Hiwi2 physically associates with actively translating ribosomes and binds processed tRFs in 231 human breast cancer cells (Keam et al., 2014). These tRF-piRNAs make up the majority (67%) of the Hiwi2-bound fraction in 231 cells and are processed preferentially from 9 tRNA isotypes representing five amino acids: glycine, glutamine, lysine, aspartic acid and valine. All but one are processed from the 5' end of the mature tRNA and are distinguished from stress-induced tRNA halves by length. In 4T1 cells, tRFs occupied a smaller proportion (12%, Figure 5-10) of all annotated Miwi2-bound RNAs, and these were predominately derived from the 5' end of tRNAs representing four amino acids: glycine, lysine, glutamine and histidine (Figure 5-17).

There is evidence that tRFs act in the repression of translation (Thompson et al., 2008, Sobala and Hutvagner, 2013). Furthermore, Hiwi2 associates with eukaryotic elongation factor 1-alpha (eEF1α) and the heat shock proteins in 231 cells, proteins known to be involved in translation (Keam et al., 2014). Furthermore, Mili has been shown to interact with eukaryotic initiation factor 3a (eIF3a) in the testis (Unhavaithaya et al., 2009). In fact, both Mili and Miwi have been linked to translational regulation in developing germ cells (Grivna et al., 2006, Vourekas et al., 2012). Furthermore, antisense suppression of piRNAs in mouse hippocampal neurons interferes with the extension of dendritic spines, indicative of dysfunctional protein translation (Lee et al., 2011).

Prior to the publication of the recent study by our laboratory (Keam et al., 2014), there was only one other report of tRF-piRNAs: 18-22nt tRFs bind to the growth-essential Piwi protein, Twi12, in *Tetrahymena thermophila* (Couvillion et al., 2010). Their function, however, remains to be determined. In the archaeon *Haloferax volcanii*, pi-sized (26nt) 5' tRFs bind to actively translating ribosomes and have been shown to inhibit translation in response to stress (Gebetsberger et al., 2012). It is unknown whether these archaeal tRFs are bound by archaeal Piwi proteins in *H. volcanii*, although the data from Couvillion et al and Keam et al suggest that this is a possibility. The binding partners of tRFs are largely uncharacterised in mammals, although Cole and colleagues have demonstrated that mammalian 5' tRFs have a weak association with the classic Ago proteins (Cole et al., 2009). This thesis, and the recently-published study from our laboratory (Keam et al., 2014), are the first reports of an association between tRFs and a Piwi protein in mammals.

It is also possible that Miwi2 is linked to the regulation of translation in the germline, as several of the tRFs present in the Miwi2-bound fractions in 4T1 cells are also present in the 20.5dpp testis, such as the LysCTT- and GlyGCC-derived tRFs, albeit at a much lower level (Figure 8-12, Appendix). My data, along with that from the aforementioned studies, suggests that Piwi/piRNA-mediated inhibition of translation could be an ancient and conserved function for the pathway.

Conclusions

The finding that Miwi2 binds piRNAs in somatic cells is unprecedented: all other studies of Miwi2/piRNA function have been performed in germ cells. In ES cells, Miwi2 associates with piRNAs produced from large clusters at intergenic loci that are recognised as canonical piRNA clusters in the pachytene testis. Repeat-derived RNAs commonly associate with Miwi2, but considering Miwi2 is cytoplasmic in ES and 4T1 cells, it is unclear whether the germline function of Miwi2 in the transcriptional silencing of transposons is conserved in somatic cells. Miwi2-bound piRNAs also cluster at the 3'UTR of selected genes involved in cell proliferation, signalling and development, such as key pluripotency genes, in ES and 4T1 cells. 3'UTR regulation is critical for transcript stability, localisation, and translation, implying a role for Miwi2/piRNAs in the regulation of expression of these genes in somatic cells. Furthermore, I have observed the association of Miwi2 with tRF-piRNAs, implying a link to translation in 4T1 cells. It is also possible that Piwi/piRNA function in the regulation of translation is also required in the germline, given that gene-derived piRNAs and tRFs are also expressed in the testis. Until now, this function has likely been overlooked, given the predominance of repeat-derived piRNAs and large intergenic clusters. The data presented in this chapter have raised some intriguing possibilities for an alternative function of somatic Miwi2 to its well-characterised role in germline transposon silencing. Further analysis of piRNA origin and targeting in Miwi2-knockout ES and cancer cells could contribute to the burgeoning field of somatic Piwi gene and protein research.

6. Future directions and conclusions

In mammals, action of the Piwi/piRNA pathway has been well-described, but only in the context of germline transposon silencing. Little is known about piRNAs that are not derived from transposable elements and other repetitive sequences, or the ping pong-independent pachytene piRNAs. Furthermore, expression of the effector Piwi proteins was, for some time, thought to be limited to germ cells. However, reports are emerging of Piwi gene and protein expression in mammalian somatic cells, cancer cell lines, and solid tumours: tissues postulated to contain a resident stem cell population (reviewed in (Ross et al., 2014)). A function for Piwi genes and proteins – and the presence of piRNAs – has been described in the stem-like somatic cells in primitive metazoans (Reddien et al., 2005, Palakodeti et al., 2008, Alie et al., 2011, Funayama et al., 2010, Rinkevich et al., 1995). Conversely, there have been few functional studies of Piwi genes and proteins in mammalian somatic stem cells. Thus, I investigated the hypothesis that the murine Piwi proteins are involved in maintaining stem-like cells in somatic tissue and cancer.

I set out to answer several questions to address this hypothesis:

1. Are the murine Piwi genes and proteins expressed in embryonic stem cells and cancer cells?
2. Does *Miwi2* knockdown affect the self-renewal or pluripotency of embryonic stem cells?
3. Does ectopic expression of *Miwi2* affect growth properties of non-metastatic 4TO7 cancer cells?
4. Is global gene and small RNA expression affected by manipulating *Miwi2* expression in these two cell types?
5. Does Miwi2 bind piRNAs in ES cells and 4T1 cancer cells and, if so, from where in the genome are they derived?

In Chapter 3 of this thesis, I showed that – while *Miwi* and *Mili* are negligibly expressed – *Miwi2* is present in ES cells at levels comparable to the testis, and it is translated into protein. Moreover, gene expression diminishes concurrent with differentiation. But, unlike in primitive metazoans such as *S. mediterranea* (Reddien et al., 2005, Palakodeti et al., 2008), it does not appear that Piwi gene expression is essential for somatic stem cell function. While Piwi genes and proteins are indispensable for somatic stem cell function in primitive organisms, there are phylogenetic differences between Piwi orthologues (particularly between bilateral and non-bilateral animals); thus, it is feasible that they could have diverse and independent functions (Seto et al., 2007, Saito

and Siomi, 2010). There were gene expression perturbations with *Miwi2* knockdown, but overall, *Miwi2*-knockdown ES cells largely retain their self-renewal properties and appear to differentiate normally *in vitro*. Other groups have shown that Miwi2-knockout mice are viable and, apart from sterility, appear to develop normally (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008). In fact, knockout of all three Piwi proteins yields viable mice with normal somatic stem cell function, at least in the haematopoietic system (Nolde et al., 2013). However, it is currently unknown whether homeostatic and regenerative functions are impaired in *aged* Miwi2-null mice. In *C. elegans* with a mutation in a Piwi protein, PRG-1, germline defects are only apparent after many generations (Simon et al., 2014). Given that Miwi2-null males are sterile, it is impossible to conduct this experiment in mice, but it is possible that long-term culture of Miwi2-knockout ES cells (generated with emerging technologies such as CRISPR) could reveal some latent, deleterious phenotype.

However, the presence of Miwi2 in ES cells, and the fact that it binds piRNAs, suggests that the Piwi/piRNA pathway has some non-essential role in stem cells that could predate its germline function. In selected instances, Miwi2 binds piRNAs produced from large intergenic clusters that are also present in the testis, suggesting that they are *bona fide* piRNAs. This is a curious finding: Miwi2 is thought to only associate with pre-pachytene piRNAs in the testis (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008), but large intergenic clusters are predominately expressed in pachytene germ cells (Gan et al., 2011). Given that I have shown Miwi2 expression in the testis persists into adulthood, it would be informative to investigate whether this Miwi2 protein binds piRNAs derived from intergenic clusters in the pachytene testis. Pachytene clusters in mammals do not yet have defined antisense regulatory targets, and are produced independently of ping pong biogenesis (Gan et al., 2011); this also appears to be the case for piRNAs in ES cells. The clusters that I have shown are not as densely populated in ES cells as the pachytene testis, but given that many of these piRNAs exhibited low copy number (or were present as single reads), greater sequencing depth in future experiments could reveal greater cluster density. The function of these selected intergenic clusters is unclear in both germ cells and ES cells, but could somehow be involved in regulating essential stem cell functions. Certain intergenic regions do exhibit regulatory function: an interacting network of long intergenic RNAs has been shown to be essential for pluripotency (Guttman et al., 2011). With the continued improvement of the sensitivity and scope

of analytical technologies such as next-generation sequencing, it is likely that more will be revealed about this complex pathway in further studies from my laboratory and others.

It could be that the intergenic clusters I have highlighted are representative of some conserved function of Piwi/piRNA in the germline and soma. It is unlikely that piRNA clusters in the pachytene testis are relics from pluripotent cells, as these clusters are not known to exist in the pre-pachytene testis. Alternatively, these piRNAs could be transmitted from the germline to soma during fertilisation. Such a phenomenon has been described in *Drosophila*, where female piRNAs in the oocyte are transmitted to the developing zygote (Shpiz and Kalmykova, 2009), although the function is unclear. Furthermore, piRNAs have been identified in mouse sperm, where they have been implicated in the transgenerational inheritance of trauma-induced behavioural and metabolic defects (Gapp et al., 2014). A future experiment could test whether any intergenic clusters are syntenic in the pachytene testis, the sperm from age-matched mice, and ES cells derived from blastocysts of the same mouse strain. If this hypothesis was true, it would be expected that earlier-stage (such as 2C) embryos and epiblast stem cells would also express these piRNA clusters; they may also bind Miwi2. Clustering software (such as F-seq) would also allow a more thorough investigation of intergenic piRNA clustering in these cell types.

Another intriguing finding from the ES cell study was that, in some instances, Miwi2-bound piRNAs cluster at the 3'UTRs of key pluripotency genes, and are absent from the 3'UTR of select lineage markers. In male germ cells, piRNAs are produced from the 3'UTR of genes involved in germline development (Robine et al., 2009). Both my study and that of Robine and colleagues revealed no correlation between piRNA abundance and parent gene expression, suggesting that piRNAs are not produced indiscriminately from the most abundant transcripts in the cells. Therefore, it is entirely possible that the Piwi/piRNA is involved in the tissue-specific regulation of gene expression. A mechanism is, as yet, unclear, but these piRNAs could function at the level of translation. Overall, most Miwi2-bound piRNAs in ES cells are derived from introns, but those that are produced from these particular clusters appear to be produced from spliced transcripts: clustered reads preferentially map to UTRs and coding regions. Furthermore, the fact that Miwi2 is cytoplasmic in ES cells implies that any role in gene regulation would likely be at the post-transcriptional level. If Miwi2 does act at the level of translation, it could explain why no large-scale changes in transcript expression were seen *Miwi2*-knockdown ES cells. In the future, I could perform proteomic analysis on these cells to investigate whether translation is affected. Polysome

profiling would also go some way to addressing whether these Miwi2-bound piRNAs are produced from genes associated with actively-translating ribosomes.

In Chapter 4, I showed variable *Miwi2* expression in a syngenic panel of breast cancer cell lines. Gene expression is enriched in highly metastatic 4T1 cells, and is negligible in non-metastatic 4TO7 cells. I was unable to achieve knockdown 4T1 cells – this could be a technical difficulty owing to doxycycline resistance, or could be indicative of a requirement for Miwi2 in these cells. Constitutive (rather than inducible) knockdown of *Miwi2*, or knockout using CRISPR, may resolve this uncertainty.

Given that the knockdown experiment was unsuccessful, I instead introduced *Miwi2* into non-expressing 4TO7 cells. While no overt phenotypic effects arose, I observed expression changes in genes and small RNAs with ontologies in metastasis-related processes. Furthermore, I showed that ectopic *Miwi2* results in increased expression of several genes that are normally enriched in 4T1 cells relative to non-metastatic syngenic cell lines. However, a broad limitation to this study is that I have focused merely on *in vitro* effects of *Miwi2* expression; the study of *in vivo* behaviour of 4TO7-Miwi2 cells would be required to unequivocally determine whether *Miwi2* confers tumorigenic properties to otherwise non-metastatic cells. When implanted into the mammary fat pad of mice, wild-type 4TO7 cells will form a tumour at the primary location and will migrate to the lungs, but appear to be unable to complete the final stage of metastasis as secondary tumours do not form (Aslakson and Miller, 1992). If ectopic *Miwi2* confers 4T1-like metastatic properties to 4TO7 cells, it is possible that these cells would be able to undergo the metastatic growth phase at distant locations from the primary tumour.

Immunoprecipitation of endogenous Miwi2 from 4T1 cells revealed by visual examination that – unlike in ES cells – the Piwi protein tends not associate with piRNAs produced from large intergenic clusters, suggesting that this particular function of Miwi2 – whatever it may be – is not active in cancer. Instead, Miwi2 IP from 4T1 cells enriches for piRNAs derived from clusters at the 3'UTR of selected genes. Further investigation into genic piRNAs in 4T1 cells could utilise clustering algorithms (such as F-seq) to determine the proportion of genes producing *bona fide* piRNA clusters. This finding is in common with the Miwi2 IP from ES cells, and could indicate that Miwi2 is involved in gene regulation. This hypothesis is further supported by the discovery that Miwi2 binds tRF-piRNAs, as tRFs have previously been implicated in the regulation of gene expression in cancer

cells (Lee et al., 2009) and the inhibition of translation (Thompson et al., 2008, Sobala and Hutvagner, 2013). Antisense suppression of the most abundant tRFs might provide insight into the function of these tRF-piRNAs. Although the relationship between the gene-derived piRNAs and tRFs is currently unclear, these results point to a *bona fide* role of the Piwi/piRNA pathway in the regulation of translation in cancer cells.

I commenced this study with the intention of describing a function for the Piwi/piRNA pathway in the stem-like cells of normal tissue and malignancies, and my data suggest that Miwi2 has a non-essential function in ES cells and metastatic breast cancer cells. Although the precise function of Miwi2 in somatic cells is yet to be understood, this thesis reveals a broader scope of action for the Piwi/piRNA pathway in metazoans and provides a foundation for further functional studies.

7. List of references

ABDEEN, S. K., SALAH, Z., KHAWALED, S. & AQEILAN, R. I. 2013. Characterization of WWOX inactivation in murine mammary gland development. *J Cell Physiol,* 228**,** 1391-6.

ABDEEN, S. K., SALAH, Z., MALY, B., SMITH, Y., TUFAIL, R., ABU-ODEH, M., ZANESI, N., CROCE, C. M., NAWAZ, Z. & AQEILAN, R. I. 2011. Wwox inactivation enhances mammary tumorigenesis. *Oncogene,* 30**,** 3900-6.

ACAMPORA, D., DI GIOVANNANTONIO, L. G. & SIMEONE, A. 2013. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development,* 140**,** 43-55.

ALDER, H. & SCHMID, V. 1987. Cell cycles and in vitro transdifferentiation and regeneration of isolated, striated muscle of jellyfish. *Dev Biol,* 124**,** 358-69.

ALI, S. H. & DECAPRIO, J. A. 2001. Cellular transformation by SV40 large T antigen: interaction with host proteins. *Semin Cancer Biol,* 11**,** 15-23.

ALIE, A., LECLERE, L., JAGER, M., DAYRAUD, C., CHANG, P., LE GUYADER, H., QUEINNEC, E. & MANUEL, M. 2011. Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: ancient association of "germline genes" with stemness. *Dev Biol,* 350**,** 183-97.

AMIT, M., CARPENTER, M. K., INOKUMA, M. S., CHIU, C. P., HARRIS, C. P., WAKNITZ, M. A., ITSKOVITZ-ELDOR, J. & THOMSON, J. A. 2000. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol,* 227**,** 271-8.

ARAVIN, A., GAIDATZIS, D., PFEFFER, S., LAGOS-QUINTANA, M., LANDGRAF, P., IOVINO, N., MORRIS, P., BROWNSTEIN, M. J., KURAMOCHI-MIYAGAWA, S., NAKANO, T., CHIEN, M., RUSSO, J. J., JU, J., SHERIDAN, R., SANDER, C., ZAVOLAN, M. & TUSCHL, T. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature,* 442**,** 203-7.

ARAVIN, A. A. & BOURC'HIS, D. 2008. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev,* 22**,** 970-5.

ARAVIN, A. A., HANNON, G. J. & BRENNECKE, J. 2007a. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science,* 318**,** 761-4.

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

ARAVIN, A. A., SACHIDANANDAM, R., BOURC'HIS, D., SCHAEFER, C., PEZIC, D., TOTH, K. F., BESTOR, T. & HANNON, G. J. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell,* 31**,** 785-99.

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

ASHE, A., SAPETSCHNIG, A., WEICK, E. M., MITCHELL, J., BAGIJN, M. P., CORDING, A. C., DOEBLEY, A. L., GOLDSTEIN, L. D., LEHRBACH, N. J., LE PEN, J., PINTACUDA, G., SAKAGUCHI, A., SARKIES, P., AHMED, S. & MISKA, E. A. 2012. piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C. elegans. *Cell,* 150**,** 88-99.

ASLAKSON, C. J. & MILLER, F. R. 1992. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res,* 52**,** 1399-405.

ATTIA, M., RACHEZ, C., DE PAUW, A., AVNER, P. & ROGNER, U. C. 2007. Nap1l2 promotes histone acetylation activity during neuronal differentiation. *Mol Cell Biol,* 27**,** 6093-102.

BABIARZ, J. E., RUBY, J. G., WANG, Y., BARTEL, D. P. & BLELLOCH, R. 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev,* 22**,** 2773-85.

BAGIJN, M. P., GOLDSTEIN, L. D., SAPETSCHNIG, A., WEICK, E. M., BOUASKER, S., LEHRBACH, N. J., SIMARD, M. J. & MISKA, E. A. 2012. Function, targets, and evolution of Caenorhabditis elegans piRNAs. *Science,* 337**,** 574-8.

BAGUNA, J., SALO. E., AND AULADELL, C. 1989. Regeneration and pattern formation in planarians. III. that neoblasts are totipotent stem cells and the cells. *Development,* 107**,** 77-86.

BAO, L., HAQUE, A., JACKSON, K., HAZARI, S., MOROZ, K., JETLY, R. & DASH, S. 2011. Increased expression of P-glycoprotein is associated with doxorubicin chemoresistance in the metastatic 4T1 breast cancer model. *Am J Pathol,* 178**,** 838-52.

BATISTA, P. J., RUBY, J. G., CLAYCOMB, J. M., CHIANG, R., FAHLGREN, N., KASSCHAU, K. D., CHAVES, D. A., GU, W., VASALE, J. J., DUAN, S., CONTE, D., JR., LUO, S., SCHROTH, G. P., CARRINGTON, J. C., BARTEL, D. P. & MELLO, C. C. 2008. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell,* 31**,** 67-78.

BATLLE, E., HENDERSON, J. T., BEGHTEL, H., VAN DEN BORN, M. M., SANCHO, E., HULS, G., MEELDIJK, J., ROBERTSON, J., VAN DE WETERING, M., PAWSON, T. & CLEVERS, H. 2002. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell,* 111**,** 251-63.

BERNINGER, P., JASKIEWICZ, L., KHORSHID, M. & ZAVOLAN, M. 2011. Conserved generation of short products at piRNA loci. *BMC Genomics,* 12**,** 46.

BEYRET, E., LIU, N. & LIN, H. 2012. piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism. *Cell Res,* 22**,** 1429-39.

BHARDWAJ, G., MURDOCH, B., WU, D., BAKER, D. P., WILLIAMS, K. P., CHADWICK, K., LING, L. E., KARANU, F. N. & BHATIA, M. 2001. Sonic hedgehog induces the proliferation of primitive human hematopoietic cells via BMP regulation. *Nat Immunol,* 2**,** 172-80.

BHATIA, M., WANG, J. C., KAPP, U., BONNET, D. & DICK, J. E. 1997. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci U S A,* 94**,** 5320-5.

BIECHELE, T. L., KULIKAUSKAS, R. M., TORONI, R. A., LUCERO, O. M., SWIFT, R. D., JAMES, R. G., ROBIN, N. C., DAWSON, D. W., MOON, R. T. & CHIEN, A. J. 2012. Wnt/beta-catenin signaling and AXIN1 regulate apoptosis triggered by inhibition of the mutant kinase BRAFV600E in human melanoma. *Sci Signal,* 5**,** ra3.

BONNET, D. & DICK, J. E. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med,* 3**,** 730-7.

BOSSE, M. D. 1940. Rhabdomyosarcomatous Pulmonary Metastases from a Teratoma Testis. *Cancer Research,* 39**,** 343.

BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

BOYLE, A. P., GUINNEY, J., CRAWFORD, G. E. & FUREY, T. S. 2008. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics,* 24**,** 2537-8.

BRENNECKE, J., ARAVIN, A. A., STARK, A., DUS, M., KELLIS, M., SACHIDANANDAM, R. & HANNON, G. J. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell,* 128**,** 1089-103.

BRIDGE, A. J., PEBERNARD, S., DUCRAUX, A., NICOULAZ, A. L. & IGGO, R. 2003. Induction of an interferon response by RNAi vectors in mammalian cells. *Nat Genet,* 34**,** 263-4.

BROWER-TOLAND, B., FINDLEY, S. D., JIANG, L., LIU, L., YIN, H., DUS, M., ZHOU, P., ELGIN, S. C. & LIN, H. 2007. Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev,* 21**,** 2300-11.

BROWN, B. D., SITIA, G., ANNONI, A., HAUBEN, E., SERGI, L. S., ZINGALE, A., RONCAROLO, M. G., GUIDOTTI, L. G. & NALDINI, L. 2007. In vivo administration of lentiviral vectors triggers a type I interferon response that restricts hepatocyte gene transfer and promotes vector clearance. *Blood,* 109**,** 2797-805.

BROWN, F. D., KEELING, E. L., LE, A. D. & SWALLA, B. J. 2009. Whole body regeneration in a colonial ascidian, Botrylloides violaceus. *J Exp Zool B Mol Dev Evol,* 312**,** 885-900.

BUTCHER, S. E. & BROW, D. A. 2005. Towards understanding the catalytic core structure of the spliceosome. *Biochem Soc Trans,* 33**,** 447-9.

CALDERWOOD, D. A., SHATTIL, S. J. & GINSBERG, M. H. 2000. Integrins and actin filaments: reciprocal regulation of cell adhesion and signaling. *J Biol Chem,* 275**,** 22607-10.

CARMELL, M. A., GIRARD, A., VAN DE KANT, H. J., BOURC'HIS, D., BESTOR, T. H., DE ROOIJ, D. G. & HANNON, G. J. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell,* 12**,** 503-14.

CASTANEDA, J., GENZOR, P. & BORTVIN, A. 2011. piRNAs, transposon silencing, and germline genome integrity. *Mutat Res,* 714**,** 95-104.

CHAMBERS, I., COLBY, D., ROBERTSON, M., NICHOLS, J., LEE, S., TWEEDIE, S. & SMITH, A. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell,* 113**,** 643-55.

CHAMBERS, I., SILVA, J., COLBY, D., NICHOLS, J., NIJMEIJER, B., ROBERTSON, M., VRANA, J., JONES, K., GROTEWOLD, L. & SMITH, A. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature,* 450**,** 1230-4.

CHAMBERS, I. & SMITH, A. 2004. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene,* 23**,** 7150-60.

CHAWLA, G. & SOKOL, N. S. 2014. ADAR mediates differential expression of polycistronic microRNAs. *Nucleic Acids Res,* 42**,** 5245-55.

CHAZAUD, C., YAMANAKA, Y., PAWSON, T. & ROSSANT, J. 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell,* 10**,** 615-24.

CHEN, C., LIU, J. & XU, G. 2013. Overexpression of PIWI proteins in human stage III epithelial ovarian cancer with lymph node metastasis. *Cancer Biomark,* 13**,** 315-21.

CHEN, L., SHEN, R., YE, Y., PU, X. A., LIU, X., DUAN, W., WEN, J., ZIMMERER, J., WANG, Y., LIU, Y., LASKY, L. C., HEEREMA, N. A., PERROTTI, D., OZATO, K., KURAMOCHI-MIYAGAWA, S., NAKANO, T., YATES, A. J., CARSON, W. E., 3RD, LIN, H., BARSKY, S. H. & GAO, J. X. 2007. Precancerous stem cells have the potential for both benign and malignant differentiation. *PLoS One,* 2**,** e293.

CHENG, J., GUO, J. M., XIAO, B. X., MIAO, Y., JIANG, Z., ZHOU, H. & LI, Q. N. 2011. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clin Chim Acta,* 412**,** 1621-5.

COLE, C., SOBALA, A., LU, C., THATCHER, S. R., BOWMAN, A., BROWN, J. W., GREEN, P. J., BARTON, G. J. & HUTVAGNER, G. 2009. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA,* 15**,** 2147-60.

COLLER, J. M., GRAY, N. K. & WICKENS, M. P. 1998. mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev,* 12**,** 3226-35.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

COUVILLION, M. T., SACHIDANANDAM, R. & COLLINS, K. 2010. A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. *Genes Dev,* 24**,** 2742-7.

COX, D. N., CHAO, A., BAKER, J., CHANG, L., QIAO, D. & LIN, H. 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev,* 12**,** 3715-27.

COX, D. N., CHAO, A. & LIN, H. 2000. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development,* 127**,** 503-14.

CREIGHTON, C. J., REID, J. G. & GUNARATNE, P. H. 2009. Expression profiling of microRNAs by deep sequencing. *Brief Bioinform,* 10**,** 490-7.

CRIPPA, S., CASSANO, M., MESSINA, G., GALLI, D., GALVEZ, B. G., CURK, T., ALTOMARE, C., RONZONI, F., TOELEN, J., GIJSBERS, R., DEBYSER, Z., JANSSENS, S., ZUPAN, B., ZAZA, A., COSSU, G. & SAMPAOLESI, M. 2011. miR669a and miR669q prevent skeletal muscle differentiation in postnatal cardiac progenitors. *J Cell Biol,* 193**,** 1197-212.

DALERBA, P., DYLLA, S. J., PARK, I. K., LIU, R., WANG, X., CHO, R. W., HOEY, T., GURNEY, A., HUANG, E. H., SIMEONE, D. M., SHELTON, A. A., PARMIANI, G., CASTELLI, C. & CLARKE, M. F. 2007. Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A,* 104**,** 10158-63.

DAMESHEK, W. 1951. Some speculations on the myeloproliferative syndromes. *Blood,* 6**,** 372-5.

DAS, P. P., BAGIJN, M. P., GOLDSTEIN, L. D., WOOLFORD, J. R., LEHRBACH, N. J., SAPETSCHNIG, A., BUHECHA, H. R., GILCHRIST, M. J., HOWE, K. L., STARK, R., MATTHEWS, N., BEREZIKOV, E., KETTING, R. F., TAVARE, S. & MISKA, E. A. 2008. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol Cell,* 31**,** 79-90.

DAVIS, H. E., MORGAN, J. R. & YARMUSH, M. L. 2002. Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. *Biophys Chem,* 97**,** 159-72.

DE FAZIO, S., BARTONICEK, N., DI GIACOMO, M., ABREU-GOODGER, C., SANKAR, A., FUNAYA, C., ANTONY, C., MOREIRA, P. N., ENRIGHT, A. J. & O'CARROLL, D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature,* 480**,** 259-63.

DE MULDER, K., PFISTER, D., KUALES, G., EGGER, B., SALVENMOSER, W., WILLEMS, M., STEGER, J., FAUSTER, K., MICURA, R., BORGONIE, G. & LADURNER, P. 2009. Stem cells are differentially regulated during development, regeneration and homeostasis in flatworms. *Dev Biol,* 334**,** 198-212.

DE SUTTER, D. A. V. D. V., G. 1977. Aggregative properties of different cell types of the fresh-water sponge Ephydatia fluviatilis isolated on ficoll gradients. *Roux's Archives of Developmental Biology,* 181**,** 151-161.

DENG, W. & LIN, H. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell,* 2**,** 819-30.

DENKER, E., MANUEL, M., LECLERE, L., LE GUYADER, H. & RABET, N. 2008. Ordered progression of nematogenesis from stem cells through differentiation stages in the tentacle bulb of Clytia hemisphaerica (Hydrozoa, Cnidaria). *Dev Biol,* 315**,** 99-113.

DHAHBI, J. M., SPINDLER, S. R., ATAMNA, H., YAMAKAWA, A., BOFFELLI, D., MOTE, P. & MARTIN, D. I. 2013. 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics,* 14**,** 298.

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

DICK, J. E. 2009. Looking ahead in cancer stem cell research. *Nat Biotechnol,* 27**,** 44-6.

DJIKENG, A., FERREIRA, L., D'ANGELO, M., DOLEZAL, P., LAMB, T., MURTA, S., TRIGGS, V., ULBERT, S., VILLARINO, A., RENZI, S., ULLU, E. & TSCHUDI, C. 2001. Characterization of a candidate Trypanosoma brucei U1 small nuclear RNA gene. *Mol Biochem Parasitol,* 113**,** 109-15.

DRABEK, K., GUTIERREZ, L., VERMEIJ, M., CLAPES, T., PATEL, S. R., BOISSET, J. C., VAN HAREN, J., PEREIRA, A. L., LIU, Z., AKINCI, U., NIKOLIC, T., VAN IJCKEN, W., VAN DEN HOUT, M., MEINDERS, M., MELO, C., SAMBADE, C., DRABEK, D., HENDRIKS, R. W., PHILIPSEN, S., MOMMAAS, M., GROSVELD, F., MAIATO, H., ITALIANO, J. E., JR., ROBIN, C. & GALJART, N. 2012. The microtubule plus-end tracking protein CLASP2 is required for hematopoiesis and hematopoietic stem cell maintenance. *Cell Rep,* 2**,** 781-8.

DUNN, G. P., KOEBEL, C. M. & SCHREIBER, R. D. 2006. Interferons, immunity and cancer immunoediting. *Nat Rev Immunol,* 6**,** 836-48.

EDEN, E., NAVON, R., STEINFELD, I., LIPSON, D. & YAKHINI, Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics,* 10**,** 48.

ELLISEN, L. W., BIRD, J., WEST, D. C., SORENG, A. L., REYNOLDS, T. C., SMITH, S. D. & SKLAR, J. 1991. TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell,* 66**,** 649-61.

ENCODE PROJECT CONSORTIUM, M. R., STAMATOYANNOPOULOS J, SNYDER M, DUNHAM I, HARDISON RC, BERNSTEIN BE, GINGERAS TR, KENT WJ, BIRNEY E ET AL. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol,* 9**,** e1001046.

ENDER, C., KREK, A., FRIEDLANDER, M. R., BEITZINGER, M., WEINMANN, L., CHEN, W., PFEFFER, S., RAJEWSKY, N. & MEISTER, G. 2008. A human snoRNA with microRNA-like functions. *Mol Cell,* 32**,** 519-28.

ESTELLER, M. 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet,* 8**,** 286-98.

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

FATICA, A. & BOZZONI, I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet,* 15**,** 7-21.

FAUSTO, N. 2000. Liver regeneration. *J Hepatol,* 32**,** 19-31.

FEARON, E. R., BURKE, P. J., ZEHNBAUER, B. A., VOGELSTEIN, B. & SCHIFFER, C. A. 1986. Differentiation of blast cells in acute nonlymphocytic leukemia. *N Engl J Med,* 315**,** 1488.

FIALKOW, P. J., GARTLER, S. M. & YOSHIDA, A. 1967. Clonal origin of chronic myelocytic leukemia in man. *Proc Natl Acad Sci U S A,* 58**,** 1468-71.

FIDLER, I. J. 1978. Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res,* 38**,** 2651-60.

FIDLER, I. J. & KRIPKE, M. L. 1977. Metastasis results from preexisting variant cells within a malignant tumor. *Science,* 197**,** 893-5.

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

FONG, S., ITAHANA, Y., SUMIDA, T., SINGH, J., COPPE, J. P., LIU, Y., RICHARDS, P. C., BENNINGTON, J. L., LEE, N. M., DEBS, R. J. & DESPREZ, P. Y. 2003. Id-1 as a molecular target in therapy for breast cancer cell invasion and metastasis. *Proc Natl Acad Sci U S A,* 100**,** 13543-8.

FORD, L. P., BAGGA, P. S. & WILUSZ, J. 1997. The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. *Mol Cell Biol,* 17**,** 398-406.

FRANC, J. M. 1978. Organization and Function of Ctenophore Colloblasts - Ultrastructural-Study. *Biological Bulletin,* 155**,** 527-541.

FU, H., FENG, J., LIU, Q., SUN, F., TIE, Y., ZHU, J., XING, R., SUN, Z. & ZHENG, X. 2009. Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett,* 583**,** 437-42.

FUNAYAMA, N. 2010. The stem cell system in demosponges: insights into the origin of somatic stem cells. *Dev Growth Differ,* 52**,** 1-14.

FUNAYAMA, N., NAKATSUKASA, M., MOHRI, K., MASUDA, Y. & AGATA, K. 2010. Piwi expression in archeocytes and choanocytes in demosponges: insights into the stem cell system in demosponges. *Evol Dev,* 12**,** 275-87.

GAILANI, M. R. & BALE, A. E. 1999. Acquired and inherited basal cell carcinomas and the patched gene. *Adv Dermatol,* 14**,** 261-83; discussion 284.

GALBRAITH, C. G. & SHEETZ, M. P. 1998. Forces on adhesive contacts affect cell function. *Curr Opin Cell Biol,* 10**,** 566-71.

GAN, H., LIN, X., ZHANG, Z., ZHANG, W., LIAO, S., WANG, L. & HAN, C. 2011. piRNA profiling during specific stages of mouse spermatogenesis. *RNA,* 17**,** 1191-203.

GAPP, K., JAWAID, A., SARKIES, P., BOHACEK, J., PELCZAR, P., PRADOS, J., FARINELLI, L., MISKA, E. & MANSUY, I. M. 2014. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci,* 17**,** 667-9.

GEBETSBERGER, J., ZYWICKI, M., KUNZI, A. & POLACEK, N. 2012. tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii. *Archaea,* 2012**,** 260909.

GHILDIYAL, M., SEITZ, H., HORWICH, M. D., LI, C., DU, T., LEE, S., XU, J., KITTLER, E. L., ZAPP, M. L., WENG, Z. & ZAMORE, P. D. 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science,* 320**,** 1077-81.

GHILDIYAL, M. & ZAMORE, P. D. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet,* 10**,** 94-108.

GIRARD, A., SACHIDANANDAM, R., HANNON, G. J. & CARMELL, M. A. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature,* 442**,** 199-202.

GOU, L. T., DAI, P., YANG, J. H., XUE, Y., HU, Y. P., ZHOU, Y., KANG, J. Y., WANG, X., LI, H., HUA, M. M., ZHAO, S., HU, S. D., WU, L. G., SHI, H. J., LI, Y., FU, X. D., QU, L. H., WANG, E. D. & LIU, M. F. 2014. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res,* 24**,** 680-700.

GRAY, M. J., DALLAS, N. A., VAN BUREN, G., XIA, L., YANG, A. D., SOMCIO, R. J., GAUR, P., MANGALA, L. S., VIVAS-MEJIA, P. E., FAN, F., SANGUINO, A. M., GALLICK, G. E., LOPEZ-BERESTEIN, G., SOOD, A. K. & ELLIS, L. M. 2008. Therapeutic targeting of Id2 reduces growth of human colorectal carcinoma in the murine liver. *Oncogene,* 27**,** 7192-200.

GRISHOK, A., PASQUINELLI, A. E., CONTE, D., LI, N., PARRISH, S., HA, I., BAILLIE, D. L., FIRE, A., RUVKUN, G. & MELLO, C. C. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell,* 106**,** 23-34.

GRIVNA, S. T., PYHTILA, B. & LIN, H. 2006. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc Natl Acad Sci U S A,* 103**,** 13415-20.

GROCHOLA, L. F., GREITHER, T., TAUBERT, H., MOLLER, P., KNIPPSCHILD, U., UDELNOW, A., HENNE-BRUNS, D. & WURL, P. 2008. The stem cell-associated Hiwi gene in human

adenocarcinoma of the pancreas: expression and risk of tumour-related death. *Br J Cancer,* 99**,** 1083-8.

GRUDZIEN, E., KALEK, M., JEMIELITY, J., DARZYNKIEWICZ, E. & RHOADS, R. E. 2006. Differential inhibition of mRNA degradation pathways by novel cap analogs. *J Biol Chem,* 281**,** 1857-67.

GU, W., LEE, H. C., CHAVES, D., YOUNGMAN, E. M., PAZOUR, G. J., CONTE, D., JR. & MELLO, C. C. 2012. CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. *Cell,* 151**,** 1488-500.

GUPTA, S., VERFAILLIE, C., CHMIELEWSKI, D., KIM, Y. & ROSENBERG, M. E. 2002. A role for extrarenal cells in the regeneration following acute renal failure. *Kidney Int,* 62**,** 1285-90.

GUTTMAN, M., DONAGHEY, J., CAREY, B. W., GARBER, M., GRENIER, J. K., MUNSON, G., YOUNG, G., LUCAS, A. B., ACH, R., BRUHN, L., YANG, X., AMIT, I., MEISSNER, A., REGEV, A., RINN, J. L., ROOT, D. E. & LANDER, E. S. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature,* 477**,** 295-300.

GUTTMAN, M., GARBER, M., LEVIN, J. Z., DONAGHEY, J., ROBINSON, J., ADICONIS, X., FAN, L., KOZIOL, M. J., GNIRKE, A., NUSBAUM, C., RINN, J. L., LANDER, E. S. & REGEV, A. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol,* 28**,** 503-10.

HAILESELLASSE SENE, K., PORTER, C. J., PALIDWOR, G., PEREZ-IRATXETA, C., MURO, E. M., CAMPBELL, P. A., RUDNICKI, M. A. & ANDRADE-NAVARRO, M. A. 2007. Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics,* 8**,** 85.

HARDING, J. L., HORSWELL, S., HELIOT, C., ARMISEN, J., ZIMMERMAN, L. B., LUSCOMBE, N. M., MISKA, E. A. & HILL, C. S. 2014. Small RNA profiling of Xenopus embryos reveals novel miRNAs and a new class of small RNAs derived from intronic transposable elements. *Genome Res,* 24**,** 96-106.

HARRIS, A. N. & MACDONALD, P. M. 2001. Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. *Development,* 128**,** 2823-32.

HAUSSECKER, D., HUANG, Y., LAU, A., PARAMESWARAN, P., FIRE, A. Z. & KAY, M. A. 2010. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA,* 16**,** 673-95.

HE, W., WANG, Z., WANG, Q., FAN, Q., SHOU, C., WANG, J., GIERCKSKY, K. E., NESLAND, J. M. & SUO, Z. 2009. Expression of HIWI in human esophageal squamous cell carcinoma is significantly associated with poorer prognosis. *BMC Cancer,* 9**,** 426.

HEPPNER, G. H. 1984. Tumor heterogeneity. *Cancer Res,* 44**,** 2259-65.

HERTWIG, R. 1880. Ueber den Bau der Ctenophoren. *Jena. Z. Naturwiss,* 14.

HOLLNAGEL, A., OEHLMANN, V., HEYMER, J., RUTHER, U. & NORDHEIM, A. 1999. Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J Biol Chem,* 274**,** 19838-45.

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

HRUZ, T., LAULE, O., SZABO, G., WESSENDORP, F., BLEULER, S., OERTLE, L., WIDMAYER, P., GRUISSEM, W. & ZIMMERMANN, P. 2008. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics,* 2008**,** 420747.

HSIA, D. A., MITRA, S. K., HAUCK, C. R., STREBLOW, D. N., NELSON, J. A., ILIC, D., HUANG, S., LI, E., NEMEROW, G. R., LENG, J., SPENCER, K. S., CHERESH, D. A. & SCHLAEPFER, D. D. 2003. Differential regulation of cell motility and invasion by FAK. *J Cell Biol,* 160**,** 753-67.

ISHIZU, H., SIOMI, H. & SIOMI, M. C. 2012. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev,* 26**,** 2361-73.

JACKSON, A. L. & LINSLEY, P. S. 2010. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov,* 9**,** 57-67.

JACKSON, E. B. A. B., A. M. 1941. Studies on a transplantable embryoma of the mouse. *Cancer Research,* 1**,** 494-498.

JACOBSON, S. L. & PIPER, H. M. 1986. Cell cultures of adult cardiomyocytes as models of the myocardium. *J Mol Cell Cardiol,* 18**,** 661-78.

JAENISCH, R. & YOUNG, R. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell,* 132**,** 567-82.

JANIC, A., MENDIZABAL, L., LLAMAZARES, S., ROSSELL, D. & GONZALEZ, C. 2010. Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. *Science,* 330**,** 1824-7.

JEONG, W. J., YOON, J., PARK, J. C., LEE, S. H., KADUWAL, S., KIM, H., YOON, J. B. & CHOI, K. Y. 2012. Ras stabilization through aberrant activation of Wnt/beta-catenin signaling promotes intestinal tumorigenesis. *Sci Signal,* 5**,** ra30.

JI, Z., LEE, J. Y., PAN, Z., JIANG, B. & TIAN, B. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A,* 106**,** 7028-33.

JULIANO, C., WANG, J. & LIN, H. 2011. Uniting Germline and Stem Cells: The Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. *Annu Rev Genet,* 45**,** 447-69.

KAELIN, W. G., JR. 2012. Molecular biology. Use and abuse of RNAi to study mammalian gene function. *Science,* 337**,** 421-2.

KALLIO, M. A., TUIMALA, J. T., HUPPONEN, T., KLEMELA, P., GENTILE, M., SCHEININ, I., KOSKI, M., KAKI, J. & KORPELAINEN, E. I. 2011. Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics,* 12**,** 507.

KALMYKOVA, A. I., KLENOV, M. S. & GVOZDEV, V. A. 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res,* 33**,** 2052-9.

KATSUNO, Y., LAMOUILLE, S. & DERYNCK, R. 2013. TGF-beta signaling and epithelial-mesenchymal transition in cancer progression. *Curr Opin Oncol,* 25**,** 76-84.

KAWAJI, H., NAKAMURA, M., TAKAHASHI, Y., SANDELIN, A., KATAYAMA, S., FUKUDA, S., DAUB, C. O., KAI, C., KAWAI, J., YASUDA, J., CARNINCI, P. & HAYASHIZAKI, Y. 2008. Hidden layers of human small RNAs. *BMC Genomics,* 9**,** 157.

KAWAOKA, S., IZUMI, N., KATSUMA, S. & TOMARI, Y. 2011. 3' end formation of PIWI-interacting RNAs in vitro. *Mol Cell,* 43**,** 1015-22.

KEAM, S. P., YOUNG, P. E., MCCORKINDALE, A. L., DANG, T. H., CLANCY, J. L., HUMPHREYS, D. T., PREISS, T., HUTVAGNER, G., MARTIN, D. I., CROPLEY, J. E. & SUTER, C. M. 2014. The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. *Nucleic Acids Res*.

KEENE, J. D., KOMISAROW, J. M. & FRIEDERSDORF, M. B. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc,* 1**,** 302-7.

KHODAREV, N. N., ROIZMAN, B. & WEICHSELBAUM, R. R. 2012. Molecular pathways: interferon/stat1 pathway: role in the tumor resistance to genotoxic stress and aggressive growth. *Clin Cancer Res,* 18**,** 3015-21.

KIM, J., CHU, J., SHEN, X., WANG, J. & ORKIN, S. H. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell,* 132**,** 1049-61.

KIM, J. B., URBAN, K., COCHRAN, E., LEE, S., ANG, A., RICE, B., BATA, A., CAMPBELL, K., COFFEE, R., GORODINSKY, A., LU, Z., ZHOU, H., KISHIMOTO, T. K. & LASSOTA, P. 2010. Non-invasive detection of a small number of bioluminescent cancer cells in vivo. *PLoS One,* 5**,** e9364.

KING, T. H., LIU, B., MCCULLY, R. R. & FOURNIER, M. J. 2003. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell,* 11**,** 425-35.

KINGSTON, R. E., CHEN, C. A. & OKAYAMA, H. 2003. Calcium phosphate transfection. *Curr Protoc Cell Biol,* Chapter 20**,** Unit 20 3.

KISS-LASZLO, Z., HENRY, Y., BACHELLERIE, J. P., CAIZERGUES-FERRER, M. & KISS, T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell,* 85**,** 1077-88.

KLATTENHOFF, C., XI, H., LI, C., LEE, S., XU, J., KHURANA, J. S., ZHANG, F., SCHULTZ, N., KOPPETSCH, B. S., NOWOSIELSKA, A., SEITZ, H., ZAMORE, P. D., WENG, Z. & THEURKAUF, W. E. 2009. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell,* 138**,** 1137-49.

KLEEFF, J., ISHIWATA, T., FRIESS, H., BUCHLER, M. W., ISRAEL, M. A. & KORC, M. 1998. The helix-loop-helix protein Id2 is overexpressed in human pancreatic cancer. *Cancer Res,* 58**,** 3769-72.

KLEIN, C. A. 2003. The systemic progression of human cancer: a focus on the individual disseminated cancer cell--the unit of selection. *Adv Cancer Res,* 89**,** 35-67.

KLEINSMITH, L. J. & PIERCE, G. B., JR. 1964. Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res,* 24**,** 1544-51.

KLENOV, M. S., SOKOLOVA, O. A., YAKUSHEV, E. Y., STOLYARENKO, A. D., MIKHALEVA, E. A., LAVROV, S. A. & GVOZDEV, V. A. 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci U S A*.

KNOWLES, B. B., ADEN, D. P. & SOLTER, D. 1978. Monoclonal antibody detecting a stage-specific embryonic antigen (SSEA-1) on preimplantation mouse embryos and teratocarcinoma cells. *Curr Top Microbiol Immunol,* 81**,** 51-3.

KORINEK, V., BARKER, N., MOERER, P., VAN DONSELAAR, E., HULS, G., PETERS, P. J. & CLEVERS, H. 1998. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet,* 19**,** 379-83.

KOSAKA, Y., KOBAYASHI, N., FUKAZAWA, T., TOTSUGAWA, T., MARUYAMA, M., YONG, C., ARATA, T., IKEDA, H., KOBAYASHI, K., UEDA, T., KURABAYASHI, Y. & TANAKA, N. 2004. Lentivirus-based gene delivery in mouse embryonic stem cells. *Artif Organs,* 28**,** 271-7.

KOTANI, H., NEWTON, P. B., 3RD, ZHANG, S., CHIANG, Y. L., OTTO, E., WEAVER, L., BLAESE, R. M., ANDERSON, W. F. & MCGARRITY, G. J. 1994. Improved methods of retroviral vector transduction and production for gene therapy. *Hum Gene Ther,* 5**,** 19-28.

KURAMOCHI-MIYAGAWA, S., KIMURA, T., IJIRI, T. W., ISOBE, T., ASADA, N., FUJITA, Y., IKAWA, M., IWAI, N., OKABE, M., DENG, W., LIN, H., MATSUDA, Y. & NAKANO, T. 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development,* 131**,** 839-49.

KURAMOCHI-MIYAGAWA, S., WATANABE, T., GOTOH, K., TOTOKI, Y., TOYODA, A., IKAWA, M., ASADA, N., KOJIMA, K., YAMAGUCHI, Y., IJIRI, T. W., HATA, K., LI, E., MATSUDA, Y., KIMURA, T., OKABE, M., SAKAKI, Y., SASAKI, H. & NAKANO, T. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev,* 22**,** 908-17.

LABERGE, R. M., AMBADIPUDI, R. & GEORGES, E. 2009. P-glycoprotein (ABCB1) modulates collateral sensitivity of a multidrug resistant cell line to verapamil. *Arch Biochem Biophys,* 491**,** 53-60.

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

LANGMEAD, B. 2010. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics,* Chapter 11**,** Unit 11 7.

LAPLANTE, M. & SABATINI, D. M. 2012. mTOR signaling in growth control and disease. *Cell,* 149**,** 274-93.

LAU, N. C., SETO, A. G., KIM, J., KURAMOCHI-MIYAGAWA, S., NAKANO, T., BARTEL, D. P. & KINGSTON, R. E. 2006. Characterization of the piRNA complex from rat testes. *Science,* 313**,** 363-7.

LAWSON, K. A., MENESES, J. J. & PEDERSEN, R. A. 1991. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development,* 113**,** 891-911.

LEE, E. J., BANERJEE, S., ZHOU, H., JAMMALAMADAKA, A., ARCILA, M., MANJUNATH, B. S. & KOSIK, K. S. 2011. Identification of piRNAs in the central nervous system. *RNA,* 17**,** 1090-9.

LEE, J. H., JUNG, C., JAVADIAN-ELYADERANI, P., SCHWEYER, S., SCHUTTE, D., SHOUKIER, M., KARIMI-BUSHERI, F., WEINFELD, M., RASOULI-NIA, A., HENGSTLER, J. G., MANTILLA, A., SOLEIMANPOUR-LICHAEI, H. R., ENGEL, W., ROBSON, C. N. & NAYERNIA, K. 2010. Pathways of proliferation and antiapoptosis driven in breast cancer stem cells by stem cell protein piwil2. *Cancer Res,* 70**,** 4569-79.

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

LEE, S. G., KIM, S., ROBBINS, P. D. & KIM, B. G. 1996. Optimization of environmental factors for the production and handling of recombinant retrovirus. *Appl Microbiol Biotechnol,* 45**,** 477-83.

LEE, Y. S., SHIBATA, Y., MALHOTRA, A. & DUTTA, A. 2009. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev,* 23**,** 2639-49.

LEHMANN, R. 2012. Germline stem cells: origin and destiny. *Cell Stem Cell,* 10**,** 729-39.

LEHMANN, R. & NUSSLEIN-VOLHARD, C. 1991. The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. *Development,* 112**,** 679-91.

LENGNER, C. J., CAMARGO, F. D., HOCHEDLINGER, K., WELSTEAD, G. G., ZAIDI, S., GOKHALE, S., SCHOLER, H. R., TOMILIN, A. & JAENISCH, R. 2007. Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell,* 1**,** 403-15.

LI, C., VAGIN, V. V., LEE, S., XU, J., MA, S., XI, H., SEITZ, H., HORWICH, M. D., SYRZYCKA, M., HONDA, B. M., KITTLER, E. L., ZAPP, M. L., KLATTENHOFF, C., SCHULZ, N., THEURKAUF, W. E., WENG, Z. & ZAMORE, P. D. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell,* 137**,** 509-21.

LI, L. & NEAVES, W. B. 2006. Normal stem cells and cancer stem cells: the niche matters. *Cancer Res,* 66**,** 4553-7.

LI, L. & XIE, T. 2005. Stem cell niche: structure and function. *Annu Rev Cell Dev Biol,* 21**,** 605-31.

LI, L., YU, C., GAO, H. & LI, Y. 2010. Argonaute proteins: potential biomarkers for human colon cancer. *BMC Cancer,* 10**,** 38.

LI, X. Z., ROY, C. K., DONG, X., BOLCUN-FILAS, E., WANG, J., HAN, B. W., XU, J., MOORE, M. J., SCHIMENTI, J. C., WENG, Z. & ZAMORE, P. D. 2013. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol Cell,* 50**,** 67-81.

LI, Z., ENDER, C., MEISTER, G., MOORE, P. S., CHANG, Y. & JOHN, B. 2012. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res,* 40**,** 6787-99.

LIM, E., VAILLANT, F., WU, D., FORREST, N. C., PAL, B., HART, A. H., ASSELIN-LABAT, M. L., GYORKI, D. E., WARD, T., PARTANEN, A., FELEPPA, F., HUSCHTSCHA, L. I., THORNE, H. J., FOX, S. B., YAN, M., FRENCH, J. D., BROWN, M. A., SMYTH, G. K., VISVADER, J. E. & LINDEMAN, G. J. 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med,* 15**,** 907-13.

LIM, R. S., ANAND, A., NISHIMIYA-FUJISAWA, C., KOBAYASHI, S. & KAI, T. 2013. Analysis of Hydra PIWI proteins and piRNAs uncover early evolutionary origins of the piRNA pathway. *Dev Biol*.

LIN, C. Q., SINGH, J., MURATA, K., ITAHANA, Y., PARRINELLO, S., LIANG, S. H., GILLETT, C. E., CAMPISI, J. & DESPREZ, P. Y. 2000. A role for Id-1 in the aggressive phenotype and steroid hormone response of human breast cancer cells. *Cancer Res,* 60**,** 1332-40.

LIN, E. A., KONG, L., BAI, X. H., LUAN, Y. & LIU, C. J. 2009. miR-199a, a bone morphogenic protein 2-responsive MicroRNA, regulates chondrogenesis via direct targeting to Smad1. *J Biol Chem,* 284**,** 11326-35.

LIN, H. 2002. The stem-cell niche theory: lessons from flies. *Nat Rev Genet,* 3**,** 931-40.

LIN, H. & SPRADLING, A. C. 1997. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Development,* 124**,** 2463-76.

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

LIU, J. J., SHEN, R., CHEN, L., YE, Y., HE, G., HUA, K., JARJOURA, D., NAKANO, T., RAMESH, G. K., SHAPIRO, C. L., BARSKY, S. H. & GAO, J. X. 2010. Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. *Int J Clin Exp Pathol,* 3**,** 328-37.

LIU, W., ZHANG, L. & WU, R. 2013. Differential expression of STAT1 and IFN-gamma in primary and invasive or metastatic wilms tumors. *J Surg Oncol,* 108**,** 152-6.

LIU, X., SUN, Y., GUO, J., MA, H., LI, J., DONG, B., JIN, G., ZHANG, J., WU, J., MENG, L. & SHOU, C. 2006. Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. *Int J Cancer,* 118**,** 1922-9.

LOH, Y. H., WU, Q., CHEW, J. L., VEGA, V. B., ZHANG, W., CHEN, X., BOURQUE, G., GEORGE, J., LEONG, B., LIU, J., WONG, K. Y., SUNG, K. W., LEE, C. W., ZHAO, X. D., CHIU, K. P., LIPOVICH, L., KUZNETSOV, V. A., ROBSON, P., STANTON, L. W., WEI, C. L., RUAN, Y., LIM, B. & NG, H. H. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet,* 38**,** 431-40.

LU, Y., LI, C., ZHANG, K., SUN, H., TAO, D., LIU, Y., ZHANG, S. & MA, Y. 2010. Identification of piRNAs in Hela cells by massive parallel sequencing. *BMB Rep,* 43**,** 635-41.

MA, J. B., YUAN, Y. R., MEISTER, G., PEI, Y., TUSCHL, T. & PATEL, D. J. 2005. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature,* 434**,** 666-70.

MACFARLAN, T. S., GIFFORD, W. D., DRISCOLL, S., LETTIERI, K., ROWE, H. M., BONANOMI, D., FIRTH, A., SINGER, O., TRONO, D. & PFAFF, S. L. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*.

MALANCHI, I., SANTAMARIA-MARTINEZ, A., SUSANTO, E., PENG, H., LEHR, H. A., DELALOYE, J. F. & HUELSKEN, J. 2012. Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature,* 481**,** 85-9.

MALONE, C. D., BRENNECKE, J., DUS, M., STARK, A., MCCOMBIE, W. R., SACHIDANANDAM, R. & HANNON, G. J. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell,* 137**,** 522-35.

MALONE, J. H. & OLIVER, B. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol,* 9**,** 34.

MARGOLIS, J. & SPRADLING, A. 1995. Identification and behavior of epithelial stem cells in the Drosophila ovary. *Development,* 121**,** 3797-807.

MARSON, A., LEVINE, S. S., COLE, M. F., FRAMPTON, G. M., BRAMBRINK, T., JOHNSTONE, S., GUENTHER, M. G., JOHNSTON, W. K., WERNIG, M., NEWMAN, J., CALABRESE, J. M.,

DENNIS, L. M., VOLKERT, T. L., GUPTA, S., LOVE, J., HANNETT, N., SHARP, P. A., BARTEL, D. P., JAENISCH, R. & YOUNG, R. A. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell,* 134**,** 521-33.

MASUI, S., NAKATAKE, Y., TOYOOKA, Y., SHIMOSATO, D., YAGI, R., TAKAHASHI, K., OKOCHI, H., OKUDA, A., MATOBA, R., SHAROV, A. A., KO, M. S. & NIWA, H. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol,* 9**,** 625-35.

MATSUI, T., LEUNG, D., MIYASHITA, H., MAKSAKOVA, I. A., MIYACHI, H., KIMURA, H., TACHIBANA, M., LORINCZ, M. C. & SHINKAI, Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature,* 464**,** 927-31.

MAUTE, R. L., DALLA-FAVERA, R. & BASSO, K. 2014. RNAs with multiple personalities. *Wiley Interdiscip Rev RNA,* 5**,** 1-13.

MAUTE, R. L., SCHNEIDER, C., SUMAZIN, P., HOLMES, A., CALIFANO, A., BASSO, K. & DALLA-FAVERA, R. 2013. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl Acad Sci U S A,* 110**,** 1404-9.

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

MCTAGGART, S. & AL-RUBEAI, M. 2000. Effects of culture parameters on the production of retroviral vectors by a human packaging cell line. *Biotechnol Prog,* 16**,** 859-65.

MEALEY, K. L., BARHOUMI, R., BURGHARDT, R. C., SAFE, S. & KOCHEVAR, D. T. 2002. Doxycycline induces expression of P glycoprotein in MCF-7 breast carcinoma cells. *Antimicrob Agents Chemother,* 46**,** 755-61.

MEDEMA, J. P. 2013. Cancer stem cells: the challenges ahead. *Nat Cell Biol,* 15**,** 338-44.

MEISTER, G. 2013. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet,* 14**,** 447-59.

MELLICK, A. S., PLUMMER, P. N., NOLAN, D. J., GAO, D., BAMBINO, K., HAHN, M., CATENA, R., TURNER, V., MCDONNELL, K., BENEZRA, R., BRINK, R., SWARBRICK, A. & MITTAL, V. 2010. Using the transcription factor inhibitor of DNA binding 1 to selectively target endothelial progenitor cells offers novel strategies to inhibit tumor angiogenesis and growth. *Cancer Res,* 70**,** 7273-82.

METEOGLU, I., MEYDAN, N. & ERKUS, M. 2008. Id-1: regulator of EGFR and VEGF and potential target for colorectal cancer therapy. *J Exp Clin Cancer Res,* 27**,** 69.

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

MIRI, K., LATHAM, K., PANNING, B., ZHONG, Z., ANDERSEN, A. & VARMUZA, S. 2013. The imprinted polycomb group gene Sfmbt2 is required for trophoblast maintenance and placenta development. *Development,* 140**,** 4480-9.

MITSUI, K., TOKUZAWA, Y., ITOH, H., SEGAWA, K., MURAKAMI, M., TAKAHASHI, K., MARUYAMA, M., MAEDA, M. & YAMANAKA, S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell,* 113**,** 631-42.

MOHN, F., SIENSKI, G., HANDLER, D. & BRENNECKE, J. 2014. The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. *Cell,* 157**,** 1364-79.

MOLYNEUX, G., GEYER, F. C., MAGNAY, F. A., MCCARTHY, A., KENDRICK, H., NATRAJAN, R., MACKAY, A., GRIGORIADIS, A., TUTT, A., ASHWORTH, A., REIS-FILHO, J. S. & SMALLEY, M. J. 2010. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell,* 7**,** 403-17.

NAGY, A., ROSSANT, J., NAGY, R., ABRAMOW-NEWERLY, W. & RODER, J. C. 1993. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A,* 90**,** 8424-8.

NALDINI, L., BLOMER, U., GALLAY, P., ORY, D., MULLIGAN, R., GAGE, F. H., VERMA, I. M. & TRONO, D. 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science,* 272**,** 263-7.

NAM, H. S. & BENEZRA, R. 2009. High levels of Id1 expression define B1 type adult neural stem cells. *Cell Stem Cell,* 5**,** 515-26.

NGUYEN, D. X., BOS, P. D. & MASSAGUE, J. 2009. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer,* 9**,** 274-84.

NGUYEN, L. V., VANNER, R., DIRKS, P. & EAVES, C. J. 2012. Cancer stem cells: an evolving concept. *Nat Rev Cancer,* 12**,** 133-43.

NI, J., TIEN, A. L. & FOURNIER, M. J. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell,* 89**,** 565-73.

NICHOLS, J., ZEVNIK, B., ANASTASSIADIS, K., NIWA, H., KLEWE-NEBENIUS, D., CHAMBERS, I., SCHOLER, H. & SMITH, A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell,* 95**,** 379-91.

NIWA, H., TOYOOKA, Y., SHIMOSATO, D., STRUMPF, D., TAKAHASHI, K., YAGI, R. & ROSSANT, J. 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell,* 123**,** 917-29.

NOLDE, M. J., CHENG, E. C., GUO, S. & LIN, H. 2013. Piwi genes are dispensable for normal hematopoiesis in mice. *PLoS One,* 8**,** e71950.

NOWELL, P. C. 1986. Mechanisms of tumor progression. *Cancer Res,* 46**,** 2203-7.

O'BRIEN, C. A., KRESO, A., RYAN, P., HERMANS, K. G., GIBSON, L., WANG, Y., TSATSANIS, A., GALLINGER, S. & DICK, J. E. 2012. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. *Cancer Cell,* 21**,** 777-92.

O'BRIEN, C. A., POLLETT, A., GALLINGER, S. & DICK, J. E. 2007. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature,* 445**,** 106-10.

ODORICO, J. S., KAUFMAN, D. S. & THOMSON, J. A. 2001. Multilineage differentiation from human embryonic stem cell lines. *Stem Cells,* 19**,** 193-204.

OEY, H. M., YOUNGSON, N. A. & WHITELAW, E. 2011. The characterisation of piRNA-related 19mers in the mouse. *BMC Genomics,* 12**,** 315.

OGAWA, M., FRIED, J., SAKAI, Y., STRIFE, A. & CLARKSON, B. D. 1970. Studies of cellular proliferation in human leukemia. VI. The proliferative activity, generation time, and emergence time of neutrophilic granulocytes in chronic granulocytic leukemia. *Cancer,* 25**,** 1031-49.

OHNISHI, Y., TOTOKI, Y., TOYODA, A., WATANABE, T., YAMAMOTO, Y., TOKUNAGA, K., SAKAKI, Y., SASAKI, H. & HOHJOH, H. 2010. Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. *Nucleic Acids Res,* 38**,** 5141-51.

OLIVIERI, D., SYKORA, M. M., SACHIDANANDAM, R., MECHTLER, K. & BRENNECKE, J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J,* 29**,** 3301-17.

OUYANG, X. S., WANG, X., LEE, D. T., TSAO, S. W. & WONG, Y. C. 2002. Over expression of ID-1 in prostate cancer. *J Urol,* 167**,** 2598-602.

OVERTURF, K., AL-DHALIMY, M., OU, C. N., FINEGOLD, M. & GROMPE, M. 1997. Serial transplantation reveals the stem-cell-like regenerative potential of adult mouse hepatocytes. *Am J Pathol,* 151**,** 1273-80.

PALAKODETI, D., SMIELEWSKA, M., LU, Y. C., YEO, G. W. & GRAVELEY, B. R. 2008. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA,* 14**,** 1174-86.

PANE, A., JIANG, P., ZHAO, D. Y., SINGH, M. & SCHUPBACH, T. 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *EMBO J,* 30**,** 4601-15.

PARSONS, J. T., MARTIN, K. H., SLACK, J. K., TAYLOR, J. M. & WEED, S. A. 2000. Focal adhesion kinase: a regulator of focal adhesion dynamics and cell movement. *Oncogene,* 19**,** 5606-13.

PAUKLIN, S., PEDERSEN, R. A. & VALLIER, L. 2011. Mouse pluripotent stem cells at a glance. *J Cell Sci,* 124**,** 3727-32.

PAULUS, W. A. W., N. 1986. The spermatogenesis of Ephydatia fluviatilis (Porifera). *Zoomorphology,* 106**,** 155-162.

PAYA, M., SEGOVIA, J. C., SANTIAGO, B., GALINDO, M., DEL RIO, P., PABLOS, J. L. & RAMIREZ, J. C. 2006. Optimising stable retroviral transduction of primary human synovial fibroblasts. *J Virol Methods,* 137**,** 95-102.

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

PELISSON, A., SAROT, E., PAYEN-GROSCHENE, G. & BUCHETON, A. 2007. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. *J Virol,* 81**,** 1951-60.

PERRAT, P. N., DASGUPTA, S., WANG, J., THEURKAUF, W., WENG, Z., ROSBASH, M. & WADDELL, S. 2013. Transposition-driven genomic heterogeneity in the Drosophila brain. *Science,* 340**,** 91-5.

PETIT, V. & THIERY, J. P. 2000. Focal adhesions: structure and dynamics. *Biol Cell,* 92**,** 477-94.

PEZIC, D., MANAKOV, S. A., SACHIDANANDAM, R. & ARAVIN, A. A. 2014. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev*.

POSTE, G. & FIDLER, I. J. 1980. The pathogenesis of cancer metastasis. *Nature,* 283**,** 139-46.

POTU, H., SGORBISSA, A. & BRANCOLINI, C. 2010. Identification of USP18 as an important regulator of the susceptibility to IFN-alpha and drug-induced apoptosis. *Cancer Res,* 70**,** 655-65.

PRATT, A. J. & MACRAE, I. J. 2009. The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem,* 284**,** 17897-901.

QIAO, D., ZEEMAN, A. M., DENG, W., LOOIJENGA, L. H. & LIN, H. 2002. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. *Oncogene,* 21**,** 3988-99.

QU, J. & BISHOP, J. M. 2012. Nucleostemin maintains self-renewal of embryonic stem cells and promotes reprogramming of somatic cells to pluripotency. *J Cell Biol,* 197**,** 731-45.

RAJASETHUPATHY, P., ANTONOV, I., SHERIDAN, R., FREY, S., SANDER, C., TUSCHL, T. & KANDEL, E. R. 2012. A Role for Neuronal piRNAs in the Epigenetic Control of Memory-Related Synaptic Plasticity. *Cell,* 149**,** 693-707.

RECZKO, M., MARAGKAKIS, M., ALEXIOU, P., GROSSE, I. & HATZIGEORGIOU, A. G. 2012. Functional microRNA targets in protein coding sequences. *Bioinformatics,* 28**,** 771-6.

REDDIEN, P. W., OVIEDO, N. J., JENNINGS, J. R., JENKIN, J. C. & SANCHEZ ALVARADO, A. 2005. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science,* 310**,** 1327-30.

REINHART, B. J. & BARTEL, D. P. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science,* 297**,** 1831.

REINHART, B. J., SLACK, F. J., BASSON, M., PASQUINELLI, A. E., BETTINGER, J. C., ROUGVIE, A. E., HORVITZ, H. R. & RUVKUN, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature,* 403**,** 901-6.

REYA, T. & CLEVERS, H. 2005. Wnt signalling in stem cells and cancer. *Nature,* 434**,** 843-50.

RICCI-VITIANI, L., LOMBARDI, D. G., PILOZZI, E., BIFFONI, M., TODARO, M., PESCHLE, C. & DE MARIA, R. 2007. Identification and expansion of human colon-cancer-initiating cells. *Nature,* 445**,** 111-5.

RIEDMANN, L. T. & SCHWENTNER, R. 2010. miRNA, siRNA, piRNA and argonautes: news in small matters. *RNA Biol,* 7**,** 133-9.

RINGROSE, L. & PARO, R. 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet,* 38**,** 413-43.

RINKEVICH, B., SHLEMBERG, Z. & FISHELSON, L. 1995. Whole-body protochordate regeneration from totipotent blood cells. *Proc Natl Acad Sci U S A,* 92**,** 7695-9.

RINKEVICH, Y., ROSNER, A., RABINOWITZ, C., LAPIDOT, Z., MOISEEVA, E. & RINKEVICH, B. 2010. Piwi positive cells that line the vasculature epithelium, underlie whole body regeneration in a basal chordate. *Dev Biol,* 345**,** 94-104.

RINKEVICH, Y., VOSKOBOYNIK, A., ROSNER, A., RABINOWITZ, C., PAZ, G., OREN, M., DOUEK, J., ALFASSI, G., MOISEEVA, E., ISHIZUKA, K. J., PALMERI, K. J., WEISSMAN, I. L. & RINKEVICH, B. 2013. Repeated, long-term cycling of putative stem cells between niches in a basal chordate. *Dev Cell,* 24**,** 76-88.

ROBINE, N., LAU, N. C., BALLA, S., JIN, Z., OKAMURA, K., KURAMOCHI-MIYAGAWA, S., BLOWER, M. D. & LAI, E. C. 2009. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol,* 19**,** 2066-76.

ROSA, A., SPAGNOLI, F. M. & BRIVANLOU, A. H. 2009. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. *Dev Cell,* 16**,** 517-27.

ROSENBLOOM, K. R., DRESZER, T. R., LONG, J. C., MALLADI, V. S., SLOAN, C. A., RANEY, B. J., CLINE, M. S., KAROLCHIK, D., BARBER, G. P., CLAWSON, H., DIEKHANS, M., FUJITA, P. A., GOLDMAN, M., GRAVELL, R. C., HARTE, R. A., HINRICHS, A. S., KIRKUP, V. M., KUHN, R. M., LEARNED, K., MADDREN, M., MEYER, L. R., POHL, A., RHEAD, B., WONG, M. C., ZWEIG, A. S., HAUSSLER, D. & KENT, W. J. 2012. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res,* 40**,** D912-7.

ROSS, R. J., WEINER, M. M. & LIN, H. 2014. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature,* 505**,** 353-9.

ROUGET, C., PAPIN, C., BOUREUX, A., MEUNIER, A. C., FRANCO, B., ROBINE, N., LAI, E. C., PELISSON, A. & SIMONELIG, M. 2010. Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature,* 467**,** 1128-32.

RUBINFELD, B., ALBERT, I., PORFIRI, E., FIOL, C., MUNEMITSU, S. & POLAKIS, P. 1996. Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science,* 272**,** 1023-6.

RUBY, J. G., JAN, C., PLAYER, C., AXTELL, M. J., LEE, W., NUSBAUM, C., GE, H. & BARTEL, D. P. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell,* 127**,** 1193-207.

RUZINOVA, M. B. & BENEZRA, R. 2003. Id proteins in development, cell cycle and cancer. *Trends Cell Biol,* 13**,** 410-8.

SACHS, A. B. 1993. Messenger RNA degradation in eukaryotes. *Cell,* 74**,** 413-21.

SAI LAKSHMI, S. & AGRAWAL, S. 2008. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res,* 36**,** D173-7.

SAITO, K., NISHIDA, K. M., MORI, T., KAWAMURA, Y., MIYOSHI, K., NAGAMI, T., SIOMI, H. & SIOMI, M. C. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev,* 20**,** 2214-22.

SAITO, K., SAKAGUCHI, Y., SUZUKI, T., SIOMI, H. & SIOMI, M. C. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev,* 21**,** 1603-8.

SAITO, K. & SIOMI, M. C. 2010. Small RNA-mediated quiescence of transposable elements in animals. *Dev Cell,* 19**,** 687-97.

SAKURAI, T., RAMOZ, N., BARRETO, M., GAZDOIU, M., TAKAHASHI, N., GERTNER, M., DORR, N., GAMA SOSA, M. A., DE GASPERI, R., PEREZ, G., SCHMEIDLER, J., MITROPOULOU, V., LE, H. C., LUPU, M., HOF, P. R., ELDER, G. A. & BUXBAUM, J. D. 2010. Slc25a12 disruption alters myelination and neurofilaments: a model for a hypomyelination syndrome and childhood neurodevelopmental disorders. *Biol Psychiatry,* 67**,** 887-94.

SANCHEZ ALVARADO, A., NEWMARK, P. A., ROBB, S. M. & JUSTE, R. 2002. The Schmidtea mediterranea database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development,* 129**,** 5659-65.

SANSONE, P. & BROMBERG, J. 2012. Targeting the interleukin-6/Jak/stat pathway in human malignancies. *J Clin Oncol,* 30**,** 1005-14.

SASAKI, H. & MATSUI, Y. 2008. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet,* 9**,** 129-40.

SCHINDL, M., OBERHUBER, G., OBERMAIR, A., SCHOPPMANN, S. F., KARNER, B. & BIRNER, P. 2001. Overexpression of Id-1 protein is a marker for unfavorable prognosis in early-stage cervical cancer. *Cancer Res,* 61**,** 5703-6.

SCHIRLE, N. T. & MACRAE, I. J. 2012. The crystal structure of human Argonaute2. *Science,* 336**,** 1037-40.

SCHLAEPFER, D. D., MITRA, S. K. & ILIC, D. 2004. Control of motile and invasive cell phenotypes by focal adhesion kinase. *Biochim Biophys Acta,* 1692**,** 77-102.

SCHMID, V. 1975. Cell transformation in isolated striated muscle of hydromedusae independent of DNA synthesis. *Exp Cell Res,* 94**,** 401-8.

SCHMID, V. & ALDER, H. 1984. Isolated, mononucleated, striated muscle can undergo pluripotent transdifferentiation and form a complex regenerate. *Cell,* 38**,** 801-9.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc,* 3**,** 1101-8.

SCHUMACKER, P. T. 2006. Reactive oxygen species in cancer cells: live by the sword, die by the sword. *Cancer Cell,* 10**,** 175-6.

SCHUPBACH, T. & WIESCHAUS, E. 1991. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics,* 129**,** 1119-36.

SEIPEL, K., YANZE, N. & SCHMID, V. 2004. The germ line and somatic stem cell gene Cniwi in the jellyfish Podocoryne carnea. *Int J Dev Biol,* 48**,** 1-7.

SELL, S. 2004. Stem cell origin of cancer and differentiation therapy. *Crit Rev Oncol Hematol,* 51**,** 1-28.

SETO, A. G., KINGSTON, R. E. & LAU, N. C. 2007. The coming of age for Piwi proteins. *Mol Cell,* 26**,** 603-9.

SHAHBAZI, M. N., MEGIAS, D., EPIFANO, C., AKHMANOVA, A., GUNDERSEN, G. G., FUCHS, E. & PEREZ-MORENO, M. 2013. CLASP2 interacts with p120-catenin and governs microtubule dynamics at adherens junctions. *J Cell Biol,* 203**,** 1043-61.

SHARMA, A. K., NELSON, M. C., BRANDT, J. E., WESSMAN, M., MAHMUD, N., WELLER, K. P. & HOFFMAN, R. 2001. Human CD34(+) stem cells express the hiwi gene, a human homologue of the Drosophila gene piwi. *Blood,* 97**,** 426-34.

SHIROSHIMA, T., OKA, C. & KAWAICHI, M. 2009. Identification of LRP1B-interacting proteins and inhibition of protein kinase Calpha-phosphorylation of LRP1B by association with PICK1. *FEBS Lett,* 583**,** 43-8.

SHPIZ, S. & KALMYKOVA, A. 2009. Epigenetic transmission of piRNAs through the female germline. *Genome Biol,* 10**,** 208.

SIDDIQI, S. & MATUSHANSKY, I. 2012. Piwis and piwi-interacting RNAs in the epigenetics of cancer. *J Cell Biochem,* 113**,** 373-80.

SIMON, M., SARKIES, P., IKEGAMI, K., DOEBLEY, A. L., GOLDSTEIN, L. D., MITCHELL, J., SAKAGUCHI, A., MISKA, E. A. & AHMED, S. 2014. Reduced insulin/IGF-1 signaling restores germ cell immortality to caenorhabditis elegans Piwi mutants. *Cell Rep,* 7**,** 762-73.

SINGH, S. K., CLARKE, I. D., TERASAKI, M., BONN, V. E., HAWKINS, C., SQUIRE, J. & DIRKS, P. B. 2003. Identification of a cancer stem cell in human brain tumors. *Cancer Res,* 63**,** 5821-8.

SIOMI, M. C., SATO, K., PEZIC, D. & ARAVIN, A. A. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol,* 12**,** 246-58.

SLOMA, I., JIANG, X., EAVES, A. C. & EAVES, C. J. 2010. Insights into the stem cells of chronic myeloid leukemia. *Leukemia,* 24**,** 1823-33.

SMIBERT, P., MIURA, P., WESTHOLM, J. O., SHENKER, S., MAY, G., DUFF, M. O., ZHANG, D., EADS, B. D., CARLSON, J., BROWN, J. B., EISMAN, R. C., ANDREWS, J., KAUFMAN, T., CHERBAS, P., CELNIKER, S. E., GRAVELEY, B. R. & LAI, E. C. 2012. Global patterns of tissue-specific alternative polyadenylation in Drosophila. *Cell Rep,* 1**,** 277-89.

SMIT, A. F. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res,* 21**,** 1863-72.

SOBALA, A. & HUTVAGNER, G. 2013. Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. *RNA Biol,* 10**,** 553-63.

SOLTER, D. & KNOWLES, B. B. 1978. Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc Natl Acad Sci U S A,* 75**,** 5565-9.

SONG, J., SMITH, S. & HANNON, G. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*.

SONG, R., HENNIG, G. W., WU, Q., JOSE, C., ZHENG, H. & YAN, W. 2011. Male germ cells express abundant endogenous siRNAs. *Proc Natl Acad Sci U S A,* 108**,** 13159-64.

SPRADLING, A., DRUMMOND-BARBOSA, D. & KAI, T. 2001. Stem cells find their niche. *Nature,* 414**,** 98-104.

SPRUCE, T., PERNAUTE, B., DI-GREGORIO, A., COBB, B. S., MERKENSCHLAGER, M., MANZANARES, M. & RODRIGUEZ, T. A. 2010. An early developmental role for miRNAs in the maintenance of extraembryonic stem cells in the mouse embryo. *Dev Cell,* 19**,** 207-19.

STANKIC, M., PAVLOVIC, S., CHIN, Y., BROGI, E., PADUA, D., NORTON, L., MASSAGUE, J. & BENEZRA, R. 2013. TGF-beta-Id1 Signaling Opposes Twist1 and Promotes Metastatic Colonization via a Mesenchymal-to-Epithelial Transition. *Cell Rep,* 5**,** 1228-42.

STEIN, N. 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography,* 41**,** 641-643.

STEPHANOU, A. & LATCHMAN, D. S. 2003. STAT-1: a novel regulator of apoptosis. *Int J Exp Pathol,* 84**,** 239-44.

SU, H., TROMBLY, M. I., CHEN, J. & WANG, X. 2009. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev,* 23**,** 304-17.

SUGIMOTO, K., KAGE, H., AKI, N., SANO, A., KITAGAWA, H., NAGASE, T., YATOMI, Y., OHISHI, N. & TAKAI, D. 2007. The induction of H3K9 methylation by PIWIL4 at the p16Ink4a locus. *Biochem Biophys Res Commun,* 359**,** 497-502.

SUN, G., WANG, Y., SUN, L., LUO, H., LIU, N., FU, Z. & YOU, Y. 2011. Clinical significance of Hiwi gene expression in gliomas. *Brain Res,* 1373**,** 183-8.

SURANI, M. A., ANCELIN, K., HAJKOVA, P., LANGE, U. C., PAYER, B., WESTERN, P. & SAITOU, M. 2004. Mechanism of mouse germ cell specification: a genetic program regulating epigenetic reprogramming. *Cold Spring Harb Symp Quant Biol,* 69**,** 1-9.

SURANI, M. A., HAYASHI, K. & HAJKOVA, P. 2007. Genetic and epigenetic regulators of pluripotency. *Cell,* 128**,** 747-62.

SWARBRICK, A., ROY, E., ALLEN, T. & BISHOP, J. M. 2008. Id1 cooperates with oncogenic Ras to induce metastatic mammary carcinoma by subversion of the cellular senescence response. *Proc Natl Acad Sci U S A,* 105**,** 5402-7.

TAFT, R. J., GLAZOV, E. A., CLOONAN, N., SIMONS, C., STEPHEN, S., FAULKNER, G. J., LASSMANN, T., FORREST, A. R., GRIMMOND, S. M., SCHRODER, K., IRVINE, K., ARAKAWA, T., NAKAMURA, M., KUBOSAKI, A., HAYASHIDA, K., KAWAZU, C., MURATA, M., NISHIYORI, H., FUKUDA, S., KAWAI, J., DAUB, C. O., HUME, D. A., SUZUKI, H., ORLANDO, V., CARNINCI, P., HAYASHIZAKI, Y. & MATTICK, J. S. 2009. Tiny RNAs associated with transcription start sites in animals. *Nat Genet,* 41**,** 572-8.

TAM, O. H., ARAVIN, A. A., STEIN, P., GIRARD, A., MURCHISON, E. P., CHELOUFI, S., HODGES, E., ANGER, M., SACHIDANANDAM, R., SCHULTZ, R. M. & HANNON, G. J. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature,* 453**,** 534-8.

TAO, K., FANG, M., ALROY, J. & SAHAGIAN, G. G. 2008. Imagable 4T1 model for the study of late stage breast cancer. *BMC Cancer,* 8**,** 228.

TAUBERT, H., GREITHER, T., KAUSHAL, D., WURL, P., BACHE, M., BARTEL, F., KEHLEN, A., LAUTENSCHLAGER, C., HARRIS, L., KRAEMER, K., MEYE, A., KAPPLER, M., SCHMIDT, H., HOLZHAUSEN, H. J. & HAUPTMANN, S. 2007. Expression of the stem cell self-renewal gene Hiwi and risk of tumour-related death in patients with soft-tissue sarcoma. *Oncogene,* 26**,** 1098-100.

THOMPSON, D. M., LU, C., GREEN, P. J. & PARKER, R. 2008. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA,* 14**,** 2095-103.

THOMSON, T. & LIN, H. 2009. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol,* 25**,** 355-76.

TILL, J. E. & MC, C. E. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res,* 14**,** 213-22.

TILL, J. E. & MCCULLOCH, E. A. 1964. Repair Processes in Irradiated Mouse Hematopoietic Tissue. *Ann N Y Acad Sci,* 114**,** 115-25.

UNHAVAITHAYA, Y., HAO, Y., BEYRET, E., YIN, H., KURAMOCHI-MIYAGAWA, S., NAKANO, T. & LIN, H. 2009. MILI, a PIWI-interacting RNA-binding protein, is required for germ line stem cell self-renewal and appears to positively regulate translation. *J Biol Chem,* 284**,** 6507-19.

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

VALENCIA-SANCHEZ, M. A., LIU, J., HANNON, G. J. & PARKER, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev,* 20**,** 515-24.

VAN DE WETERING, M., SANCHO, E., VERWEIJ, C., DE LAU, W., OVING, I., HURLSTONE, A., VAN DER HORN, K., BATLLE, E., COUDREUSE, D., HARAMIS, A. P., TJON-PON-FONG, M., MOERER, P., VAN DEN BORN, M., SOETE, G., PALS, S., EILERS, M., MEDEMA, R. & CLEVERS, H. 2002. The

beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell,* 111**,** 241-50.

VARNUM-FINNEY, B., XU, L., BRASHEM-STEIN, C., NOURIGAT, C., FLOWERS, D., BAKKOUR, S., PEAR, W. S. & BERNSTEIN, I. D. 2000. Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. *Nat Med,* 6**,** 1278-81.

VASUDEVAN, S. 2012. Posttranscriptional upregulation by microRNAs. *Wiley Interdiscip Rev RNA,* 3**,** 311-30.

VISVADER, J. E. & LINDEMAN, G. J. 2008. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer,* 8**,** 755-68.

VLACHOS, I. S., KOSTOULAS, N., VERGOULIS, T., GEORGAKILAS, G., RECZKO, M., MARAGKAKIS, M., PARASKEVOPOULOU, M. D., PRIONIDIS, K., DALAMAGAS, T. & HATZIGEORGIOU, A. G. 2012. DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res,* 40**,** W498-504.

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

VOUREKAS, A., ZHENG, Q., ALEXIOU, P., MARAGKAKIS, M., KIRINO, Y., GREGORY, B. D. & MOURELATOS, Z. 2012. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat Struct Mol Biol,* 19**,** 773-81.

WALTHER, T. N., WITTOP KONING, T. H., SCHUMPERLI, D. & MULLER, B. 1998. A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. *RNA,* 4**,** 1034-46.

WANG, G. & REINKE, V. 2008. A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol,* 18**,** 861-7.

WANG, Y., ZAYAS, R. M., GUO, T. & NEWMARK, P. A. 2007. nanos function is essential for development and regeneration of planarian germ cells. *Proc Natl Acad Sci U S A,* 104**,** 5901-6.

WANG, Z., DAY, N., TRIFILLIS, P. & KILEDJIAN, M. 1999. An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. *Mol Cell Biol,* 19**,** 4552-60.

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

WATANABE, T., TOTOKI, Y., TOYODA, A., KANEDA, M., KURAMOCHI-MIYAGAWA, S., OBATA, Y., CHIBA, H., KOHARA, Y., KONO, T., NAKANO, T., SURANI, M. A., SAKAKI, Y. & SASAKI, H. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature,* 453**,** 539-43.

WATT, F. M. & HOGAN, B. L. 2000. Out of Eden: stem cells and their niches. *Science,* 287**,** 1427-30.

WEBB, D. J., PARSONS, J. T. & HORWITZ, A. F. 2002. Adhesion assembly, disassembly and turnover in migrating cells -- over and over and over again. *Nat Cell Biol,* 4**,** E97-100.

WECHSLER-REYA, R. & SCOTT, M. P. 2001. The developmental biology of brain tumors. *Annu Rev Neurosci,* 24**,** 385-428.

WECHSLER-REYA, R. J. & SCOTT, M. P. 1999. Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron,* 22**,** 103-14.

WEICK, E. M., SARKIES, P., SILVA, N., CHEN, R. A., MOSS, S. M., CORDING, A. C., AHRINGER, J., MARTINEZ-PEREZ, E. & MISKA, E. A. 2014. PRDE-1 is a nuclear factor essential for the biogenesis of Ruby motif-dependent piRNAs in C. elegans. *Genes Dev,* 28**,** 783-96.

WIDSCHWENDTER, M., FIEGL, H., EGLE, D., MUELLER-HOLZNER, E., SPIZZO, G., MARTH, C., WEISENBERGER, D. J., CAMPAN, M., YOUNG, J., JACOBS, I. & LAIRD, P. W. 2007. Epigenetic stem cell signature in cancer. *Nat Genet,* 39**,** 157-8.

WILCZYNSKA, A., MINSHALL, N., ARMISEN, J., MISKA, E. A. & STANDART, N. 2009. Two Piwi proteins, Xiwi and Xili, are expressed in the Xenopus female germline. *RNA,* 15**,** 337-45.

WILLIAMS, T. M., SELEGUE, J. E., WERNER, T., GOMPEL, N., KOPP, A. & CARROLL, S. B. 2008. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell,* 134**,** 610-23.

WILSON, H. V. 1907. A New Method by Which Sponges May Be Artificially Reared. *Science,* 25**,** 912-5.

WILSON, J. E., CONNELL, J. E. & MACDONALD, P. M. 1996. aubergine enhances oskar translation in the Drosophila ovary. *Development,* 122**,** 1631-9.

WOLBER, F. M., LEONARD, E., MICHAEL, S., ORSCHELL-TRAYCOFF, C. M., YODER, M. C. & SROUR, E. F. 2002. Roles of spleen and liver in development of the murine hematopoietic system. *Exp Hematol,* 30**,** 1010-9.

WONG, K. K., ENGELMAN, J. A. & CANTLEY, L. C. 2010. Targeting the PI3K signaling pathway in cancer. *Curr Opin Genet Dev,* 20**,** 87-90.

WU, Q., MA, Q., SHEHADEH, L. A., WILSON, A., XIA, L., YU, H. & WEBSTER, K. A. 2010. Expression of the Argonaute protein PiwiL2 and piRNAs in adult mouse mesenchymal stem cells. *Biochem Biophys Res Commun,* 396**,** 915-20.

YAMANAKA, Y., RALSTON, A., STEPHENSON, R. O. & ROSSANT, J. 2006. Cell and molecular regulation of the mouse blastocyst. *Dev Dyn,* 235**,** 2301-14.

YAMASHITA, M., NITTA, E., NAGAMATSU, G., IKUSHIMA, Y. M., HOSOKAWA, K., ARAI, F. & SUDA, T. 2013. Nucleostemin is indispensable for the maintenance and genetic stability of hematopoietic stem cells. *Biochem Biophys Res Commun,* 441**,** 196-201.

YAN, Z., HU, H. Y., JIANG, X., MAIERHOFER, V., NEB, E., HE, L., HU, Y., HU, H., LI, N., CHEN, W. & KHAITOVICH, P. 2011. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res,* 39**,** 6596-607.

YARDEN, Y. & PINES, G. 2012. The ERBB network: at last, cancer therapy meets systems biology. *Nat Rev Cancer,* 12**,** 553-63.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13**,** 134.

YU, Y., MAGGI, L. B., JR., BRADY, S. N., APICELLI, A. J., DAI, M. S., LU, H. & WEBER, J. D. 2006. Nucleophosmin is essential for ribosomal protein L5 nuclear export. *Mol Cell Biol,* 26**,** 3798-809.

ZHANG, F., WANG, J., XU, J., ZHANG, Z., KOPPETSCH, B. S., SCHULTZ, N., VREVEN, T., MEIGNIN, C., DAVIS, I., ZAMORE, P. D., WENG, Z. & THEURKAUF, W. E. 2012. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell,* 151**,** 871-84.

ZHANG, Z., WANG, J., SCHULTZ, N., ZHANG, F., PARHAD, S. S., TU, S., VREVEN, T., ZAMORE, P. D., WENG, Z. & THEURKAUF, W. E. 2014. The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell,* 157**,** 1353-63.

ZHAO, S., FUNG-LEUNG, W. P., BITTNER, A., NGO, K. & LIU, X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One,* 9**,** e78644.

ZHAO, Y. M., ZHOU, J. M., WANG, L. R., HE, H. W., WANG, X. L., TAO, Z. H., SUN, H. C., WU, W. Z., FAN, J., TANG, Z. Y. & WANG, L. 2012. HIWI is associated with prognosis in patients with hepatocellular carcinoma after curative resection. *Cancer,* 118**,** 2708-17.

ZHENG, G. X., RAVI, A., GOULD, G. M., BURGE, C. B. & SHARP, P. A. 2011. Genome-wide impact of a recently expanded microRNA cluster in mouse. *Proc Natl Acad Sci U S A,* 108**,** 15804-9.

ZHOU, H., ARCILA, M. L., LI, Z., LEE, E. J., HENZLER, C., LIU, J., RANA, T. M. & KOSIK, K. S. 2012. Deep annotation of mouse iso-miR and iso-moR variation. *Nucleic Acids Res,* 40**,** 5864-75.

ZHU, W., PAO, G. M., SATOH, A., CUMMINGS, G., MONAGHAN, J. R., HARKINS, T. T., BRYANT, S. V., RANDAL VOSS, S., GARDINER, D. M. & HUNTER, T. 2012. Activation of germline-specific genes is required for limb regeneration in the Mexican axolotl. *Dev Biol,* 370**,** 42-51.

8. Appendix

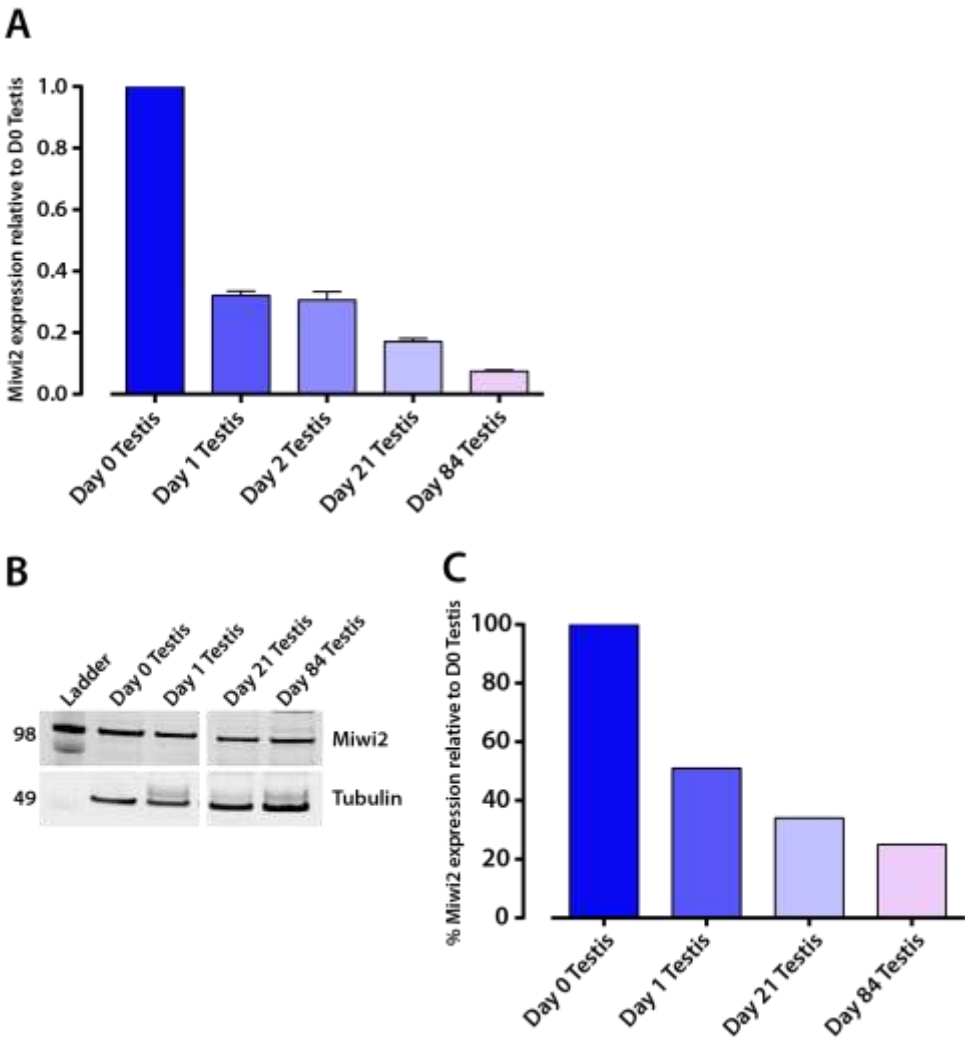
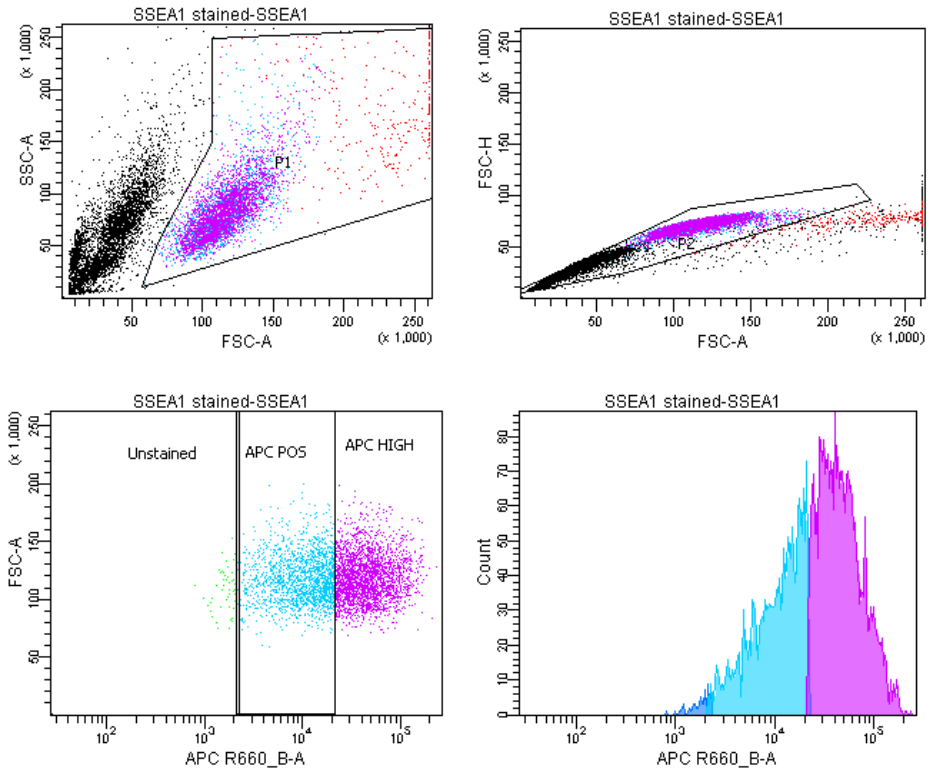


Figure 8-1: Miwi2 expression in murine testis persists into adulthood

(A) Quantitative PCR data for *Miwi2* expression. All gene expression is shown relative to Day 0 Testis. All values are averaged from four replicate PCR reactions. Error bars represent standard error of the mean. (B) Western blot showing Miwi2 expression (top panel). β-tubulin (bottom panel) was used as a loading control. (C) Graphical representation of densitometry measurements of Western blot data in (B). Miwi2 expression was corrected for β-tubulin expression and shown relative to Day 0 Testis.



Tube: SSEA1

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
P1	4,783	47.8	47.8
P2	4,504	94.2	45.0
Unstained	63	1.4	0.6
APC POS	1,790	39.7	17.9
APC HIGH	2,668	59.2	26.7

Figure 8-2: Gating strategy for SSEA1-positive ES cells using the BD FACS Aria.

ES cells were selected based on side scatter area (SSC-A) versus forward scatter area (FSC-A, top left plot) and single cells enriched using forward scatter height (FSC-H) and FSC-A (top right plot). SSEA1-APC-positive (APC HIGH) cells were selected using FSC-A versus fluorescence intensity using the 660nm filter (APC R660_B-A, bottom left plot). Negative controls included unstained cells (green, Unstained) and normal IgM-APC-stained cells (blue, APC POS, bottom left plot).

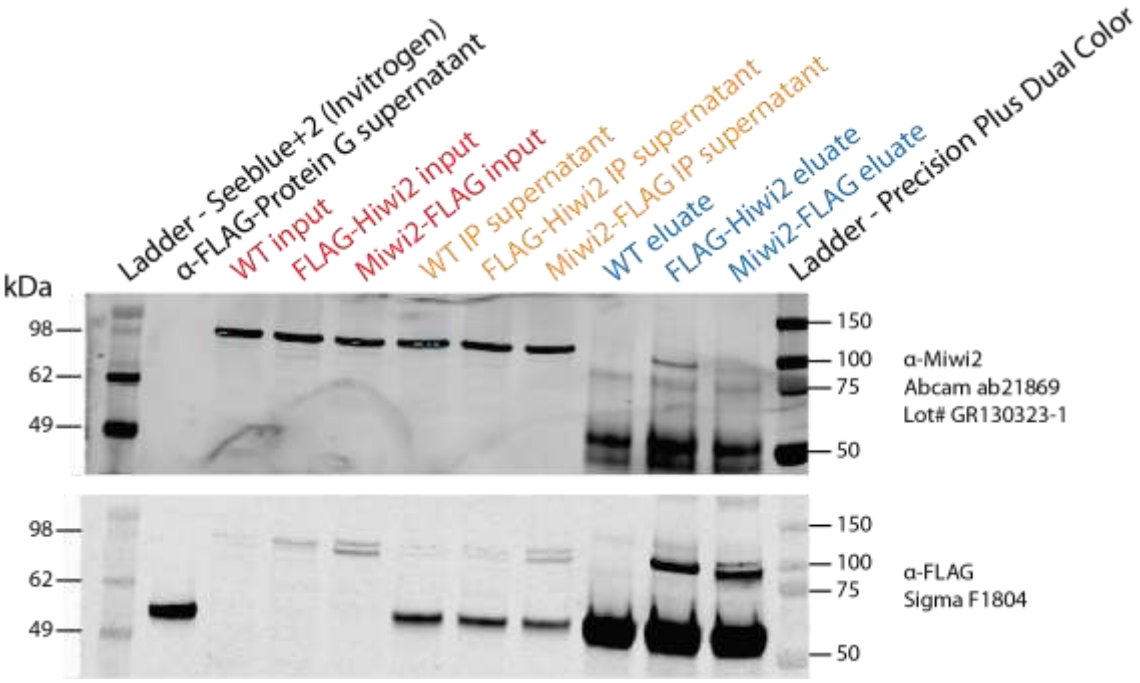


Figure 8-3: Specificity testing of rabbit polyclonal α-Miwi2 antibody

Western blot showing Miwi2 and FLAG expression. 293FT cells were transfected with pIRES-neo-FLAG/HA-Hiwi2 or pCMV6-Myc-FLAG-Miwi2 for 24h and immunoprecipitation was performed of FLAG-Hiwi2 or FLAG-Miwi2 from lysate. Immunodetection was performed using input lysate, IP supernatant and eluate with α-Miwi2 or α-FLAG primary antibody, followed by a fluorescent secondary antibody. Membrane was visualised at 680nm and 800nm with the LI-COR Odyssey Imaging System.

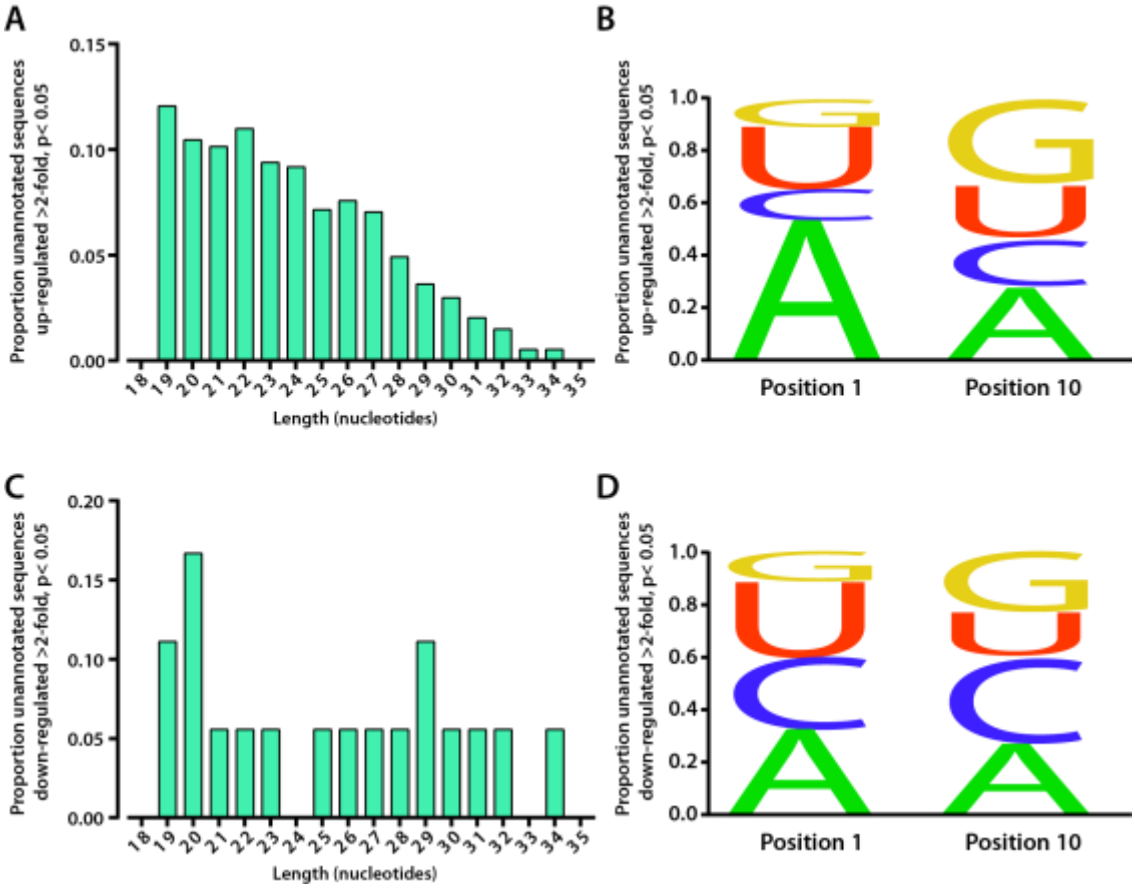


Figure 8-4: Sequence characteristics of small RNAs derived from unannotated regions

Sequence features of small RNAs from unannotated regions that were (A-B) up-regulated and (B-C) down-regulated >2-fold after 3 days *in vitro* differentiation. (A and C) Length distribution. (B and D) Sequence alignment at positions 1 and 10.

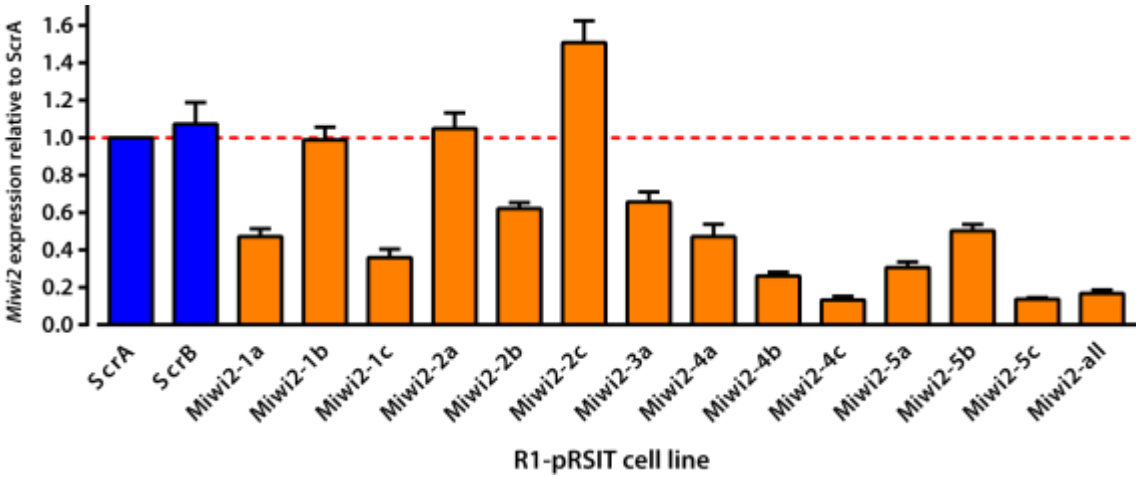


Figure 8-5: Miwi2 expression in R1-pRSIT mouse ES cells – all samples

Quantitative PCR data for *Miwi2* expression from FACS sorted samples with high quality RNA extracted. All gene expression is shown relative to R1-pRSIT-ScrA; this value is depicted by the red dotted line. All values are averaged from four triplicate PCR reactions. Error bars represent standard error of the mean.

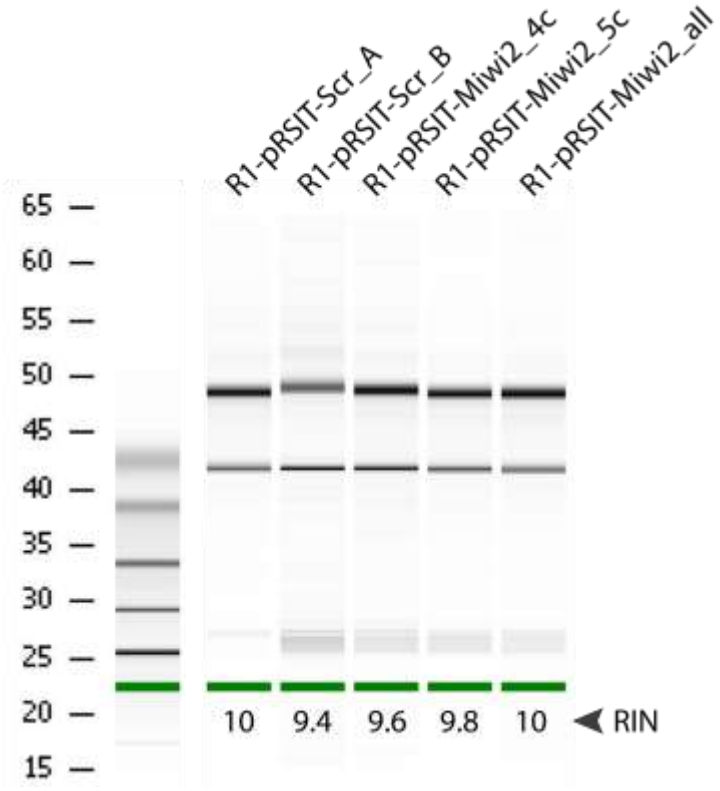


Figure 8-6: RNA integrity of R1-pRSIT ES cell RNA samples

RNA from Agilent 2100 Bioanalyzer RNA 6000 Nano Chip. RNA integrity number (RIN) indicates RNA integrity (max 10).

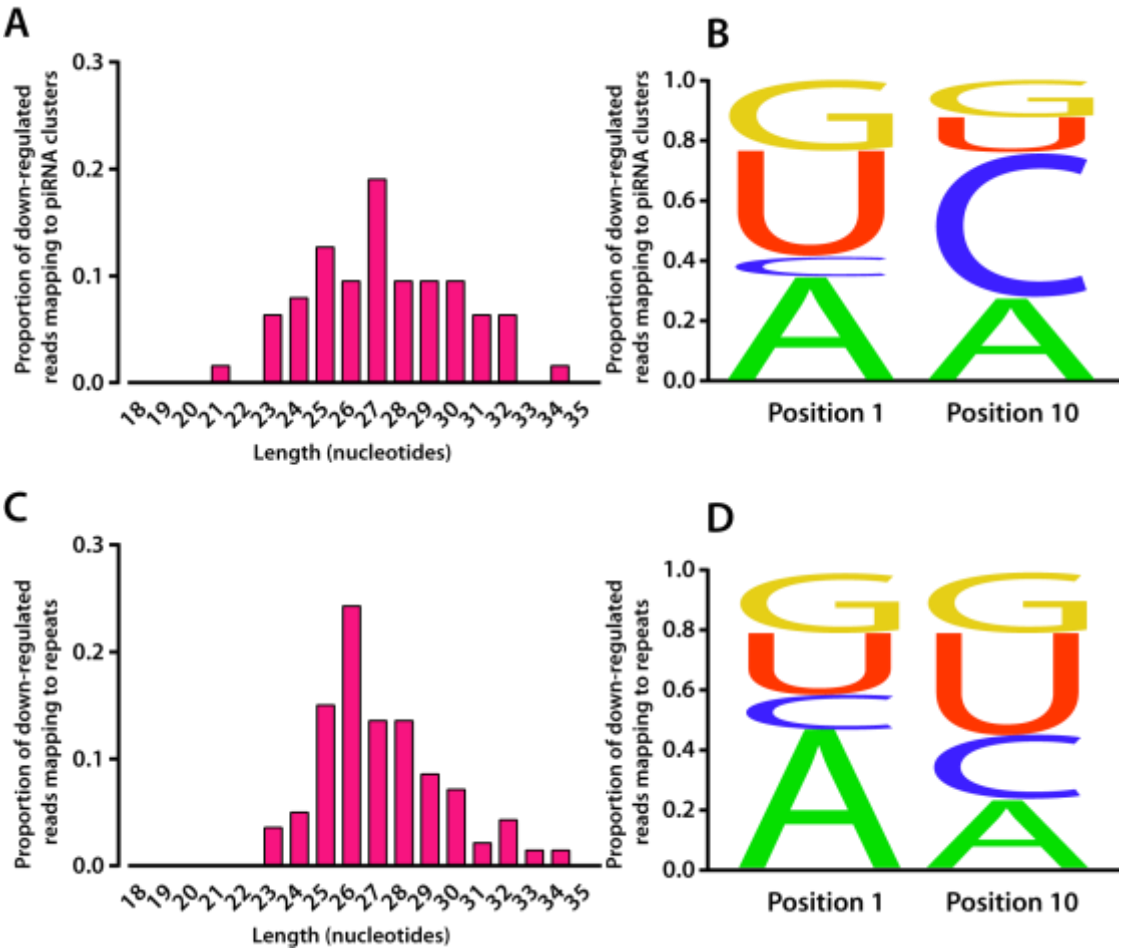


Figure 8-7: Sequence characteristics of piRNA cluster-derived small RNAs that are down-regulated following *Miwi2* knockdown

(A-B) Small RNAs that map to piRNA clusters. (C-D) Small RNAs derived from repeat elements. (A and C) Length distribution and (B) Sequence alignment at positions 1 and 10 of small RNAs that were i) down-regulated >2-fold in R1-pRSIT-Miwi2 cells

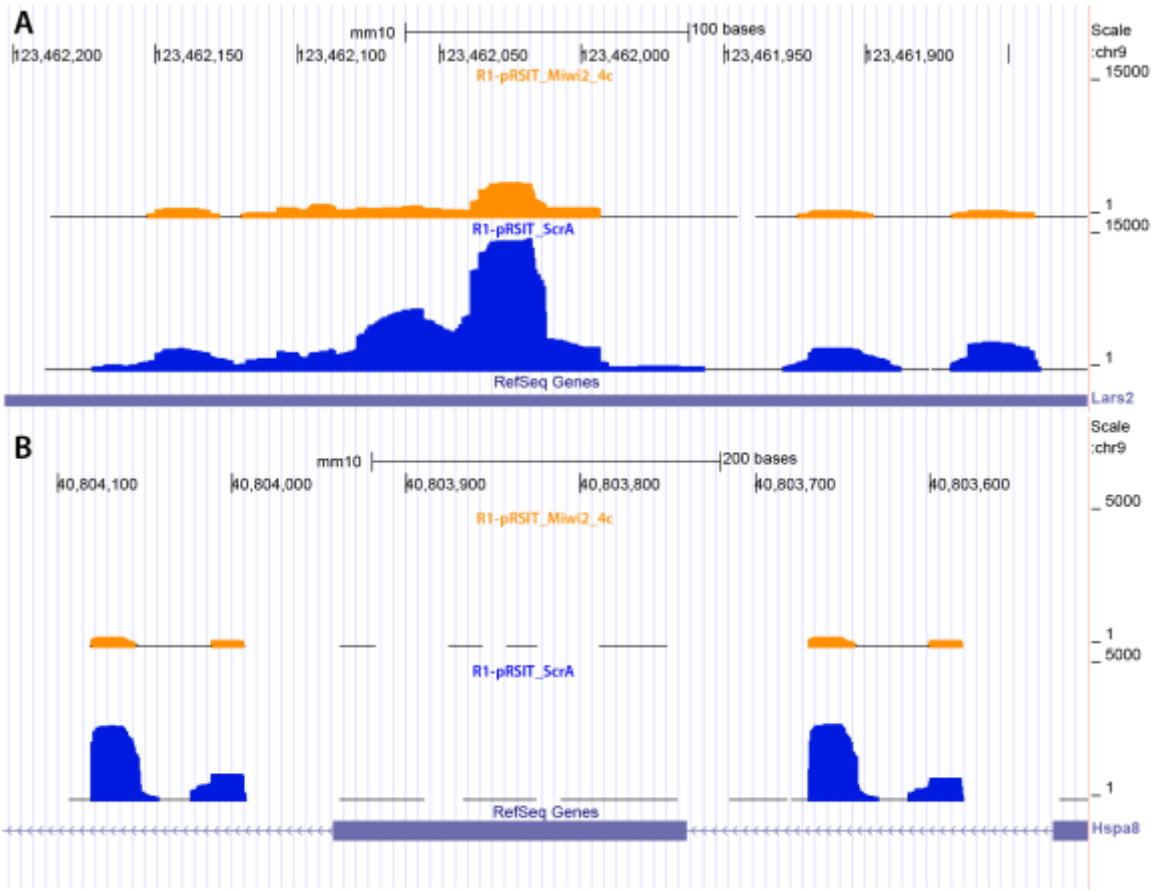


Figure 8-8: Reads mapping to 3' UTR of Lars2 and Hspa8

Graphical representation of small RNAs from scrambled control ("R1-pRSIT-Scr") and *Miwi2* knockdown ("R1-pRSIT-Miwi2") cells in WIG format aligned with the mouse genome (UCSC Genome Browser, GRCm38/mm10). (A) Reads aligned with 3' UTR of *Lars2.* Y-axis max = 15000. (B) Reads aligned with intronic regions of *Hspa8*. Y-axis max = 5000.

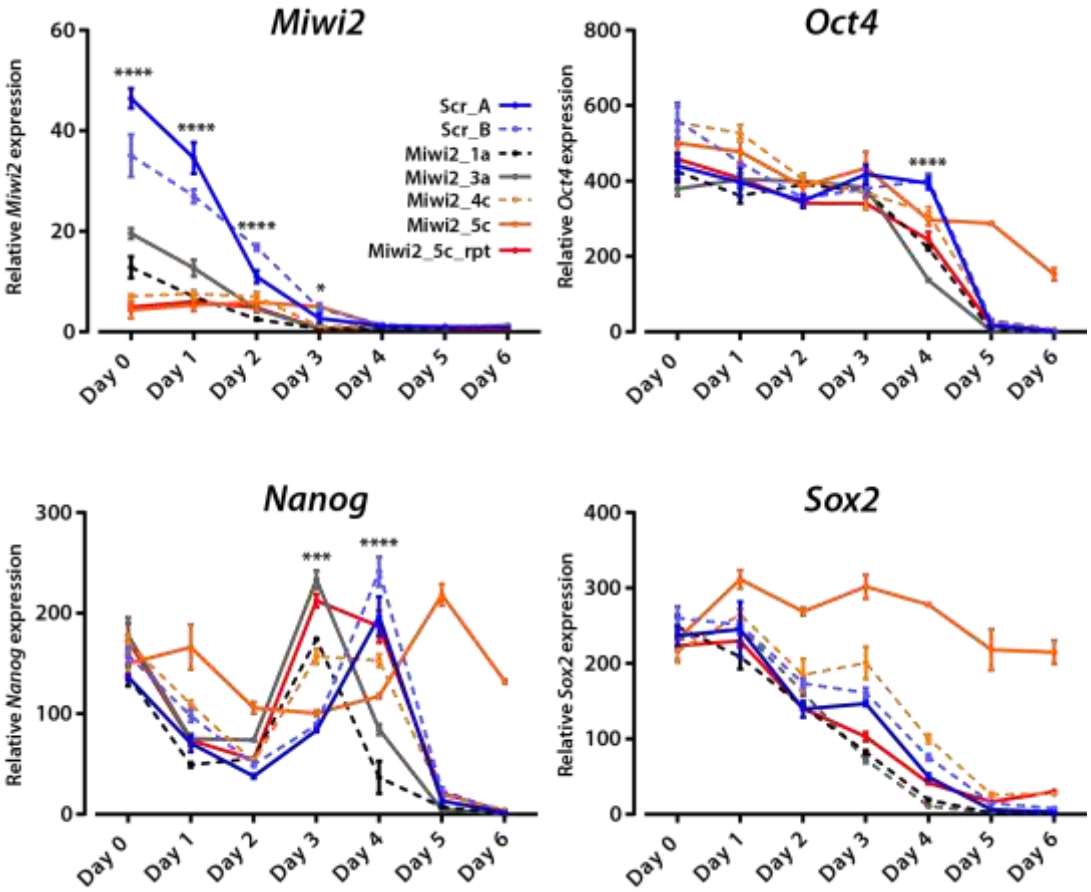


Figure 8-9: Expression of pluripotency genes during *in vitro* differentiation of control and *Miwi2* knockdown ES cell lines

Quantitative RT-PCR data for *Miwi2* and pluripotency gene expression in scrambled control cell lines versus *Miwi2* knockdown cell lines over 6 days *in vitro* differentiation. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant differences in gene expression between scrambled control and *Miwi2* knockdown cell lines (* $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$). Y-axis values are normalised expression values derived from Comparative C_T analysis. Values are different on each axis as transcript abundance varies for each gene. Error bars represent standard error of the mean of replicate cell line expression.

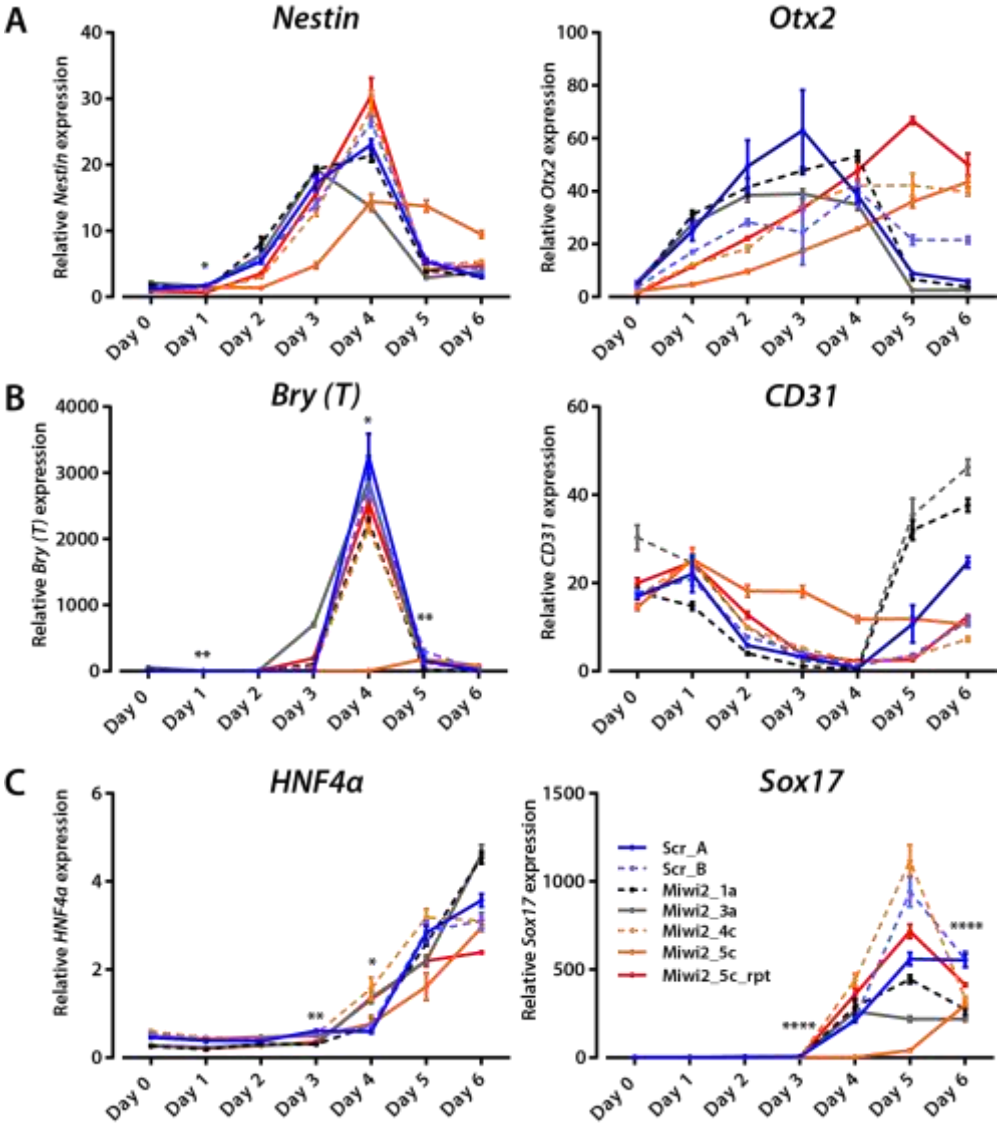


Figure 8-10: Expression of lineage markers during *in vitro* differentiation of R1-pRSIT cell lines

Quantitative RT-PCR data for lineage marker gene expression in scrambled control ("Scr") cell lines versus *Miwi2* knockdown ("Miwi2") cell lines over 6 days *in vitro* differentiation. P-values were determined via student's t-test (two-tailed, unpaired) analysis and represent significant differences in gene expression between scrambled control and *Miwi2* knockdown cell lines (* $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$). Y-axis values are normalised expression values derived from Comparative C_T analysis. Values are different on each axis as transcript abundance varies for each gene. Error bars represent standard error of the mean of replicate cell line expression. (A) Ectodermal markers *Nestin* and *Otx2*. (B) Mesodermal markers *Brachyury (Bry/T)* and *CD31 (PECAM)*. (C) Endodermal markers *HNF4α* and *Sox17*

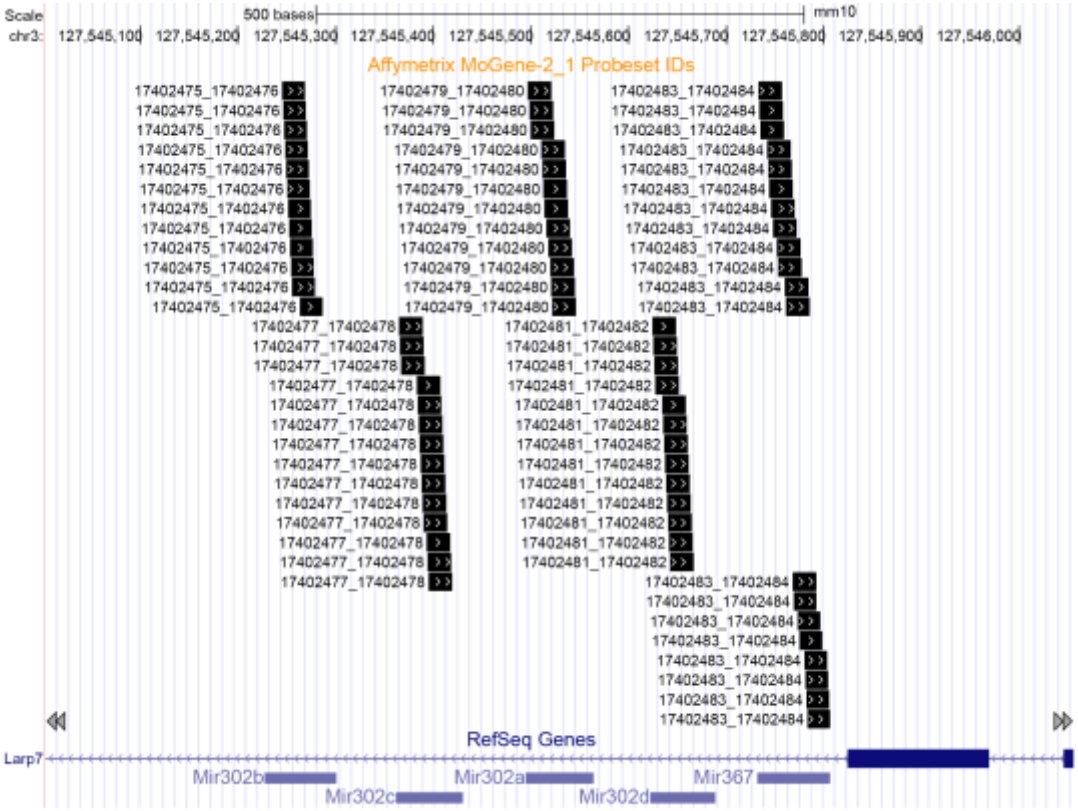


Figure 8-11: Affymetrix probeset aligned with miR-302/367 cluster (UCSC)

Graphical representation of individual Affymetrix MoGene-2_1 probes aligned with the miR-302/367 cluster on the mouse genome (UCSC Genome Browser, GRCm38/mm10).

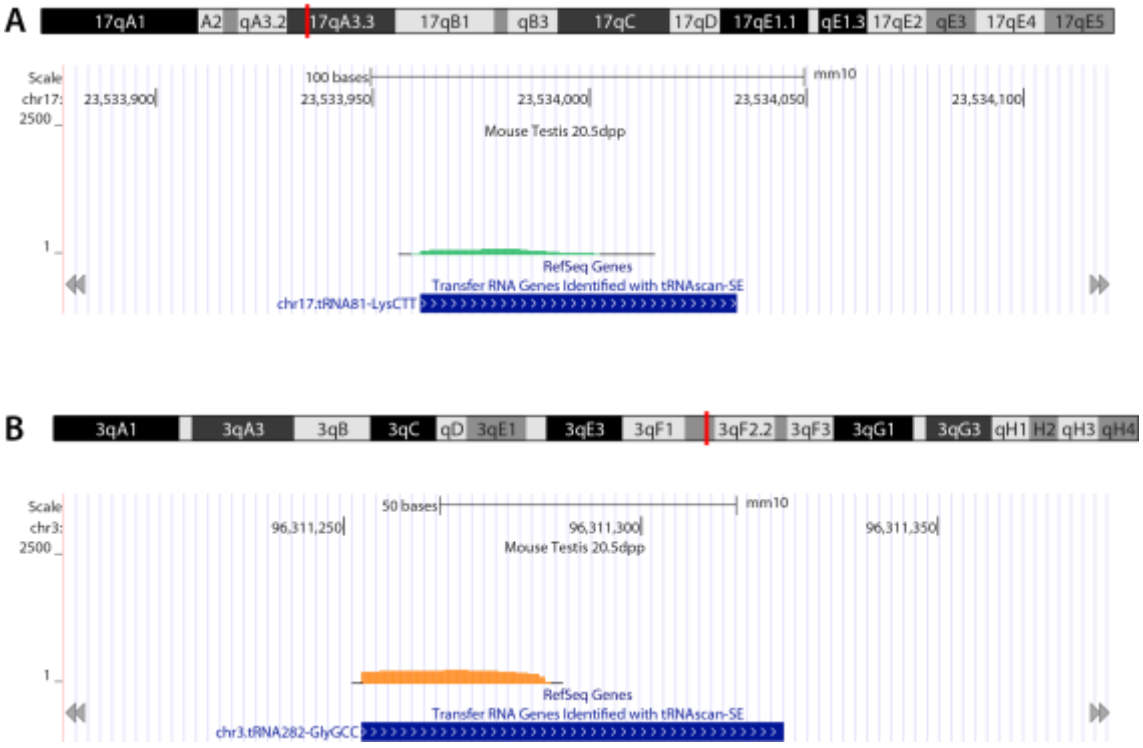


Figure 8-12: Positioning of tRFs at genomic loci in pachytene testis

Graphical representation of tRFs aligned with the genome in 20.5dpp mouse testis (Li et al., 2013). Tracks visualised in UCSC Genome Browser (GRCm38/mm10). (A) LysCTT and (B) GlyGCC. Small RNA reads displayed in WIG format.

Table 8-1: Basic sequencing statistics for all samples

Sample	Processed reads	Too short	% too short	Reads that went into mapping	% that went into mapping	Mapped reads	% Mapped reads	Unmapped reads	% Un-mapped
Day 0 EB	31873571	11600194	36%	20273377	64%	14,917,003	74%	5356374	26%
Day 3 EB	16694728	3208432	19%	13486296	81%	10,255,402	76%	3230894	24%
R1-pRSIT-Scr_A	5255425	2782636	53%	2472789	47%	1,934,085	78%	538704	22%
R1-pRSIT-Scr_B	6319842	3526435	56%	2793407	44%	1,916,987	69%	876420	31%
R1-pRSIT-Miwi2_4c	3317693	2946839	89%	370854	11%	256,322	69%	114532	31%
R1-pRSIT-Miwi2_5c	6124209	4920614	80%	1203595	20%	954,821	79%	248774	21%
R1-pRSIT-Miwi2_all	5103184	4469028	88%	634156	12%	515,681	81%	118475	19%
4TO7-GFPa	10,406,134	1713662	16%	8,692,472	84%	7,693,429	89%	999043	11%
4TO7-GFPb	15,098,553	3226360	21%	11,872,193	79%	10,355,557	87%	1516636	13%
4TO7-Miwi2-GFPa	13,406,473	3169171	24%	10,237,302	76%	8,251,904	81%	1985398	19%
4TO7-Miwi2-GFPb	13,559,740	2065472	15%	11,494,268	85%	9,822,603	85%	1671665	15%

Sample	Processed reads	Too short	% too short	Reads that went into mapping	% that went into mapping	Mapped reads	% Mapped reads	Unmapped reads	% Un-mapped
ES whole	4,555,515	696,867	15%	3,858,648	85%	3,527,345	91%	331,303	9%
ES-piRNA_1	2,536,187	754,225	30%	1,781,962	70%	901,412	51%	880,550	49%
ES-piRNA_2	2,999,452	1,273,825	42%	1,725,627	58%	999,181	58%	726,446	42%
4T1 whole	5,568,309	876,218	16%	4,692,091	84%	4,132,543	88%	559,548	12%
4T1-piRNA_2	5,216,042	199,476	4%	5,016,566	96%	4,502,031	90%	514,535	10%
4T1-piRNA_1	6,235,879	749,604	12%	5,486,275	88%	5,039,403	92%	446,872	8%

Table 8-2: Top genes with expression changes following *Miwi2* knockdown in ES cells and putative function (NCBI Gene)

Up-regulated gene name	Putative function	Down-regulated gene name	Putative function
Nkx6-3	Cell fate determination, regulation of transcription	**Uty**	Wnt signalling, cardiac muscle cell contraction, chromatin modification, embryonic organ development, heart development and morphogenesis, histone H3/K36 demethylation
Psg18	Regulation of interleukin-10 secretion	**Ddx3y**	ATP-dependent helicase activity, RNA binding
SLC17A6	Biomolecular transport	**Calb2**	Cytosolic calcium ion homeostasis
DNAJC15	Negative regulation of mitochondrial electron transport and protein complex assembly, protein transport, regulation of lipid metabolism	**Mmp13**	Bone morphogenesis, collagen catabolism, extracellular matrix disassembly, proteolysis
Mbnl2	Regulation of RNA splicing	**Mmp3**	Collagen catabolism, proteolysis, regulation of cell migration
Ank1	Vesicle transport, erythrocyte development	**Otx2**	Cell differentiation and fate specification, endoderm development, positive regulation of embryonic development and gastrulation
PLA2G5	Lipid catabolism and metabolism, synthesis of platelet activating factor	**Enpp1**	Bone re-modelling, negative regulation of fat cell differentiation, immune response
		Mcpt8	Granzyme-mediated apoptotic signalling, proteolysis
		Rasgrp1	Activation of Ras/Rho GTPase activity, cell differentiation, inflammatory response